Appendix D – Methods

1. INTRODUCTION

This appendix describes the analytical approach and methods, including data acquisition and data analysis details, that were used for the analyses and results reported in the 2018 Impact Report to address the research questions, which guided the work on this report.

Section 1890A(a)(6) of the Social Security Act (as added by section 3014(b) of the Affordable Care Act) provides that the Secretary shall, not later than March 1, 2012, and at least once every three years thereafter, conduct an assessment of the quality and efficiency impact of the use of endorsed measures described in section 1890(b)(7)(B) of the Social Security Act and make such assessment available to the public.[i,1] Previous assessments were reported in the first Impact Report, published in 2012,[2] and the second, published in 2015.[3] The third assessment is documented in this 2018 Impact Report.

2. CONCEPTUAL APPROACH

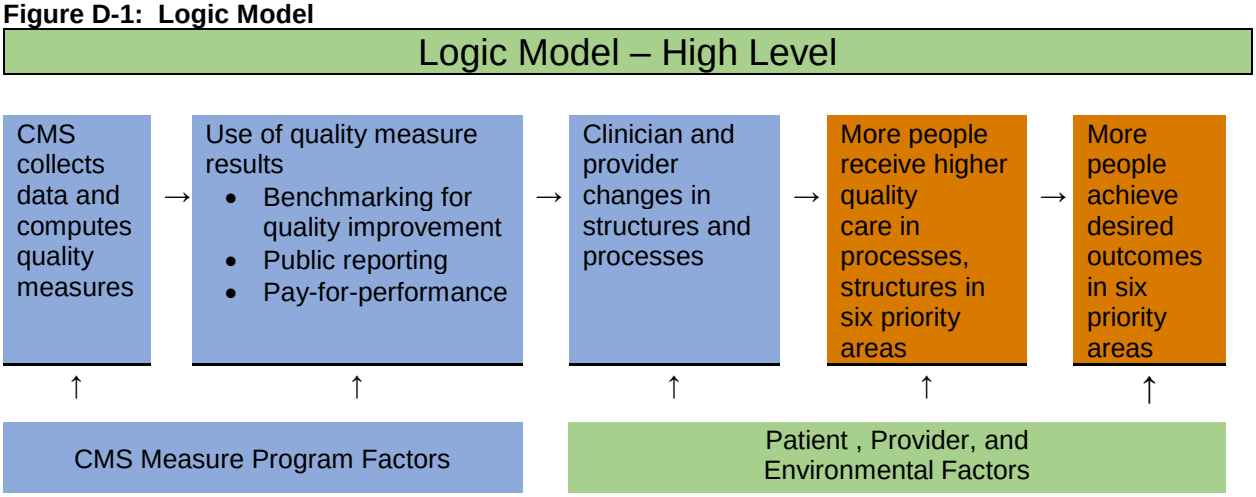
The CMS quality priorities[4] serve as the overarching framework for the 2018 Impact Report. The six quality priorities are:

1: Make care safer by reducing harm caused in the delivery of care.
2: Strengthen persons and their families as partners in their care.
3: Promote effective communication and coordination of care.
4: Promote effective prevention and treatment of chronic disease.
5: Work with communities to promote best practices of healthy living.
6: Make care affordable.

The use of performance measures is one strategy that CMS is employing to achieve the overall goal to put patients first. Performance measures empower patients and doctors to make decisions about their health care, support innovative approaches to quality, accessibility and affordability and lead to an improved experience for the CMS customer. The 2018 Impact Report assesses the impact of the use of measures in CMS Medicare programs that apply to a wide variety of clinicians, providers (acute care hospitals, home health agencies, nursing homes, etc.), and patients. The intended purpose of performance measures is to improve the quality of health care and ultimately patient health, as depicted in Figure D-1. The logic model shows that program factors such as public release of measure analysis results and financial incentives can influence the impact of measures. Likewise, patient, provider, and environmental factors also directly influence provider changes, patients receiving quality care, and desired outcomes.

[i] Section 1890A(a)(6) of the Social Security Act

Figure D-1: Logic Model



For the purposes of the 2018 Impact Report, CMS defines impact as progress in terms of meeting CMS quality priorities and objectives as demonstrated by:

- Trends in performance and disparities on Key Indicators (i.e., selected measures used in CMS Medicare programs that represent the six quality priorities). Both patient and economic impacts (expressed as avoided costs) associated with changes in measure rates were quantified.
- Aggregate trends in performance and disparities on measures used in CMS Medicare programs that meet inclusion criteria, mapped to each quality priority.
- Actions taken by hospitals and nursing homes in response to the use of performance measures, as identified from national surveys of quality leaders from hospitals and nursing homes conducted in 2016.

3. SCOPE OF THE ASSESSMENT

The 2018 Impact Report is a comprehensive assessment of the use of quality measures from CMS Medicare reporting programs. The data for this assessment focus on the measures used in the CMS Medicare programs in 2015 and encompass the period from January 2006 through December 2015. For purposes of data acquisition, data by program was requested by coordinating with the CMS program leads. For additional program detail please see *Description of CMS Quality Measure Programs* (Appendix B).

For some aspects of the methodology involving assessment of performance measure gaps, measures identified in final rules as of December 2016 are also included. For purposes of quantifying measures in use, an unduplicated count by setting or plan was performed. For example, the CHF readmission measure, which is in two hospital measure programs (Hospital Inpatient Quality Reporting and Hospital Readmissions Reduction), is counted only once.

This assessment does not duplicate individual program evaluations that assess the effect of specific program features. The assessment also does not duplicate the evaluation of the reliability and validity of the measures implemented in CMS programs that occurs during measure development and consensus-based endorsement.

4. GENERAL APPROACH

For the impact assessment, national results were used for the implemented measures—either for all patients or for various subpopulations—to analyze trends in improvement and identify disparities. This approach focused on the impact on patients rather than profiling of providers, which is conducted at the program level.

The assessment relied upon data provided by CMS personnel and contractors that collect data and compute the measures for CMS programs. Measure data received at the beneficiary or provider level were aggregated to calculate national rates as well as stratified subpopulation rates for each year of data availability. Data received at the national level were used as provided by CMS data owners.

Measures that had at least three annual results were included in the analyses of trends in improvement and trends in disparities. The most recent results were used to test for any disparities. Measures with fewer than three data points were included in the analysis for disparities. When feasible, analysis of outcome measures included an adjustment using age and sex to account for changes in the population over time or among various population subgroups.

Analysis of measures and the 2018 Impact Report was organized by the six quality priorities. The report has separate chapters for each quality priority, featuring National Quality Dashboards with analysis results for selected measures (Key Indicators) as well as aggregate results for all measures within the quality priority. Separate chapters report the findings of the national provider survey of hospitals and nursing homes as well as overall trends and disparities analysis results for all measures analyzed. The final chapter summarizes the findings presented in the preceding chapters and discusses potential future directions.

Standardized classification rules were applied to assign the measures to six CMS quality priorities. Measures may meet the criteria for more than one priority. The objectives and desired outcomes from the patient perspective were considered for each of the CMS quality priorities when assigning the measures. Additionally, feedback from the project technical expert panel (TEP) that was convened by the contractor and the Federal Assessment Steering Committee (FASC) was considered when assigning measures to a quality priority. Measures used in CMS programs are often categorized by CMS quality priority in final rules; however, different CMS programs may categorize the same or similar measures differently. All the same or similar measures were assigned to the same priority; therefore, the quality priority assignment may differ from what has been published in the *Federal Register*.

5. RESEARCH QUESTIONS

The following research questions guided the work for this report. Table D-1 lists the research questions and identifies the basic analytic approach associated with each question.

Table D-1: Research Questions and Basic Analytic Approach

#	Research Question	Basic Analytic Approach
1	What measures currently implemented in CMS programs are most closely aligned with the CMS quality priorities and related goals and objectives and could be considered Key Indicators to assess progress?	• Categorize measures according to HHS Decision Rules. • Select Key Indicators with TEP and FASC input.
2	To what extent did performance on measures implemented in CMS programs improve over time?	• Conduct a quantitative analysis of trends in national rates over time (minimum of three data points). • Develop a National Quality Dashboard of results for Key Indicators. • Aggregate trends on all measures by quality priority. • Summarize results across all quality priorities.
3	For Key Indicators, what is the impact of changes in performance on patients (e.g., number of patients affected) and health care costs likely avoided?	• Estimate the impact, by measuring the difference between the observed and expected outcomes, for the relevant Key Indicators. • Conduct a targeted literature review to identify studies that have estimated the health care costs associated with the outcomes for the relevant Key Indicators. • Obtain TEP and FASC input related to identifying Key Indicators appropriate for impact, health care cost analysis, priorities for analysis, methods, and insights into previous studies. • Conduct a quantitative analysis combining the impact estimate with the health care cost estimate associated with the health outcome, derived from previous studies where feasible.
4	For measures implemented in CMS programs, did disparities among identified subpopulation groups exist, and did the disparities change over time?	• Conduct a quantitative analysis of differences by identified subpopulations (age, sex, race/ethnicity, urbanicity, geographic location, and income level, where data was available). • Aggregate trends on all measures by quality priority. • Summarize results across all quality priorities.

#	Research Question	Basic Analytic Approach
5	What is the environmental context that may affect interpretation of the impact of measures used in CMS programs?	• Analyze descriptive information to enhance interpretation and understanding of the findings, including *Related Programs and Initiatives* – (Appendix H). ○ A quantitative summary of the number and types of measures as well as the settings and programs in which they are used. ○ A targeted environmental scan and/or literature review to identify other initiatives besides implementing measures that can influence achievement of CMS quality priorities (e.g., legislation/regulation, quality improvement [QI] focus, and Innovation Center payment models). ○ For the Key Indicators, a targeted literature review to identify similar trends.
6	What additional measures are needed to assess progress toward achievement of the CMS quality priorities (i.e., meaningful measure gaps)?	• Conduct a qualitative analysis and synthesis of TEP and FASC input during the selection of Key Indicators.
National Provider Surveys		
7	What changes are providers making in response to use of performance measures by CMS? • What types of changes or innovations have hospitals/nursing homes made to improve their performance on CMS measures? • If a change or innovation was made, has it helped the hospital/nursing home improve its performance on one or more CMS measures? • What challenges or barriers do hospitals/nursing homes face in reporting CMS quality measures? • What challenges or barriers do hospitals/nursing homes face in improving performance on CMS quality measures? • What unintended consequences do hospitals/nursing homes report associated with implementation of CMS quality measures?	• Conduct a survey of hospital and nursing home providers. • Conduct qualitative interviews of smaller group of hospitals and nursing homes. • Link performance data for the respondent providers.

6. LIMITATIONS

The limitations of this impact assessment are acknowledged as follows:

- Data: Data required to perform quantitative analyses and to adjust outcomes over time may be incomplete because of varying data collection requirements across programs or varying ability of CMS personnel contractors to provide the data as required.
- Performance measure gaps: Measures used in CMS programs that are available for analysis do not align with every stated objective in the CMS quality priorities or may not include all the affected patients and health care settings.

- Attribution: The CMS quality priorities include a variety of strategies beyond performance measurement; however, this assessment cannot empirically isolate the effect of performance measures from other factors that may influence the achievement of goals for quality of care and desired outcomes.
- Limited number of studies that include estimates of health care costs: Valid published literature may not be available to quantify health care costs avoided for the Key Indicators identified appropriate for cost-avoided analysis.
- Patients represented multiple times: Measure rates are likely to include multiple encounters for some patients and may therefore be a biased representation of true measure rates at the patient level to the extent that there is double counting of sicker patients in the measure population. Additionally, double counting may occur if the same patient appears in multiple measures for the same topic area (e.g., influenza immunization measures for physician, specialist, and hospital). The 2018 Impact Report assesses national rates for the measures as specified without adjustment for double counting of beneficiaries. If aggregate rates are considered for a specific topic, whether this bias could be addressed was considered.
- The PQRS measures used in this report were limited to measures being considered as potential Key Indicators. For PQRS individual clinicians, only data from 2014 and 2015 were included to align with the reporting periods established for Physician Compare. PQRS clinician group results in this report are based on the measures submitted through the web-based tool, beginning in 2012, to align with Medicare Shared Savings Program ACO data, which uses the same data submission mechanism and measure specifications.[ii]

7. DATA

The data for the analyses used in the 2018 Impact Report came from measures implemented in CMS Medicare programs. This section offers an overview of the identification and acquisition of the data utilized for this report.

7.1. DATA IDENTIFICATION

To determine which data sources were needed, measures used in CMS Medicare programs were identified by searching Medicare program websites, Compare sites, final rules published in the *Federal Register*, and the CMS Measures Inventory.

The criteria for inclusion were:

- Used in a CMS program as of 2015.
- Included at least one year of data expected by the end of 2015.
- Used for public reporting or payment incentive programs.

[ii] Prior to 2014, in the early years of the PQRS, data submission requirements and methods changed significantly, while clinician participation increased significantly each year. By 2014, the program and data had stabilized and CMS began reporting individual clinician data performance rates on Physician Compare.

The criteria for exclusion were:

- Used in reporting-only programs, designated as display measures, or available only to clinicians and providers.

The *Overview of CMS Measures Included in Analyses* (Appendix A) shows the measures included in the 2018 Impact Report and includes a flag indicating which measures were in quantitative analyses.

7.2. DATA ACQUISITION

To acquire the large amount of data for implemented measures and ensure data integrity and quality, a detailed data request was made accompanied by table shells that each CMS data owner and associated contractor(s) completed. Table shells represented suggested file layouts, which included all the data elements required for analyses. All annual data available between 2006 and 2015 were requested.

To allow for flexibility in completion of a table shell, while still enabling collection of the necessary data to support the proposed analyses, data at the beneficiary level that would allow for data aggregation and stratification for disparity analyses were requested. If contractors could not provide beneficiary-level data, provider-level or national-level data were accepted if beneficiary data were not able to be provided. Where feasible, the CMS Enrollment Database (EDB) was used for patient demographic information that was missing from the files received from CMS and its contractors. If CMS and contractor data could not feasibly be obtained, data available for public reporting, such as found Data.Medicare.gov or Part C and D Performance Data were used.

7.3. EXTERNAL DATA

The project scope requires quantitative analysis of measures implemented in CMS Medicare programs; however, other sources of information were used to assist with interpreting the findings and quantifying impact. The literature was reviewed to identify studies that include estimates of health care costs that align with the relevant Key Indicators, to estimate health care costs avoided, as called for in Research Question 3. Published results for similar measures were used to discuss the findings on the Key Indicators, as called for in Research Question 5.

8. RESEARCH QUESTION 1

What measures currently implemented in CMS programs are most closely aligned with the CMS quality priorities and related goals and objectives and could be considered Key Indicators to assess progress?

Sources of information to identify programs for the 2018 Impact Report included final rules published in the *Federal Register*, CMS press releases, and the CMS.gov website. Programs that met the following inclusion criteria were used in an environmental scan and are included in the 2018 Impact Report. The criteria for inclusion were:

- The CMS Medicare program uses quality measures.
- The program was implemented by December 31, 2015.

- The program was nationally implemented (i.e., programs of limited scope such as Innovation Center pilot and demonstration projects were not included).

A list of measures used in CMS programs by identifying measures finalized in rules and/or publicly displayed was assembled. With input from the TEP and FASC, standardized classification rules were applied to assign the measures to six CMS quality priorities.[5] The assignments were made for the purposes of the 2018 Impact Report and may vary in some instances from prior classifications of the measures in CMS rulemaking.

An unduplicated count of measures by setting or plan (e.g., hospital, nursing home, physician office, health plan, ACO) was used. For example, if one measure is used in three hospital programs, such as Hospital Inpatient Quality Reporting, Hospital Value-Based Purchasing, and Hospital-Acquired Condition Reduction Program (HACRP), the measure is counted once.

A rule for counting patient experience measures from Consumer Assessment of Healthcare Providers and Systems (CAHPS®) was established. Each individual measure from the CAHPS® surveys was counted as a measure, such as the composite and global rating measures used on Compare sites.

8.1. KEY INDICATORS

A list of 762 measures used in CMS Medicare quality programs in 2015 was compiled by the CMS project contractor from information posted on the CMS website and from final rules posted in the *Federal Register*. Of these 762 measures, 253 were expected to have a minimum of three years of data available for trend analysis. The 253 measures, categorized by CMS priority area, were analyzed by CMS and the contractor for possible use as Key Indicators. A Key Indicator is defined as a measure or group of measures used to gauge performance on aspects of CMS quality priorities. The Key Indicators represent each of the CMS quality priorities and are used to create a National Quality Dashboard to report progress. CMS and the contractor evaluated the 253 measures using the following criteria:

- A sound conceptual basis exists for representing achievement of a specific CMS quality priority or related strategic result or objective.
- Policymakers, clinicians and providers, patients, and caregivers can easily understand the overall goal of the measure.
- The measure includes a large or representative portion of the population or a smaller subpopulation that provides a strong signal of overall quality or efficiency.
- The measure is useful for monitoring progress over time.

The resulting 128 candidate Key Indicator measures were aligned with 24 Key Indicator concepts and presented to the TEP and FASC. Using the same four criteria, the TEP and the FASC reviewed and rated the 24 proposed Key Indicator concepts using a 1–9 point Likert scale. The TEP and the FASC discussed their initial assessments of the proposed Key Indicators at meetings in Baltimore, MD, on June 18, 2015, and November 12, 2015 (for quality priorities 1–3 and 4–6, respectively). After their discussion, the panels again rated the Key Indicators on the same criteria to prioritize for further consideration.

Following the in-person meeting, workgroups composed of five to seven TEP and FASC members reviewed additional information about the prioritized potential Key Indicators and

considered issues raised by the TEP and the FASC. The workgroups made recommendations on whether to retain measures as Key Indicators and discussed potential issues for the analysis and interpretation of findings. They also recommended additional measures as Key Indicators. The workgroups presented their recommendations for Key Indicators for quality priorities 1–3 to the full FASC and TEP for additional input at the November 2015 meeting and for quality priorities 4–6 at the May 16, 2016, meeting. Final recommendations for Key Indicators were discussed and approved by the TEP and FASC at the August 2017 meeting. The following exclusion criteria were applied to develop the final list of measures reported by programs to be included in the Key Indicators:

- Topped out as defined by the CMS program in a final rule
- Removed from the CMS program for any reason; includes being removed from a similar or related program
- No data received or data not in the format requested
- Fewer than three annual reporting periods of data were available for use
- Measures submitted to the National Quality Forum (NQF) that either failed endorsement, failed to receive endorsement at the maintenance review because the measures failed the maintenance review, or were not submitted for endorsement maintenance; includes measures put on reserve status by NQF during maintenance
- Denominator of the measure includes few Medicare beneficiaries or is reported by few providers
- Does not directly relate to a CMS quality priority objective
- Measures address a single aspect of care for a complex condition
- Other NQF-endorsed measures or outcome measure Key Indicators address the same condition.

After iterative deliberations and application of the exclusion criteria, 62 measures aligned with 28 Key Indicators were selected.

Overview of CMS Measures Included in Analyses (Appendix A) identifies all measures included in the analysis, including measures used by Medicare quality programs that were selected as Key Indicators.

9. RESEARCH QUESTION 2

To what extent did performance on measures implemented in CMS programs improve over time?

The trends for measures with at least three annual data points were examined. For health outcome measures, the national measure rates were age- and sex-adjusted to control for changing population demographics over time, where feasible. For other measures, the observed national rate was used for trending analysis. To ensure that the measure rates being analyzed are reliable estimates for performance, measures were excluded for which the sample sizes were small, and measures for which the observed data are unreliable. Measures based on means, medians, and ratios were excluded if more than 50% of annual data points had fewer than 30 observations. Measures estimated as means, medians, and ratios with a relative standard error (RSE) greater than 30% for at least half of the annual data points were excluded from analysis due to unreliability in the data. The RSE is calculated as:

$$RSE = \left(\frac{Standard\ Error}{Rate}\right) * 100 \quad (1)$$

Finally, for proportion-based measures there must have been a minimum of 5 successes and 5 failures: n(p) >= 5 and n(1-p) >= 5, where n is the sample size and p is the proportion for the measure rate. Measures based on proportions were excluded if fewer than 50% of annual data points have more than 30 observations, and met the n(p) and n(1-p) criteria.

9.1. ADJUSTMENT OF OUTCOMES FOR POPULATION CHANGES OVER TIME

The outcome measures data used for trend analyses were adjusted for changes in the age and sex composition of the population over time, where feasible. Age and sex are associated with differences in clinical conditions and comorbidities, and age is especially relevant to the development of chronic illness. Direct standardization for age and sex across measurement periods controlled for the influence of major changes in the population on rates over time. The composition of the measure data in the baseline year for age and sex was used as the reference population for the standardization process. Age standardization was based on the age categories specified for the disparity analysis discussed in Research Question 4. Age and sex standardization was performed for outcome measures where beneficiary level data were available and were not already risk-adjusted, but was not performed on process or intermediate outcome measures.

9.2. TRENDING ANALYSIS

Trending analysis was performed for measures with at least three annual data points using a log-linear regression model with the following form:

$$ln(y_t) = \beta_0 + \beta_1(Time_t) + \varepsilon_t \quad (2)$$

In Equation 2, the dependent variable is $ln(y_t)$, the natural log of the measure rate *y* at time *t*, where *Time* ranges from 0 to 1 less than the number of annual data points provided for the measure.[iii] The constant, β_0, represents the measure rate in the initial period of observation during the study period. The coefficient, β_1, represents the average annual change in the logged outcome rate. The error term ε_t is assumed to be i.i.d. $N(0, \sigma^2)$.[iv]

A measure trend was defined as improving if [exp(β_1) – 1]*100 was larger than +/- 1%. Additionally, a trend was defined as improving rapidly if [exp(β_1) – 1]*100 was larger than +/- 10%.

[iii] Where y_t is zero, the natural log is undefined. Therefore, if a quality measure exhibits a rate of zero, the natural log transformation includes the addition of a small constant to all records: ln(y_t) + (.05* $\overline{y_t}$), where $\overline{y_t}$ is the average rate. The addition of this small constant prevents the logged rate of zero from being undefined, and by generating the constant using the average rate, the final constant was appropriately small so as to not unduly influence the results of any analysis.

[iv] This methodology is similar to the methodology used by the AHRQ 2014 National Healthcare Disparities Report, Introduction and Methods, Publication No. 15-0007-3-EF, April 2015 http://www.ahrq.gov/research/findings/nhqrdr/nhqdr14/intro.html.

9.3. EXPRESSING CHANGE

In Equation 2, β_1 can be transformed to represent the percentage change in the outcome measure rate associated with a one-year increase in time. Specifically, the transformation [exp(β_1) – 1]*100 yields the percentage change for the outcome rate resulting from a one-year increase in time. The β_1 coefficient accounts for the change in the average level of the measure rate, allows for direct comparisons of the relative changes exhibited by measures over time.

9.4. BENCHMARKING

Benchmark levels of performance were calculated using data from the first available data point for each measure. Benchmarks were calculated for measures where provider-level data was available, the sample size was large enough to warrant the analysis, and the measures were not cost related. The benchmark rates were calculated using the Achievable Benchmarks of Care™ (ABC™) methodology.[6] The ABC methodology ranks providers in order of performance rates from best to worst. Next, the patient populations, beginning with the best-performing providers, are summed until at least 10% of the total patient population or sample is included. The benchmark rate is then calculated as the performance rate for patients in this subset, representing the approximately 10% of the patient population or sample. Of note, these benchmark results may differ from benchmark results used by various CMS programs.

Changes in measure rates after the first data point were compared with the benchmark rate as an assessment of improvement. The comparison consisted of the difference between the measure rate and the benchmark level. Benchmark levels were calculated using only the national-level rates and were not calculated using subgroups in the disparity analysis. Benchmark results are labeled as "Achievable Results" in the dashboards in the report.

10. RESEARCH QUESTION 3

For the Key Indicators, what is the impact of changes in performance on patients (e.g., number of patients affected) and health care costs likely avoided?

The analytic framework described in Research Question 2 allowed for the assessment of the direction, magnitude, and statistical significance of measure trends related to changes in rates of performance. While analyzing changes in performance rates over time is useful, it is also important to understand the absolute impact of the changes on patients and on health care costs. Therefore, this report also evaluated the impact of changes in performance by describing the number of patients affected and lives saved over time, as well as by estimating the health care costs avoided associated with reductions in safety events and improved disease conditions. For any indicators where the impact represents an adverse change, the number of patients affected and associated health care cost increases were described and estimated, as appropriate.

Therefore, two primary types of impact were estimated from changes in performance: patient impact and health care costs avoided. This section describes the approach used to estimate both types of results, as well as the limitations to the analytic framework related to this research question.

10.1. PATIENT IMPACT

Patient impact is described in terms of patient-level events, such as receiving a treatment, avoiding an adverse event, or experiencing a disease condition. The most direct way to estimate the impact on patients for a change in measure rate is to calculate the number of patients affected. Patient impact estimates were calculated only for measures where trending analyses showed an improving or declining performance. Limitations to this analysis are included below.

An estimate of patients impacted was extrapolated by applying the measure rate from the beginning of the data series to the denominator number of eligible cases in each period. The extrapolation generated the expected number of numerator cases for that period of data, if the measure rate had not changed since the beginning of the data series, as shown in Equation 9:

$$Expected\ Numerator\ Patients_t = Rate_{baseline} * Eligible\ Patients_t \quad (9)$$

The difference between the observed number of patients and the expected number of patients generated through extrapolation represents the increase or decrease in patients either exposed to a new process of care, exhibiting a disease condition, or expressing an experience of care, as shown in Equation 10:

$$Patients\ Impacted_t = Observed\ Patients_t - Expected\ Patients_t \quad (10)$$

The number of patients impacted derived from Equation 10 was then summed across the annual time points for which data is available from the second observed time point until the last to determine the net number of patients impacted as defined in Equation 11 below:

$$Net\ Patients\ Impacted = \sum_{t=2}^{t} Patients\ Impacted_t \quad (11)$$

Interpreting the number of patients impacted depends on the desired direction of change for the measure. In the case of most outcome measures, for which a lower rate is generally preferred, a negative number for *Patients Impacted* in Equation 10 indicates an improvement in performance, or fewer patients experiencing the outcome than would have been expected based on rates in previous years. In the case of many process measures, for which a higher rate is often preferred, a positive number of *Patients Impacted* in Equation 10 would indicate an improvement in performance. When the direction of change for either an outcome measure or a process measure reflects a deterioration in performance, the impact on the number of patients was estimated.

For some measures, the eligible population represents patients with a chronic condition, where inclusion in the numerator in one year is likely to lead to inclusion in the numerator in subsequent years (e.g., hypertension under control). For other measures, multi-year rolling averages were used to calculate the rates for the last year in the rolling window. In both circumstances, calculating a cumulative patient impact would require double counting patients that appear in multiple years of the measure. In these situations, the patient impact is estimated as the absolute difference between the first and last year in the data series.

The reported patient impact estimates were also rounded based on the following rounding rules:

- If in millions, rounded to nearest 100,000: i.e., 3.4 million
- If in 100,000s, rounded to nearest 10,000: i.e., 490,000
- If in 10,000s, rounded to nearest 1,000: i.e.: 29,000

10.2. HEALTH CARE COST AVOIDED

The health care cost avoided related to changes in performance was estimated by multiplying the results of the patient impact analysis by an estimated health care event cost related to the associated harm, prevention strategy, or disease condition measured by the Key Indicator. Health care costs were estimated from the perspective of the insurer/payer, as opposed to the perspective of the hospital, clinician, or patient or societal perspective. As the perspective of a health care cost avoided analysis drives the costs considered, this is an important distinction. The insurer/payer perspective was selected for several reasons:

1. This project is an impact assessment of CMS quality measures from the perspective of CMS, an insurer/payer.
2. The insurer/payer reimburses hospitals and clinicians for health care services across all settings.
3. Not all changes in performance measures impact the costs to the hospital or clinician only.
4. Estimating health care costs from the patient perspective is beyond the scope of the study, as this perspective would include health care cost estimates such as paid insurance premiums and deductibles and lost time from work.
5. Assessing health care costs from the societal perspective is also beyond the scope of this project, as this perspective would include the patient and the provider perspective in addition to the insurer/payer perspective.

The health care cost avoided analysis is not a cost-benefit analysis, cost-effectiveness analysis, or return on investment analysis. The costs avoided (increase) would be considered a benefit (cost) related to the improvement (decrease) in quality. Identifying the program costs related to the implementation of quality improvement programs by hospitals, clinics, CMS, and others is not part of the scope of this study. Without these quality program implementations costs, cost-benefit, cost-effectiveness or return on investment analyses cannot be conducted. Therefore, the focus of this research question is on the impact of changes in quality on health care costs from the perspective of the insurer/payer for reimbursements to providers.

Not all Key Indicators are appropriate for a health care cost avoided estimation, as some indicators measure a process that cannot be specifically linked to a health benefit or the value of the improvement is a challenge to quantify, such as reduced mortality. Therefore, the following inclusion and exclusion criteria were developed to identify relevant Key Indicators for health care cost avoided analysis:

Inclusion Criteria

- Health outcomes – including intermediate health outcomes
- Process measures closely linked to outcomes (based on evidence), such as vaccinations (treatment)

Exclusion Criteria

- Mortality outcomes
- Measures expressed as payments or costs
- Process measures that are not proximal to health outcomes
- Patient experience measures
- Indicators where patient-level data are not currently available to estimate patient impact

- Composite indicators where data are not available to distinguish patient level data between the group of indicators that comprise the composite indicator

The resulting Key Indicators represented the following quality priorities: Patient Safety, Care Coordination, Healthy Living, Effective Treatment, and Patient Engagement. Key Indicators from the Affordable Care quality priority were excluded from the cost avoided analysis as these measures represent payment and resource use measures that are not linked to a health outcome.

Once the Key Indicators eligible for health care cost avoided estimation were identified, a targeted literature search was conducted to estimate the relevant health care event cost to estimate value of provider reimbursement. Targeted literature searches provided the basis for quantifying the costs avoided for the selected performance measures. Literature searches focused on systematic reviews, meta-analyses, and frequently cited articles to identify estimates for costs per event. Literature related to health care events were identified using key search terms that align with the improvements targeted by the quality measure, such as preventing the need to treat patients experiencing specific adverse events or diagnosed conditions. Other sources of the targeted literature review included:

- NQF measure submission forms
- National Quality Measures Clearinghouse
- National Guideline Clearinghouse
- Agency for Healthcare Research and Quality (AHRQ)
- CMS published reports
- TEP and FASC workgroup recommendations

As studies and reports were identified, they were reviewed to determine whether they met the criteria outlined below.

Study Inclusion Criteria

- Studies quantifying event costs from the payer perspective

Study Exclusion Criteria

- Non-U.S.-based studies
- Studies based on data prior to 2006
- Studies with measures not aligned with Key Indicators

As studies and reports were collected, they were reviewed and assessed for quality and scored. The factors considered in reviewing studies for quality included:

1. Alignment with the Medicare population (1 = Not well aligned, 2 = partially aligned, 3 = well aligned)
2. Alignment with Key Indicator measure (age group, condition/strategy definition) (1 = Not well aligned, 2 = partially aligned, 3 = well aligned)
3. Methodological Rigor (1 = Not rigorous, 2 = some rigor (methodological design), 3 = rigor (pre-post, control group, randomization))

Based on these three characteristics, 9 possible points were available, which would represent the highest quality study. The total points per study were used to rank the studies from highest to lowest for each type of condition/event, when there was more than one study identified that meet the exclusion and inclusion criteria. While there is subjectivity in this ranking of study methodology, the ranking is intended to provide guidance into the decision between more than

one study and the value of the cost per event. In cases where there is more than one high quality study that can be used to estimate a cost per event, a range in values for the health care cost impact is reported.

Once the value for the cost per event was estimated, it was converted into 2015 dollars, using the medical care services index (MCS) report by the Bureau of Labor Statistics.[7] Though the MCS index is available to convert into 2016 dollars, 2015 was selected, as the latest data available for the patient impact was 2015.

The results of the cost per event, which were often developed as a cost range, were multiplied by the results of the patient impact analysis for the relevant measure. For measures that use patient sampling methods, population patient impacts were estimated by applying the sample weights to the relevant population.

Two approaches were applied to estimate the patient impact over time. Based on the time frame of the measure, the net patient impact was either determined by summing the differences between the observed and expected for each year or calculated as the difference between the first and last year. The first/last year differences were relevant with indicators where a significant proportion of patients could be included in the numerator in consecutive years. For both approaches, the patient impact was multiplied by the estimated range in health care cost per event, to estimate the health care cost impact. If the net patient impact resulted in an improvement in quality, this product represents health care costs avoided. If the net patient impact resulted in a reduction in quality, this product represents an increase in health care costs, from the insurer/payer perspective.

10.3. LIMITATIONS OF IMPACT AND COST-AVOIDED ANALYSES

The methods identified for use in these analyses are limited in several ways. Most notably, the estimates are only as accurate as the estimated patient impacts included in the analysis. To the extent that patient impacts calculated are not completely due to changes in quality, the estimates may overstate the impacts calculated. In contrast, to the extent to which there are other factors not included in the analysis that may have increased measure rates, this analysis understates the impacts.

Second, the data related to some patient impacts, such as those related to quality-of-life (e.g., physical mobility, ability to work, pain, and time spent interacting with family) are not readily available for analysis. These types of quality measures cannot be directly linked to the Key Indicator data without further data collection and analysis. While such metrics represent valid aspects of how the health care system influences patient lives, the 2018 Impact Report does not assess patient impacts beyond those directly captured in the available quality measure data sources.

Third, some measures capture outcomes related to patients that would be expected to be eligible for inclusion in multiple years of data. For example, a patient with well-controlled diabetes would be expected to be included in a measure of HbA1c less than 7% over multiple years. As quality measure performance rates for patients with chronic conditions improve or decline over time, the current methodology for calculating patient impacts is limited in its ability to detect patients who appear in multiple years of data. When calculating the cumulative patient impact over time, this limitation may overestimate the number patients impacted. As an alternative to

the cumulative patient impact over time, where measures are expected to capture the same patients in multiple years, the patient impact was calculated based on the difference between the baseline year and most recent year of data.

The health care cost impact is also limited in that the program costs related to quality and performance improvement efforts are not included in the analysis. While there are some cost estimates related to some individual level quality improvement initiatives, this information is not available for the specific settings that have reported results as represented by the Key Indicators. Estimation of the program implementation costs related to performance improvement would require primary data collection, which is beyond the scope of this study. Therefore, the value of the health care costs that are avoided could be overestimated or underestimated when not including quality improvement program implementation costs and when not considering changes to patient quality of life.

There are also limitations related to the availability of studies that have estimated the health care costs associated with the events targeted by the quality improvement measures. Even when studies are available that can be used to estimated health care costs, the rigor of the methods used to develop the estimate are highly variable.

For some events measured, reimbursement by CMS may be at risk, such as for hospital readmissions that meet specific criteria and for "never events." Even though a change in some of these potential non-reimbursable events may not impact health care costs from the perspective of CMS, health care costs related to changes for some of these events are included in the impact analysis. It is not clear which of these individual events would not be reimbursed based on the specific characteristics of the event, or based on whether these events would be reimbursed by a different payer.

Therefore, it is beyond the scope of this project to compile an exhaustive list of quantitative and qualitative factors that are related to patient-level and economic impacts. However, the estimates derived from this analysis provide a useful starting point for discourse related to the impact of performance changes on patients and health care costs.

The health care cost averted estimates were reported using the following rounding rules:

- If in billions, rounded to nearest $100 million.
- If in millions, rounded to nearest $100,000.
- Cost per event, no rounding.

10.4. ADJUSTMENT OF AFFORDABLE CARE PAYMENT MEASURES

Methods described above for data acquisition, data validation, use of external data, and analysis for trends and disparities were applied to these measures. Payment measures were converted into 2015 dollars, using the medical care services index (MCS) reported by the Bureau of Labor Statistics, similar to the health care cost conversion in Research Question 4.[7]

11. RESEARCH QUESTION 4

For measures implemented in CMS programs, did disparities among identified subpopulation groups exist, and did the disparities change over time?

A disparity analysis was conducted that compared the measure rates for different subgroups of patients, stratified by sex, age, race/ethnicity, income, region, urban-rural status (urbanicity). Disparity analyses were conducted on measures where beneficiary-level data was available. The disparity analysis examined both the magnitude of the disparity as well as the trend over time.

11.1. POPULATION SUBGROUPS

The characteristics of interest for the disparity analysis include sex, age, race, ethnicity, income, region, and urbanicity (urban-rural status). For each disparity analysis, a reference category was defined, and all other comparison categories were compared with the reference category to determine whether measure rate differences existed across subgroups. Table D-2 provides a list of the variables that were used in the disparity analysis, their data sources, categories, and reference groups.

Table D-2: Variables for Disparity Analysis

Variable	Source	Category*	Reference Group
Sex	Varies, depending on measure	Male, Female	Male
Age	Varies, depending on measure	18–64, 65–74, 75–84, 85+	65–74
Race/ Ethnicity	Varies, depending on measure	Varied by source – 1997 OMB definitions preferred; Unknown and Other categories were excluded	For race: Whites For ethnicity: Non-Hispanics
Region	Region defined by U.S. Census Bureau areas based on the state and ZIP code of the beneficiary, which varies, depending on the source of the beneficiary's location used for the measure. Please see the following link for category descriptions: https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf	New England, Middle Atlantic, Southern Atlantic, East North Central, East South Central, West North Central, West South Central, Mountain, Pacific	South Atlantic – chosen as an arbitrary reference point with the largest population

Variable	Source	Category*	Reference Group
Urbanicity	National Center for Health Statistics urban-rural scheme (2014)[8] based on the state and ZIP code of the beneficiary, which varies, depending on the source of the beneficiary's location used for the measure. Please see the following link for category descriptions: https://www.cdc.gov/nchs/data_access/urban_rural.htm	Large Central Metro, Large Fringe Metro, Medium Metro, Small Metro, Micropolitan, Non-Core	Large Central Metro – chosen as an arbitrary reference point
Income	U.S. Census Bureau 2010 estimates of median household income translated to ZIP Code Tabulation Areas (ZCTAs) based on the ZIP code of the beneficiary, which varies, depending on the source of the beneficiary's location used for the measure	Quartiles of household income: high (greater than $68,870), medium-high (greater than $52,136 to $68,870), medium-low (greater than $41,129 to $52,136), low ($41,129 or less)	High (wealthiest i.e., 4th quartile)

*The categories listed in this table may differ from those defined in the measure specifications collected by the data owners. For example, *Percentage of patients aged 50 years and older receiving a screening colonoscopy without biopsy* (NQF #0658) is a measure specified for beneficiaries age 50 and older. For this reason, it was not possible to examine the 18–64 age group in the analysis. In these situations, the methodology was adjusted as necessary to assess the appropriate stratifications for each variable included in the disparity analysis. In the case of NQF #0658, the 50–64 age group was compared with the 65–74 reference age group.

11.2. DISPARITY SIZE

The size of the disparity was examined by comparing the rates for the reference group to that of the comparison group. Two criteria were used to determine whether the observed differences in quality measure rates were sufficient to be defined as a disparity. First, the difference between the quality measure rates for the two groups must be statistically significant with $p < 0.05$ on a two-tailed test using the most recent annual rates for the study period. Second, the relative difference between the comparison group and the reference group must have an absolute value of at least 10% using Equation 4, where p1 is the comparison group rate, and p2 is the reference group rate:

$$\left|\frac{(p1-p2)}{p2}\right| \geq 0.1 \text{ where p1} < 0.5, \text{ or } \left|\frac{((1-p1)-(1-p2))}{(1-p2)}\right| \geq 0.1 \quad \text{where p1} >= 0.5. \qquad (4)$$

Not all measure results are proportions. The computation of the relative difference between results r1 and r2 for the most recent annual rate is virtually identical to the first part of Equation 4, and the difference also must have an absolute value of at least 10%: $\left|\frac{(r1-r2)}{r2}\right| \geq 0.1$.

The statistical difference between measure rates was tested using a z-test for proportions if the underlying measure is based on a proportion, and a t-test if the underlying measure is based on an average.

The z-test for the difference of proportions is given by the following formula:

$$z = \frac{p_1 - p_2}{\sqrt{\frac{p(1-p)}{n_1} + \frac{p(1-p)}{n_2}}} \qquad \text{where} \qquad p = \frac{(p_1 * n_1) + (p_2 * n_2)}{n_1 + n_2} \tag{5}$$

In Equation 5, p_1 is the proportion for the measure rate in the reference group, p_2 is the proportion for the measure rate in the comparison group, and n_1 and n_2 are the sample sizes for the reference and comparison groups, respectively.

For measures represented as medians, a Wilcoxon rank-sum test was used to determine whether there was a significant difference between groups regarding the overall ranking of the distributions.

11.3. DISPARITY CHANGE

To identify the difference in trends over time across two groups, the analysis included a test for the equivalence of regression coefficients derived for each of the subgroups in the comparison. Drawing from the trend regressions defined in the trending analysis, the following formula was used to calculate the z-score for the test of the equality of coefficients across independent samples:

$$z = \frac{b_1 - b_2}{\sqrt{SEb_1^2 + SEb_2^2}} \tag{6}$$

Where b_1 and b_2 are the trend coefficients from the trending analysis for group 1 and group 2, respectively, and SEb_1^2 and SEb_2^2 are the squared standard errors for the trend coefficients for group 1 and group 2, respectively. A significant change in disparities is defined as a p-value < 0.10 under a two-tailed z-test and a difference of 1% or greater in the trends.

To develop a more complete picture of the disparity trends over time, the analysis also examined the trend in the ratio between each comparison group and the reference group over time. The ratio was formed as:

$$Ratio = \frac{Rate_{comparison\ group}}{Rate_{reference\ group}} \tag{7}$$

The ratio of the rates between the two groups represents the relative difference in magnitude of the rates. To evaluate the trend in the ratio of the two groups, a linear regression was fit to the ratio of annual rates calculated in Equation 7. The regression takes the form:

$$Ratio_t = \beta_0 + \beta_1(Time_t) + \varepsilon_t \tag{8}$$

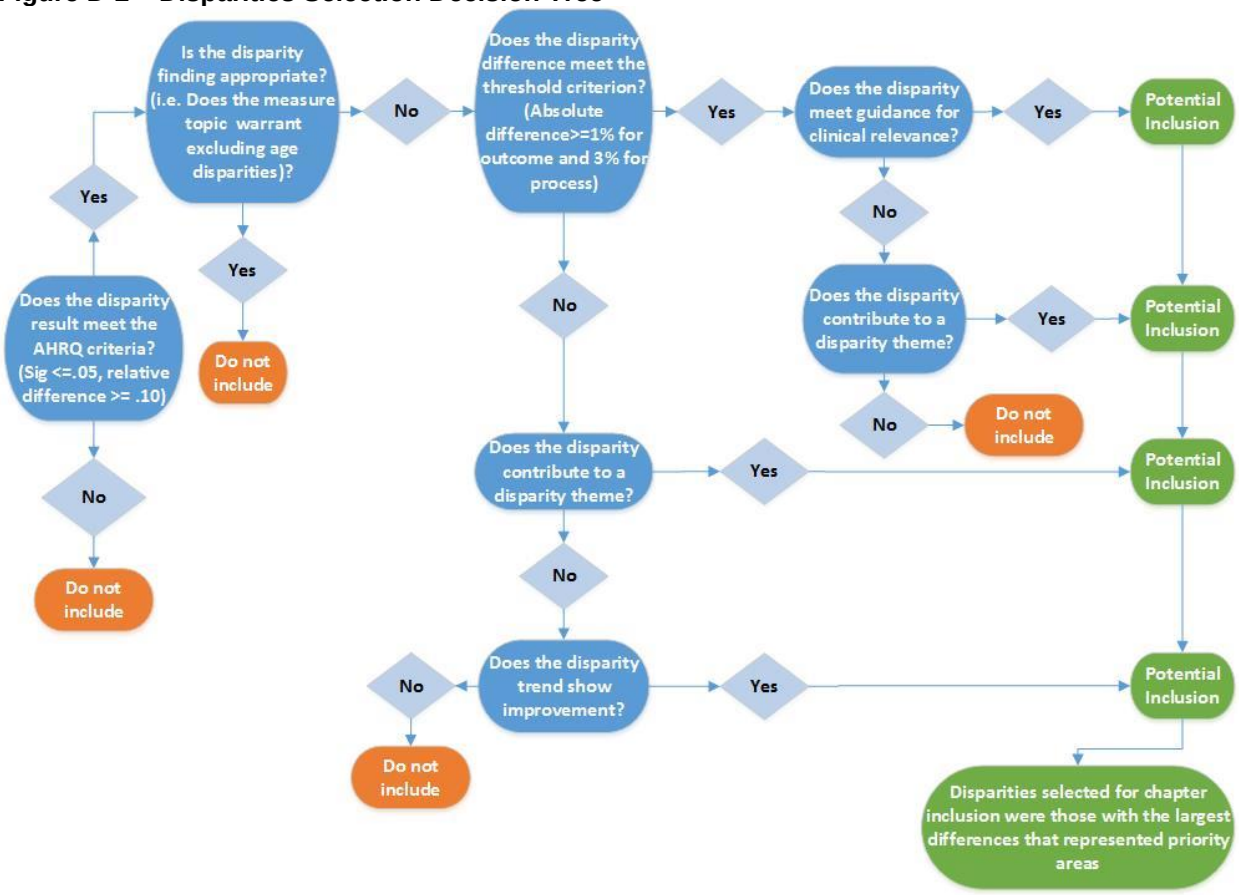
If the coefficient for the time trend, β_1, is significant with a *p*-value < 0.10, then the trend in the ratio was interpreted to be increasing if β_1 is positive and decreasing if β_1 is negative. The coefficient β_1 in Equation 8 was interpreted as the average annual proportional increase or decrease in the disparity ratio. For example, if β_1 equaled 0.20 ($p < 0.10$) then the comparison group rate would be growing at an average rate of 20% per year relative to the reference group rate.

Combining the regression coefficient for $Time_t$ from Equation 8 with the difference in the rates at the last observed data point indicates whether the ratio is growing or shrinking in size.

11.4. DETERMINING RELEVANT DISPARITY FINDINGS

Measures with observed disparities were assessed for chapter inclusion based on a set of parameters that focused on topped out status, clinical relevance, and appropriateness of the disparity. Disparity trending analysis results were included for measures where a point in time disparity was observed based on the most recent reporting period (2015). Three factors were considered when determining whether a disparity observed was clinically meaningful. These included the potential for improving the existing disparity; the topic importance to the patient and provider, such as mortality disparity findings; and the size of the affected population. The disparity category was also assessed for appropriateness. For example, age was observed consistently as a disparity for many of the measures, but the differences observed are primarily the result of physiological factors associated with age and likely not attributable to an inequity in care, treatment, or access to resources. The decision tree in Figure D-2 offers an overview of the decision-making process.

Figure D-2 – Disparities Selection Decision Tree



12. RESEARCH QUESTION 5

What is the environmental context that may affect interpretation of the impact of measures used in CMS programs?

As depicted in the Figure D-1, program factors as well as patient, provider, and environmental factors may affect achievement of goals for quality health care and health outcomes. This impact assessment cannot empirically isolate the effect of performance measurement from the other factors. CMS Medicare programs have been implemented nationally and concomitantly with numerous other public and private quality reporting/value-based payment initiatives, therefore, this impact assessment cannot attribute improvements solely to the CMS programs. However, this report includes an overview of the environmental context in which the measures are implemented to aid in interpretation of the results. Specifically, the following items are addressed in each quality priority:

- Other HHS initiatives, related programs, payment models and legislation that may influence quality and may impact the effectiveness of the quality improvement strategies.
- Published findings for similar measures to Key Indicators from literature or other HHS data (e.g., AHRQ or other HHS agencies).

12.1. OTHER INITIATIVES THAT INFLUENCE QUALITY

Performance measurement is one strategy to improve quality of care, patient outcomes, and value-based performance efforts. Other types of activities and initiatives such as legislative mandates, quality improvement initiatives, and public health initiatives could also influence performance measure results. Key federal initiatives that could influence performance measure results for the 2018 Impact Assessment were identified to facilitate interpretation of results for the measures used in CMS Medicare programs.

A scan was conducted to identify other key initiatives besides measurement that can influence performance measure results. Websites of CMS and other HHS agencies were searched to identify initiatives related to the CMS quality priorities.[4] The following information was identified for the key federal initiatives:

- Description of the initiative
- Authorizing legislation if applicable
- Implementation plans and program design described in final rules posted in the *Federal Register* or on the CMS website

12.2. PUBLISHED FINDINGS FOR MEASURES SIMILAR TO THE KEY INDICATORS

Published literature and external HHS data were used to discuss the findings on the Key Indicators in the context of other published results for similar measures.

A similar measure is a measure based conceptually on the same process or outcome as in the Key Indicator measure (e.g., vaccination, and readmission). It could have the same patient population or a broader patient population that overlaps with the population in the Key Indicator. The data source may be different (e.g., claims vs. chart abstraction) or may include different or overlapping settings (e.g., hospitals vs. home health agencies) from that of the Key Indicator. Measure descriptions, not detailed measure specifications, were used to determine similarity. Targeted external data sources for identifying similar measures include:

- AHRQ 2016 National Healthcare Quality and Disparities Report https://www.ahrq.gov/research/findings/nhqrdr/nhqdr16/index.html?utm_source=nhqrnet&utm_medium=referral&utm_campaign=nettochart
- Healthy People 2020 Leading Health Indicators http://www.healthypeople.gov/2020/Leading-Health-Indicators
- CDC Data & Statistics http://www.cdc.gov/DataStatistics/
- Medicaid https://www.medicaid.gov/medicaid/quality-of-care/index.html
- Medical Expenditure Panel Survey https://meps.ahrq.gov/mepsweb/data_stats/data_overview.jsp

Targeted literature searches were conducted for Key Indicators. Literature related to health care conditions and events were identified using key search terms that align with the quality measure, such as mortality for a condition.

13. RESEARCH QUESTION 6

What additional measures are needed to assess progress toward achievement of CMS quality priorities (i.e., meaningful measure gaps)?

One goal of performance measurement is to facilitate achievement of desired patient outcomes; another goal is to measure progress on achieving the stated objectives of CMS quality priorities.[4] A search was made for what additional measures, if any, could facilitate and gauge performance on aspects of CMS quality priorities and meaningful measurement topics. The CMS quality priorities are patient-centered and thus focus the measure gap assessment on patients.

TEP and FASC input during the selection of Key Indicators (Research Question 1) also helped identify meaningful measure topics for national improvement. The TEP and FASC input was synthesized from the panel review of measures for selection of Key Indicators.

Potential overlap or synergy with other groups that address performance measure gaps was explored, including:

- National Quality Forum (NQF)
- Measure Applications Partnership (MAP)
- Health Care Payment Learning and Action Network (LAN) on health care payment reform

14. RESEARCH QUESTION 7

What changes are providers making in response to use of performance measures by CMS?

The overarching research question was translated into five research questions that formed the content addressed by the hospital and nursing home surveys and interview guides:

- What types of changes or innovations have hospitals/nursing homes made to improve their performance on CMS measures?
- If a change or innovation was made, has it helped the hospital/nursing home improve its performance on one or more CMS measures?
- What challenges or barriers do hospitals/nursing homes face in reporting CMS quality measures?
- What challenges or barriers do hospitals/nursing homes face in improving performance on CMS quality measures?
- What unintended consequences do hospitals/nursing homes report associated with implementation of CMS quality measures?

This section provides additional details regarding the methodology for sampling design, development of survey instruments and interview guides, fielding procedures, and analytic methods for the national provider surveys and interviews.[v]

14.1. DATA SOURCES

14.1.1. Data Sources for Hospitals

The Hospital Compare downloadable database from December 2015 was used as the primary data source for identifying the universe of eligible hospitals because it includes all participants in the key quality measurement programs for hospitals: the Hospital Inpatient Quality Reporting (IQR) and Hospital Outpatient Quality Reporting (OQR) programs. The hospital information file from the database includes general contact information (e.g., name, address, phone number, and critical access hospital status), but lacks contact information for key staff. The Medicare Providers of Services File for October 2015 (the most recent available) was used to obtain the Medicare-certified bed count and other hospital characteristics (e.g., teaching status).

To obtain quality performance data for the hospital, consideration was given to using the widely used and validated composite summary score (Total Performance Score) from the FY2016 Hospital Value-Based Purchasing (VBP) Program.[9] The Total Performance Score represents the weighted average of scores for the component quality domains (clinical processes, outcomes [including patient safety outcomes], experience, and efficiency), which are also composites of several measure scores. A decision was made against directly using Total Performance Scores from the VBP Program because scores were lacking for some hospitals that report significant numbers of process and outcome measures and are of interest for the survey.

Instead, a composite performance score was designed that was strongly correlated with the FY2016 Hospital VBP score but assigned scores to additional hospitals. First, the hospitals' composite score was constructed, using the same individual measures included in the Hospital VBP Program, and calculating the same four domains prior to determining the overall performance score. Second, the domain scores were calculated by taking the mean of the available measures scores within that domain. Finally, when creating the final composite, weights were assigned to the domain scores:

[v] The Office of Management and Budget (OMB) reviewed and approved the study design, survey instruments, fielding procedures, and analytic methods; OMB assigned the following control numbers: 0938-1290 (hospital survey) and 0938-1291 (nursing home survey).

10% for clinical processes, 40% for outcomes, and 25% each for patient satisfaction and efficiency. The following changes also were made to better meet the analytic needs of the survey:

1) Instead of the VBP program's requirement of a minimum of four measures in two domains, a less stringent threshold was applied, requiring at least two of the constituent measures (assuming at least two in the domain) in three domains.
2) The improvement points incorporated in Hospital VBP were not used in this analysis because the survey assessed hospital performance and actions at the time of survey administration.
3) The entire performance distribution was used to better differentiate high and low performers instead of the Hospital VBP program's practice of assigning hospitals below the median a score of zero.
4) A data-driven standardization procedure was performed prior to calculating the hospital domain scores. Traditional standardization for the measure (subtraction of the mean, followed by division by the standard deviation) was not used because skewness in the distribution of measure scores would result in outlying scores having excessive influence. Instead, standardizing for each measure was made on the basis of the empirical distribution function[10] and conversion of the actual performance distribution into one that is bell-shaped and centered at zero with a standard deviation of 1. Specifically, each observed measure score was transformed into its percentile within the full distribution of observed scores for that measure. Next, the transformed scores were further converted using the inverse of the standard normal cumulative distribution function.

It was found that the performance metric was highly correlated ($\rho = 0.83$) with the Hospital VBP score, but was applicable to an additional 157 hospitals (of the 295 hospitals lacking a Hospital VBP score but otherwise meeting inclusion criteria). Of the 138 hospitals that could not be assigned a performance score, 41 were assigned scores on exactly two domains, 47 were assigned scores on exactly one domain, and 50 received scores on zero domains. Further, of the hospitals that received no domain scores, 34 lacked scores for any of the constituent measures.

In addition to constructing the data above for sampling purposes, facility-level demographic, socioeconomic, and risk-adjustment data were added for use in descriptive analyses using a three-step process. First, all Medicare fee-for-service beneficiaries linked to each hospital in 2015 were identified, using Medicare Part A claims from the same year. Then patient characteristics from the Medicare Master Beneficiary Summary File (e.g., age, sex, race, dual eligibility for Medicare and Medicaid, and end-stage renal disease) and hierarchical condition category (HCC) files (e.g., the HCC summary [community] score) were added. Finally, the means of these characteristics were computed and assigned to the corresponding hospital.

14.1.2. Data Sources for Nursing Homes

The Nursing Home Compare provider file was used as the primary data source for identifying the universe of eligible nursing homes because it included (as of December 2015) all nursing homes participating in the primary measurement program of interest, the Nursing Home Quality Initiative (NHQI). This file contains contact information (e.g., name, address, and phone number) as well as data for computing size (bed count) and performance (5-star quality rating), which was used in implementing the stratified sampling design described below. The quality rating was used instead of the overall nursing home rating because the latter incorporates performance on additional regulatory measures that are not addressed by CMS quality measures.

After completing data construction for sampling, facility-level demographic, socioeconomic, and risk-adjustment data were added for use in descriptive analyses, using a three-step process. First, all Medicare fee-for-service beneficiaries linked to each nursing home in 2015 were identified, using Minimum Data Set (MDS) data from the same year. Then patient characteristics from the Medicare Master Beneficiary Summary File (e.g., age, sex, race, dual-eligibility for Medicare and Medicaid, and end-stage renal disease) and hierarchical condition category (HCC) files (e.g., the HCC summary [community] score) were added. Finally, the means of these characteristics were computed and assigned to the corresponding nursing home.

14.2. RESPONDENT UNIVERSE AND SAMPLING METHODS

A stratified sampling design was used in both settings because the actions that hospitals or nursing homes take in response to quality measurement programs are likely to vary based on size (bed count, which is a proxy for greater resources and potentially more specialization among quality leadership) and because determining which features differentiate high and low performers was of interest. Within each setting (hospital or nursing home), stratification on bed count and performance (a composite of Hospital Inpatient Quality Reporting [IQR] measures in hospitals or the star rating in nursing homes) served to facilitate stratified analyses of hospitals or nursing homes and/or compare providers across strata.

To increase the likelihood that estimates can be calculated with adequate precision and that comparisons of responses of hospital/nursing homes in different strata can be performed with adequate power, certain strata were oversampled, i.e., sampled from strata that are less prevalent across the population at a rate that was higher than the sampling rate for strata that are more prevalent.

14.2.1. Sampling for Hospitals

The sample of hospitals for the qualitative interviews and standardized survey was drawn from a universe of 4,805 hospitals participating in the IQR and OQR programs, as described above. After excluding hospitals whose relationship to CMS quality measurement programs differs considerably from that of Inpatient Prospective Payment System (IPPS) hospitals participating in the IQR program, 3,336 acute care hospitals were retained: critical access hospitals (n = 1,264), acute care hospitals in Puerto Rico or other U.S. territories (n = 54), children's hospitals (n = 22), and Veterans hospitals (n = 129). After removal of a further 138 hospitals that lacked sufficient data to evaluate performance, the final sampling frame consisted of 3,198 hospitals.

14.2.1.1. Sampling Design for Hospital Survey

A sample of 2,045 hospitals was randomly drawn from this universe with the goal of achieving 900 responses. The final hospital sampling frame was stratified into nine strata defined by bed size (three categories: large: 300 or more beds; medium: 101–299 beds; and small: < 100 beds) and quality (three categories: top 20%, middle 60%, and bottom 20%, as shown in Table D-3. Based on previous surveys of healthcare quality leaders, a response rate of 44%[11-15] was anticipated and therefore sampled a total of 2,045 facilities were sampled.

To increase power to detect differences between strata, near-census sampling within top and bottom performance strata was used. This approach was intended to provide sufficient power to generate overall estimates and estimates within size and quality strata. Drawing a simple random

sample of 2,045 hospitals (with the same target of 900 respondents) was considered, but using this strategy would have led to a 22% increase in the standard error of estimates calculated across the subpopulation of high-performing hospitals over that which is yielded by the strategy described above.

Table D-3: Hospital Structured Survey Universe and Number Sampled, by Strata

Quality Rating	Small (1–100 beds)	Medium (101–300 beds)	Large (>300 beds)
Top quintile of performance*	405 (198)	190 (188)	45 (45)
Middle 2nd–4th quintiles of performance distribution*	490 (199)	847 (343)	581 (235)
Bottom quintile of performance*	111 (111)	284 (281)	245 (245)

* Facility size derived from December 2015 Medicare Providers of Services file. Composite adapted from Hospital VBP Total Performance Score and constructed using CMS Hospital Compare quality measures released in December 2015.

14.2.1.2. Sampling Design for Qualitative Interviews of Hospitals

Potential hospital quality leaders from the nine strata for the qualitative interviews also were selected. A random stratified sample for recruitment and scheduling was chosen in a sufficient size for achieving a target of four to five completed interviews per stratum. The final distribution of completed interviews by stratum is shown in Table D-4.

Table D-4: Completed Hospital Qualitative Interviews, by Strata

Quality Rating	Bed Size		
	Small (1–100 beds)	Medium (101–300 beds)	Large (300+ beds)
Top Quintile	5	3	4
Middle Quintiles	4	5	4
Bottom Quintile	5	3	7

*Facility size derived from December 2015 Medicare Providers of Services file
Composite adapted from Hospital VBP Total Performance Score and constructed using CMS Hospital Compare quality measures released in December 2015.

14.2.1.3. Sampling for Nursing Homes

The sample of nursing homes for the qualitative interviews and standardized survey was drawn from the 15,661 nursing homes participating in the NHQI. After removing 180 nursing homes that lacked a quality score and excluding nursing homes located in Puerto Rico or other U.S. territories (n = 6) because their relationship to CMS quality measurement programs differs considerably from that of most nursing homes participating in NHQI, the final sampling frame included 15,475 nursing homes.

14.2.1.4. Sampling Design for Standardized Survey in Nursing Homes

A sample of 2,045 nursing homes was randomly drawn from this universe with the goal of achieving 900 responses across nine strata, defined using the performance categories (low, medium, and high) and bed size categories (small, medium, and large) discussed above, under the assumption of a 44% response rate. With the aim of achieving 100 completed responses in each of the nine strata, 227 or 228 respondents per stratum were sampled.

The strategy of equal allocation of sample across the nine strata was preferable because it provided sufficient power to generate overall estimates and estimates within size and quality

performance strata. Drawing a simple random sample of 2,045 nursing homes (with the same target of 900 respondents) was considered, but it was projected that using this strategy would have led to reduced precision of estimates within the less populated strata over that which is yielded by the strategy described above.

Of note, response rates for staff from health care facilities can vary, with response rates as low as 20% and as high as 60% being reported.[11-14] Based on recruitment efforts in support of survey development, a response rate closer to 40% was projected. Therefore, the surveyors started with the sample described above (2,043 nursing homes) and used multiple modes of outreach to achieve a response rate of 44% (900 responses). Table D-5 shows the distribution of the sampled (and responding) nursing homes across the nine strata that result from this sampling strategy.

Table D-5: Nursing Home Structured Survey Universe and Number Sampled, by Strata

Quality Rating*	Bed Size		
	Small (1-74 beds) (N, N Sampled)	Medium (75-149 beds) (N, N sampled)	Large (150+ beds) (N, N sampled)
5 stars	1,668 (228)	2,388 (228)	746 (228)
3 or 4 stars	1,738 (227)	3,147 (228)	1,098 (228)
1 or 2 stars	1,435 (228)	2,492 (228)	763 (228)

*Quality rating obtained from CMS Nursing Home Compare quality rating, released December 2015.

14.2.1.5. Sampling Design for Qualitative Interviews of Nursing Homes

Potential nursing home quality leaders from the nine strata were selected for the qualitative interviews. A random stratified sample for recruitment and scheduling was selected that was of sufficient size for achieving a target of four to five completed interviews per stratum. The final distribution of completed interviews by stratum is shown in Table D-6.

Table D-6: Completed Nursing Home Qualitative Interviews, by Strata (n = 40 total)

Quality Rating*	Small (1–74 beds)	Medium (75–149 beds)	Large (>149 beds)
5-star quality rating*	4	6	4
3- or 4-star quality rating*	4	5	4
1- or 2-star quality rating*	4	4	5

*Based on CMS Nursing Home Compare quality measure star rating, December 2015.

14.3. PROCEDURES FOR COLLECTING INFORMATION

14.3.1. Data Collection Instruments

The survey instruments and interview guides for both hospitals and nursing homes (*Hospital Interview Guide and Provider Survey* [Appendix I] and *Nursing Home Interview Guide and Provider Survey* [Appendix J]) addressed topics that include innovations in care delivery that facilities were making to improve performance, whether facilities perceived the innovations they reported making to have been helpful in improving quality, challenges in reporting measures, difficulties improving performance, unintended consequences of CMS quality measures, supportiveness of executive and clinical leaders, facility characteristics, and respondent background information. The survey and interviews in the hospital setting focused on experiences with the IQR and OQR measures; those in the nursing home setting, on experiences with the NHQI measures.

14.3.2. Identification of Appropriate Respondents

To ensure that survey and interview respondents were comparable across facilities, a strategy was developed during formative interviews and cognitive testing for identifying a quality leader within each organization. The quality leader is defined as the individual who is most familiar with the CMS performance measures and leads quality improvement activities that the organization has undertaken in response to these measures. The senior quality leaders identified often had titles such as Chief Medical Officer, Chief Quality Officer, Vice President for Quality, or Quality Director in the hospital setting, while in the nursing home setting, titles included the Director of Nursing or Director of Quality. However, using a specific title to identify the quality leader was purposefully avoided because the exact title may vary between facilities. During the formative interviews and cognitive testing, these individuals demonstrated that they possessed the knowledge necessary to address the survey questions. No gifts or incentives were given to respondents for their participation in the survey or interviews.

14.3.3. Fielding the Standardized Surveys

When fielding the standardized surveys of hospital and nursing home quality leaders, several methods were used to reduce the risk of low (20% or lower) response rates that have been seen in recent published surveys of hospital leaders.[11-14] As recommended as best practice by Dillman, et al.,[15] non-responders were contacted, using varying modes, including modes different from the initial contact modes, according to the schedule below:

- Weeks 1–3 – Initial and follow-up email invitations to complete the survey by Web. These invitations were sent via email one week apart to all sampled facilities and contained sufficient information for informed consent as well as a personalized code that allowed access to the Web survey for that hospital.
- Week 4 – Mail survey to all non-responding quality leaders as well as quality leaders without email addresses.
- Week 7 – Phone calls to non-responding quality leaders were made to prompt return of the mail survey or completion of the Web survey. Non-responders were initially contacted by email or mail; phone outreach was reserved until later in the data collection period, when there were fewer non-responders.
- Week 16 – Facilities were sent a final email reminder to complete the Web survey.

The nursing home structured survey was fielded between June 1, 2016, and August 26, 2016. The hospital survey was fielded between June 1, 2016, and January 6, 2017. Throughout data collection, response rates and cooperation were tracked within each sample stratum, and additional efforts and added samples were employed to achieve sufficient response in each stratum.

14.3.4. Conducting Qualitative Interviews

Using information obtained by directly contacting facilities (i.e., the name, job title, mailing address, email address, and telephone extension of the facility quality leader), a letter via email was sent to the facility quality leader, which described the study and interview and invited the quality leader or a designee to take part in the interview. As needed, the letter was followed up by phone three to five days later to answer questions and to schedule an interview.

Several methods were used to maximize response rates and minimize non-response bias. First, up to 10 attempts were made, both by phone and via email, to schedule the interview. Second, the interview was scheduled for a day and time that was most convenient for the quality leader. Third, interview appointments were scheduled through each quality leader's administrative assistant. Finally, the quality leader could ask another person within the facility to participate. Facilities that refused to participate in the interview or failed to respond to the invitations were replaced with another facility with similar characteristics.

All 40 nursing home and 40 hospital qualitative interviews were conducted between April 1, 2016, and August 12, 2016. Interviews were conducted by phone and were one hour in duration. The qualitative interview guides covered similar topics as the structured survey. Interview respondents were given lists of measures displayed on Hospital Compare and Nursing Home Compare (as of April 2016) prior to the interview for reference.

14.3.5. Confidentiality and Data Safeguarding

All persons who participated in this data collection, either through the qualitative interviews or the standardized surveys, were assured that the information they provided would be kept private to the fullest extent allowed by law. Both qualitative interview guides include an informed consent and confidentiality script that was read to the respondent(s) before any interview. Quality leaders who participated in the standardized survey received consent and confidentiality information via the emails and letters inviting them to participate in the Web and mail surveys, which provided information on the nature of the research being conducted and their rights as survey or interview respondents. Respondents who had questions about the consent statement or other aspects of the study were instructed to call the contractor's staff and/or the administrator of its institutional review board (IRB).

The study had a data safeguarding plan to further ensure the privacy of the information collected. For the online survey and qualitative interviews, an anonymized data identifier was assigned to each respondent. The standardized survey was collected via an experienced vendor. All electronic files directly related to the administration of the survey were stored on a restricted drive of the vendor's or the contractor's secure local area network, and access was limited to employees working on the specific project. Additionally, no single file contained quality leaders' response data and contact information.

For the qualitative interviews, contact information that could be used to link individuals with their responses was removed from all interview instruments and notes. Files containing contact information used to conduct qualitative interviews were also stored on a centralized server with access limited to key personnel, following procedures reviewed and approved by the contractor's IRB.

14.4. ANALYTIC METHODS FOR STRUCTURED SURVEYS

The analytic approach was similar for both the nursing home and hospital surveys. Initial preparatory work included examining the distribution of answers to 21 items applicable to all respondents; using this distribution, the complete respondents were defined as those facilities that completed at least seven items. Item response scales varied between items (e.g., 2-Likert, Yes/No; or 4-Likert, Never/Sometimes/Usually/Always) and were standardized either through linear mean or top-box scoring. For linear mean scoring, possible response options were

rescaled from 0 (worst) to best (100) with equal intervals between options. For top-box scoring, the one or two most positive response options were recoded to 100 and all other response options were recoded as 0.

Sampling weights were then applied to ensure that the cohort of respondents was made representative of the entire population of nursing homes or hospitals. The sample weights are the product of design weights (which account for the study design) and non-response weights (which account for the differential rates of response across a variety of facility characteristics). Nonresponse weights were developed using logistic regression to ensure that the weighted sample resembled the facility population on characteristics such as size, performance, geographical region, patient characteristics, etc.[16] The final sampling weight represents the inverse of that facility's probability of inclusion in the set of respondents.

Weighted means and standard deviations were produced for each survey item using the sampling weights described above. Weighted means and standard deviations were also calculated across various strata. The types of subgroups that were of interest included performance strata (low, medium, high) and hospital size (e.g., number of beds). F-tests were calculated to test for significant differences in the equality of the means of the three groups within each comparison.

14.5. ANALYTIC METHODS FOR QUALITATIVE INTERVIEWS

All interviews were audio recorded, with the exception of two interviews for which the respondent declined audio recording. For all interviews, including the two for which the respondents declined to allow recording, a second surveyor listened to the call and took detailed notes. Interview responses were reviewed to extract specific examples for each topic of interest, and grouped examples into key themes. The topic areas included changes made in response to CMS measurement programs, unintended consequences, barriers to improving scores on the measures, and barriers to reporting the measures. The range of responses for each key theme was described and reported by setting (hospital or nursing home) and then compared to key findings from the structured survey in each setting.

References

1. The Patient Protection and Affordable Care Act, Pub. L. No. 111–148 §3014(b). 124 Stat. 119, 385-387, (2010) (as amended at 42 U.S.C. § 1395aaa-1).
2. Centers for Medicare & Medicaid Services. *National Impact Assessment of Medicare Quality Measures.* Baltimore, MD: US Department of Health and Human Services; 2012. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityMeasures/National-Impact-Assessment-of-the-Centers-for-Medicare-and-Medicaid-Services-CMS-Quality-Measures-Reports.html. Accessed December 10, 2017.
3. Centers for Medicare & Medicaid Services. *2015 National Impact Assessment of the Centers for Medicare & Medicaid Services (CMS) Quality Measures Report.* Baltimore, MD: US Department of Health and Human Services; 2015. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityMeasures/QualityMeasurementImpactReports.html. Accessed December 10, 2017.
4. Centers for Medicare & Medicaid Services. *CMS Quality Strategy 2016.* Baltimore, MD: US Department of Health and Human Services; 2015.

http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/Downloads/CMS-Quality-Strategy.pdf. Accessed December 11, 2017.

5. Centers for Medicare & Medicaid Services. *HHS Decision Rules for Categorizing Measures of Health, Health Care Quality, and Health Care Affordability.* Baltimore, MD: US Department of Health and Human Services; 2014. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityMeasures/Downloads/HHS-Decision-Rules-for-Categorizing-Measures-of-Health-Health-Care-Quality-and-Health-Care-Affordability.pdf. Accessed December 11, 2017.
6. Kiefe CI, Weissman NW, Allison JJ, Farmer R, Weaver M, Williams OD. Identifying achievable benchmarks of care: concepts and methodology. *Int J Qual Health Care.* 1998;10(5):443-447.
7. US Bureau of Labor Statistics. Databases, tables, & calculators by subject. US Department of Labor; 2017. https://data.bls.gov/timeseries/CUUR0000SAM?output_view=pct_12mths. Accessed December 10, 2017.
8. Ingram DD, Franco SJ. NCHS urban–rural classification scheme for counties. National Center for Health Statistics. *Vital Health Stat.* 2013;2(166).
9. Medicare Program; Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals and the Long-Term Care Hospital Prospective Payment System Policy Changes and Fiscal Year 2016 Rates; Revisions of Quality Reporting Requirements for Specific Providers, Including Changes Related to the Electronic Health Record Incentive Program; Extensions of the Medicare-Dependent, Small Rural Hospital Program and the Low-Volume Payment Adjustment for Hospitals; Final Rule. *Fed Regist*. 2015; 80(158):49325-49843.
10. Robbins MW. The utility of nonparametric transformations for imputation of survey data. *Journal of Official Statistics.* 2014;30(4):675-700.
11. Blendon RJ, Schoen C, DesRoches CM, Osborn R, Zapert K, Raleigh E. Confronting competing demands to improve quality: a five-country hospital survey. *Health Aff (Millwood).* 2004;23(3):119-135.
12. Weissman JS, Annas CL, Epstein AM, et al. Error reporting and disclosure systems: views from hospital leaders. *JAMA.* 2005;293(11):1359-1366. doi: 10.1001/jama.293.11.1359.
13. Cycyota C, DA H. What (not) to expect when surveying executives: a meta-analysis of top manager response rates and techniques over time. *Organizational Research Methods.* 2006;9(2):133-160. doi: 10.1177/1094428105280770.
14. Baruch B, Holton B. Survey response rate levels and trends in organizational research. *Human Relations.* 2008;61(8):1139-1160. doi: 10.1177/0018726708094863.
15. Dillman D, Smyth J, Chistian L. *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method.* Hoboken, NJ: Wiley & Sons; 2009.
16. Särndal C-E. The calibration approach in survey theory and practice. *Survey Methodology.* 2007;33(2):99-119.