NATIONAL QUALITY FORUM
Driving measurable health improvements together

Measure Worksheet

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.
To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 3667
Corresponding Measures: N/A
Measure Title: Days at Home for Patients with Complex, Chronic Conditions
Measure Steward: Centers for Medicare & Medicaid Services
sp.02. Brief Description of Measure: This is a provider group-level measure of days at home or in community settings (that is, not in acute care such as inpatient hospital or emergent care settings or post-acute settings such as Skilled Nursing Facilities (SNFs)) among adult (age 18 years or older) Medicare FFS beneficiaries with complex, chronic conditions who are aligned to participating provider groups. The measure includes risk adjustment for differences in patient mix across provider groups, with an adjustment based on patients' risk of death. An additional adjustment that accounts for patients' risk of transitioning to a long-term nursing home is also applied to encourage home- and community-based care in alignment with CMS's policy goals. A higher risk-adjusted score indicates better performance.
1b.01. Developer Rationale: This measure will directly benefit Medicare patients by reducing unnecessary hospitalizations and incentivizing days at home or in the community, which often reflect patient preferences. There are no competing measures for Medicare beneficiaries. The measure may help to incentivize care coordination between healthcare providers. The measure developers have worked to reduce the risk of the measure having negative unintended consequences on patients.

---

sp.12. Numerator Statement: The outcome measured for each eligible beneficiary is days spent "at home," adjusted for clinical and social risk factors, risk of death, and risk of transitioning to a long-term nursing home.
sp.14. Denominator Statement: Eligible beneficiaries aligned to participating provider groups.
sp.16. Denominator Exclusions: Not applicable. There are currently no denominator exclusions or exceptions for the measure. All patients meeting the denominator inclusion criteria are included.

Measure Type: Outcome
sp.28. Data Source: Claims
sp.07. Level of Analysis: Accountable Care Organization

IF Endorsement Maintenance – Original Endorsement Date:
Most Recent Endorsement Date

IF this measure is included in a composite, NQF Composite#/title:
IF this measure is paired/grouped, NQF#/title:
sp.03. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?

Criteria 1: Importance to Measure and Report

Evidence

1a. Evidence. The evidence requirements for a health outcome measure include providing empirical data that demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service; if these data not available, data demonstrating wide variation in performance, assuming the data are from a robust number of providers and results are not subject to systematic bias. For measures derived from patient report, evidence also should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.

Evidence Summary

- This is an outcome measure using claims data at the accountability care organization level that measures days at home or in community settings for Medicare FFS beneficiaries with complex, chronic conditions who are aligned to participating provider groups.
- The logic model that indicates that timely access to high quality preventive and primary care, consideration of patient preferences for care settings and improved care coordination and care transitions lead to more patient time spent at home and reduced overutilization of acute and long-term institutional care settings which lead to improved patient-centered primary care and quality of life and reduced healthcare costs for patients and healthcare systems.
- The developer conducted a literature review of relevant peer-reviewed publications that found that:
    - Most patients and families prefer spending time at home and in the community ("days at home") rather than in the hospital,1-6 and more days at home are associated with both positive clinical outcomes and lower costs for patients and providers.
    - Poor care coordination can lead to unnecessary and preventable hospital visits for patients; in contrast, improved care coordination and care transitions prevent unplanned hospital visits, leading to more days at home and high-quality timely care.
    - Given that patients with complex, chronic conditions often receive care from several clinicians and sites of care, this patient population may particularly benefit from improved care coordination.

Question for the Committee:

- Is there at least one thing that the provider can do to achieve a change in the measure results?

Guidance from the Evidence Algorithm

Measure assesses outcome (box 1) YES -> Relationship between outcome and at least one healthcare action (box 2) YES -> PASS

Preliminary rating for evidence: ☒ **Pass** ☐ **No Pass**

Gap in Care/Opportunity for Improvement and 1b. Disparities

Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- The developer reports data from 2017 – 2018 Medicare FFS claims data for 1,154,779 patients from the Shared Savings Program (SSP) Accountable Care Organizations (ACOs), which is comprised of 610

ACOs. Among the 610 ACOs, the average Adjusted Days at Home were 330.4 days (standard deviation 3.7 days), ranging from 291.0 to 345.9.

- The developer notes that due to the use of risk adjustment, this performance gap illuminates variation that can likely be attributed to ACO quality difference versus a sole reflection of differences in case mix between ACO.
- The developer references several studies that demonstrate substantial variations in time spent at home and suggests that there is opportunity to improve the quality of care and the resulting days at home for the target population.

Disparities

- Developer notes significant impact due to age and Medicaid dual-eligible status, which are included in the risk model to avoid disincentivizing care of patients at higher risk.
- Developer additionally noted statistically insignificant or comparatively minor effects for select indicators of social risk, including local density of primary care physicians or specialists, local socioeconomic status indicator, urban residence, local percent of residents unmarried or living alone, and local density of hospital and nursing home beds.
- Developer observes limitations in the connectivity of disparities data to the outcome. Based on existing studies, developer notes inconsistencies in evidence linking socioeconomic status with days at home for older patients, further sharing that some studies link indicators such as poverty, female sex, age, and/or dual-eligible status with fewer days at home while other studies find no significant difference based on age, sex, or race/ethnicity.

Questions for the Committee:

- Is there an observable gap in care that warrants a national performance measure?
- Is sufficient disparities information provided? Are you aware of evidence that disparities exist in this area of healthcare?

Preliminary rating for opportunity for improvement: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments:
Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus: For all measures (structure, process, outcome, patient-reported structure/process), empirical data are required. How does the evidence relate to the specific structure, process, or outcome being measured? Does it apply directly or is it tangential? How does the structure, process, or outcome relate to desired outcomes? For maintenance measures –are you aware of any new studies/information that changes the evidence base for this measure that has not been cited in the submission? For measures derived from a patient report: Measures derived from a patient report must demonstrate that the target population values the measured outcome, process, or structure.

- I am concerned about the "outcomes" being measured. I am sure that money saved is greater if patient is at home rather than a facility (objectively provable). I am sure that patients would subjectively prefer to be at home rather than at a facility (QoL questionnaires). I believe that care coordination by a PCP is enhanced but this is arbitrary and subjective. Are there hard data correlating patient health outcomes by risk adjustment score versus time at home? I would think that the measure result is tangential rather than directly related to the numerator.
- Developer presents evidence that majority of individuals/families prefer time at home/community versus in hospital or other care locations. They present evidence that there is a large variation in days at home across ACOs, suggesting that at the organizational level there are interventions or processes that can be fortified to keep people at home. Improving time at home would require evaluating, structural and outcome measures with a health system.

- There appears to be a moderate amount of medical evidence to suggest that this process measure is associated with improved quality.
- Pass
- The outcome measured for each eligible beneficiary is days spent "at home," adjusted for clinical and social risk factors, risk of death, and risk of transitioning to a long-term nursing home. The measure indicates that appropriate care transitions, case management, preventive and routine care, follow-up care will result in more days at home versus in a LTC setting. In most cases this would be an appropriate relationship of actions/care or service and outcomes being measured.
- An outcome measure with at least one healthcare action, so passes.
- I am concerned that the evidence presented does not fully support this measure, and because of specification and risk adjustment concerns, may actually diminish quality rather than improve it.
- Outcomes measure important to patients and families

1b. Performance Gap: Was current performance data on the measure provided? How does it demonstrate a gap in care (variability or overall, less than optimal performance) to warrant a national performance measure? Disparities: Was data on the measure by population subgroups provided? How does it demonstrate disparities in the care?

- It would appear that the gap in care (days in facility vs. days at home) is more provider group variable and risk adjustment variable than representative of a global suboptimal performance standard. Data was provided on subgroups, and it suggests that lower SE status, female gender, and dual program status were associated w/ less time at home and lower scoring on the measure.
- A performance gap exists for days at home ranging from 291 to 346 (mean 330, standard deviation 3.7 days). Risk adjustment is part of the measure, but developers note inconsistent impact of SES, poverty, age, or dual-eligible status on days in home.
- The developers present data demonstrating variability and do provide some evidence of disparities.
- The performance gap analysis utilized Medicare ACO data only and yet the measure is considered for use for all Medicare patients, not just ACOs.
- Although the measure has a risk-adjustment (three different risk adjustment models used) a performance gap may not be able to be identified. Study results had conflicting results regarding disparity impacts or disparities in care.
- Although the range for Average days at home is substantial, it is not clear that the variability among the ACOs provides a similar Performance Gap. Also, the patients are evaluated individually for the risk adjustment and not sure how that could be managed by the Accountable entity.
- Not really--I suppose that days at home could always be improved, but the question to me is about proper location--some people might be better off in SNF, nursing homes, even hospital, then home. The disparities data is a mess.
- The developer reports data from 2017 – 2018 Medicare FFS claims data for 1,154,779 patients from the Shared Savings Program (SSP) Accountable Care Organizations (ACOs), which is comprised of 610 ACOs. Among the 610 ACOs, the average Adjusted Days at Home were 330.4 days (standard deviation 3.7 days), ranging from 291.0 to 345.9.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

2b. Validity: Testing; Exclusions; Risk-Adjustment; Meaningful Differences; Comparability; Missing Data

Reliability

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented. For maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures.

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

Validity

2b2. Validity testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

2b2-2b6. Potential threats to validity should be assessed/addressed.

Complex measure evaluated by Scientific Methods Panel? ☒ **Yes** ☐ **No**

Evaluators: Patrick Romano, Sherri Kaplan, Daniel Deutscher, Joseph Hyder, John Bott, Bijan Borah, Jack Needleman, Jennifer Perloff, Susan White, Ronald Walters, David Nerenz, Sean O'Brien, Eric Weinhandl

Methods Panel Review (Combined)

Methods Panel Evaluation Summary:
This measure was reviewed by the Scientific Methods Panel and discussed on the call. A summary of the measure and the Panel discussion is provided below.

Specifications

- In their preliminary analyses, the subgroup members found the specifications confusing and occasionally arbitrary, with little evidence for the measure constructs. Members reported the potential misalignment of concept presentations within the submission and noted the denominator statement lacked an explanation of the target population, conditions, settings, and other pertinent measure constructs information. They were also concerned that several concepts included in the submission were not documented as exclusions in the specifications, which both threatens the measure's validity and may incentivize under-treatment of conditions potentially outside the locus of control of the accountable entity, including very low outliers that can never reach the expected performance gains, permanent nursing home residents
- The SMP also questioned whether the consideration of exclusions included (i.e., patients treated in emergency departments, admitted to acute care settings, and days after a death occurs), were always indicate low-quality care. Another SMP member expressed concerns with adjusting for transitions to the nursing home, which purports that moving from home to a nursing home, is always negative. Other concerning date elements included permanent nursing home admissions requiring skilled nursing care, which may include personal and community resources that are not be modifiable by the accountable entity.
- SMP members also noted that the unit of analysis reported in the measure vacillates between accountable care organizations (ACOs) and provider group.
- One SMP member questioned whether this measure, which combines multiple risk models calculations into a single overall score, should be considered a cost composite measure.

Reliability

- Reliability testing conducted at the Accountable Entity level:
    - The developer conducted a split half methodology with data from 2017-2018. They reported an ICC of 0.8326 for the final Days at Home outcome metric between the two samples. The form of ICC is not described.
    - A few SMP members stated split-half the use of split-half methodology is better suited for federal accountability programs with multiple years of data, particularly because patient assignment and ACO rules are modified annually. The SMP also noted that reporting an ICC with a split-half approach may under-estimate true reliability as scores estimates only use half of each provider's data

Validity

- Validity testing conducted at the Accountable Entity level:
    - The developer conducted construct validity with Pearson correlations to six other ACO-level measures hypothesizing that quality conceptually relates to excess days in care (EDIC) for patients with complex chronic diseases.
        - Pearson's correlations did not correlate well, ranging between -0.549 and +0.048 resulting in a *high inverse* correlation for unplanned admissions (expected), moderate correlation with other measures, no correlation with fall risk, and an unexpected inverse correlation with patient experience.
        - The developer explained that this is possibly due to endogeny of the hospital admissions and readmissions measures. The developer also reported the poor correlations may result from testing against measures using smaller sample sizes and which were not risk adjusted for clinical variables.
    - The developer also performed face validity testing of the computed measure score.
        - The TEP consisted of 19 of 21 responding members who assessed whether the "The Days at Home measure, as specified, can be used to distinguish between better or worse performance at ACOs or provider groups."
        - Two members indicated "strongly agree," 15 indicated "agree," and two indicated "somewhat agree."
- Risk-adjustment
    - Some SMP members noted that there are three different risk adjustment models used and expressed concerns about lack of clarity about whether/how they were combined to get a single score and the validity of the approach.
    - The SMP members had concerns with the model construction, which they agreed lacked vital adjustment and consideration for many variables without theoretical or empirical justifications and used arbitrary measure weighting, specifically the unexplained selection of weighting mortality days at 1.25 percent and the annual nursing home start date of January 1 that are not conceptually and empirically demonstrated or justified. The developers acknowledge these were not empirically assessed, but rather are subjective and based solely on TEP recommendation.
    - A few SMP members discussed the effect of specific chronic conditions on the risk model, such as cancer, dementia, and congestive heart failure that increase EDIC by nature of the disease states.

- The greatest concern for the risk adjustment model expressed from the SMP members was the development approach for days at home, and the mortality and nursing models. The SMP noted that faulty formulas in the approach may include doubling the EDIC estimates for enrolled ACOs and negative impacts to the penalty schematic

- Exclusions
  - The SMP questioned the process-outcome pathway that resulted in increased, rather than decreased, days in care, and the lack of exclusions for long-term nursing home residents prior to a measurement period, who have no chance of "at home" days defined in the specifications.
  - SMP members indicated the discrimination and calibration were generally acceptable but had concerns related to the low outliers. The developer described this as an unintended consequence of the measure construct as the measure attempts to balance days at home with other unintended consequences.
- Meaningful Differences
  - A few SMP members questioned the presence of meaningful differences in performance and the use of the measure for quality improvement purposes, and whether the measure could be used to identify differences in patient function or health-related quality of life.
- The SMP did not reach consensus on the validity criterion.

Questions for the Committee regarding specifications:

- Do you have any concerns about the measure specifications?

Questions for the Committee regarding reliability:

- Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?
- The Scientific Methods Panel is satisfied with the reliability testing for the measure. Does the Committee think there is a need to discuss and/or vote on reliability?

Questions for the Committee regarding validity:

- Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?

Preliminary rating for reliability: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Preliminary rating for validity: ☐ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

☒ **Consensus Not Reached**

Committee Pre-evaluation Comments:

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2c)

2a1. Reliability-Specifications: Which data elements, if any, are not clearly defined? Which codes with descriptors, if any, are not provided? Which steps, if any, in the logic or calculation algorithm or other specifications (e.g., risk/case-mix adjustment, survey/sampling instructions) are not clear? What concerns do you have about the likelihood that this measure can be consistently implemented?

- Clearly defined. Measure can be consistently implemented (providing Risk Adjustment calculations are uniform AND available) but direct correlation with "outcomes" remains sketchy in my mind.
- A full list of risk variables is tabulated and should be able to be obtained from claims data. The risk adjustments are based on three statistical models including excess days in care, mortality, and risk of transition to nursing home. If the Days at Home calculation is accomplished by multiple teams, ensuring accuracy of the statistical model and statistical programming will be important for

reproducibility. Having the ability to calculate at the ACO level may be important to some to estimate how the ACO is performing without being beholden to the timeline of the final arbiter of the Days at Home calculation.

- The data elements are clearly defined, and the logic is provided and is clear. The measure appears to be able to be consistently implemented. The exclusions are described and appear to be appropriate. Risk/case mix adjustments are described.
- Specifications unclear
- Some concern with the data which seemed to indicate that the testing may result in under-reporting.
- The SMP raised several issues but seemed "satisfied" with Reliability. It may not be worth the time for the Committee, but I have several issues: how many providers are in the 610 ACOs? What does the analysis of those data show for reliability? How does the risk adjustment appear for provider groups vs. ACOs? What appears to be the outcome of the Days at Home for non-ACOs vs. ACOs?
- Again, not very clear, and not assured that specs and adjustment are reliable at all.
- Based on the notes from the Methods Review Panel included in the Measures Worksheet, there are concerns about specification. I did not receive access to the measure specs to make my own informed judgement.

2a2. Reliability - Testing: Do you have any concerns about the reliability of the measure?

- In so much as I apologetically remain unconvinced that Days at Home equals better patient medical outcomes. The converse is logically true, i.e., that the better the patient is doing from the medical outcome's standpoint, the more time at home that they can likely spend.
- Developer used split half methodology for a short duration of data to calculate an ICC of 0.833. The methodology given to assess reliability may lead to under-estimate of reliability. Implementation should be able to be consistently implemented if claims data is available for the metric calculation.
- The reliability testing is complex, and the methodology has been reviewed by the scientific methods panel. The methods appear to be consistent with accepted methods by the NQF. I don't have concerns based on their analysis.
- Reliability testing not strong
- Testing participants seemed to lack some confidence in the reliability of the measure to distinguish quality of care.
- Not sure.
- Yes, loads.
- Based on data provided, no concerns

2b1. Validity -Testing: Do you have any concerns with the testing results?

- No
- Pearson's correlations were inversely high in expected area, moderately correlated for some and not correlated with measures that might be expected to be negatively correlated.
- The validity testing is complex and has been evaluated by the methods panel.
- Consensus not reached
- Pearson's correlations did not correlate well, ranging between -0.549 and +0.048 resulting in a high inverse correlation for unplanned admissions (expected), moderate correlation with other measures, no correlation with fall risk, and an unexpected inverse correlation with patient experience. Low number of study participants stated that "The Days at Home measure, as specified, can be used to distinguish between better or worse performance at ACOs or provider groups.

- In the submission, there is more information about the "exclusions" for the "not at home" counts. The extensive list made it less concerning about the validity.
- So, a bunch of "experts" said they agree that the measure is valid--this seems like expert opinion, a very low level of validity to me.
- High face validity but low construct validity - will be interested in the discussion. Overall, the outcomes measure seems to have high importance to patients and families.

2b2-3. Other Threats to Validity (Exclusions, Risk Adjustment)2b2. Exclusions: Are the exclusions consistent with the evidence? Are any patients or patient groups inappropriately excluded from the measure? 2b3. Risk Adjustment: If outcome (intermediate, health, or PRO-based) or resource use performance measure: Is there a conceptual relationship between potential social risk factor variables and the measure focus? How well do social risk factor variables that were available and analyzed align with the conceptual description provided? Are all of the risk-adjustment variables present at the start of care (if not, do you agree with the rationale provided)? Was the risk adjustment (case-mix adjustment) appropriately developed and tested? Do analyses indicate acceptable results? Is an appropriate risk-adjustment strategy included in the measure?

- I have the same concerns regarding validity and reliability voiced by the last committee reviewing this measure, in addition to those to which I alluded, above.
- Having multiple models that contribute to this measure, in a manner that may not be fully described/articulated may be cause for concern for the final calculations.
- The exclusions appear to be appropriate. The risk adjustment is described and appears to be complex but appropriate. The list of chronic diseases presumably allows the adjustment for diseases for which hospital utilization is higher such as active metastatic cancer. Hospice enrollment is an exclusion, and this should account for appropriate end of life care.
- Three different risk adjustment models used
- Three models for risk-adjustment. Could compromise comparability of results.
- The risk adjustment seems complicated. Does it improve the validity?
- Again, I am not very sanguine about the risk adjustment.
- Interested in hearing more about the different risk adjustment models and rationale for multiple models

2b4-7. Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data) 2b4. Meaningful Differences: How do analyses indicate this measure identifies meaningful differences about quality? 2b5. Comparability of performance scores: If multiple sets of specifications: Do analyses indicate they produce comparable results? 2b6. Missing data/no response: Does missing data constitute a threat to the validity of this measure?

- From my position dealing with Alternative Payment Models for providers with MCR and MCD patients, the Risk Adjustment calculations stem from checklist diagnoses work done primarily at the PCP level. The quality of this work is dramatically variable, when done at all. Often, third-party vendors have to be hired by managed care/insurers to perform this work who are even further removed from the patient. There is, nevertheless, significant missing patient data overall and individually by year's end.
- Using correlations with measures that are related to hospital stays, which should highly correlate with days out of hospital. Take the high correlation of these measures away, remaining correlations are week, of do not make logical sense. This is a challenge for validity.
- The results appear to product comparable results. The validity tests report face validity testing by an expert panel and they report that the measures are good but there is not an overwhelming sense that the measures are strong. There does not appear to be a problem with missing data that would

threaten the validity. The differences are relatively small and there may be some difficulty identifying meaningful differences.

- Multiple data sources
- Not enough information or data to consider measure validity to be strong.
- The "risk of transition" to nursing home could use some greater analysis and explication for understanding the validity of the measure.
- I think this measure is very obscurely defined and has a large risk of compromising its utility. I am not sure that moderate differences in performance truly represent better quality. I doubt very much the risk adjustment can appropriately account for variations in populations served by ACO's (or is it groups?)
- Missing data would be likely to inflate scores

Criterion 3. Feasibility

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- Developer indicates that the data elements used to compute the measure score are coded by someone other than the person obtaining the original information. Developer also indicates that all data elements are available in defined fields of electronic claims.
- Developer does not provide additional information that may offer insight into associated time, cost, or resource challenges, but developer indicates that there are no difficulties regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.

Questions for the Committee:

- Are the required data elements routinely generated and used during care delivery?
- Are the required data elements available in electronic form, e.g., EHR or other electronic sources?
- Has the developer adequately addressed feasibilities challenges?
- Is the data collection strategy ready to be put into operational use?

Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments:

Criteria 3: Feasibility

3. Feasibility: Which of the required data elements are not routinely generated and used during care delivery? Which of the required data elements are not available in electronic form (e.g., EHR or other electronic sources)? What are your concerns about how the data collection strategy can be put into operational use?

- Risk Adjustment, extenuating circumstances such as social risk factors, risk of death, and risk of transitioning to a long-term nursing home, patient satisfaction, etc.
- There are no specifics of time, cost or resources needed to insure accurate data collection and use of the data. Developer states there are no difficulties in the current workflow of measure development/testing.

- The data elements are based on claims data and the risk adjustment and exclusion criteria are data elements that are generated and used in the routine delivery of care. This should be available as electronic data elements and allow feasible implementation.
- Moderate
- The measure is feasible. However, it appears to be resource intensive. Data should be available in the provider EHRs. Data fields may not be standardized which could create challenges in collecting the data.
- Although feasibility is claimed from the data already collected, what is done to compute the final data set for analysis?
- We are asked to believe the developer here. I am not sure I do! I am not sure we can easily identify all the data consistently.
- No concerns

Criterion 4: Usability and Use

4a. Use (4a1. Accountability and Transparency; 4a2. Feedback on measure)

4a. Use evaluates the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

4a.1. Accountability and Transparency. Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

Planned use in an accountability program? ☒ Yes ☐ No

Accountability program details

- Developer indicates that the measure is intended to enter use on a pay-for-performance basis in both the Primary Care First and Direct Contracting models for payment determinations beginning in 2023.
- Developer also indicates that the measure is intended to be used in quality improvement benchmarking

4a.2. Feedback on the measure by those being measured or others. Three criteria demonstrate feedback: 1) those being measured have been given performance results or data, as well as assistance with interpreting the measure results and data; 2) those being measured, and other users have been given an opportunity to provide feedback on the measure performance or implementation; 3) this feedback has been considered when changes are incorporated into the measure

Feedback on the measure by those being measured or others

- This is a new measure. Developer explains that data and performance results have not yet been provided to entities in a way that would solicit feedback on the measure.
- Developer shares that the entities will have opportunities to provide feedback and ask questions about the measure specifications and interpretation of results through each model's Question & Answer mechanism.
- Developer further explains that feedback gained through this mechanism will be used by the measure steward to inform measure maintenance.

***Questions for the Committee**:*

- Do you believe that the developer has adequately demonstrated use of the measure?

Preliminary rating for Use: ☒ **Pass** ☐ **No Pass**

4b. Usability (4a1. Improvement; 4a2. Benefits of measure)

4b. Usability evaluates the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

4b.1 Improvement. Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated.

Improvement results

- Developer explains the unavailability of data on performance improvement at this time but suggests that performance on the measure will provide a quality signal to providers if their patients are spending more time in select acute and post-acute settings and out of their home or community setting than expected compared to other providers in the same program.

4b2. Benefits vs. harms. Benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

Unexpected findings (positive or negative) during implementation:

- Developer identified no unexpected findings.

Potential harms:

- No harms identified by the developer

***Questions for the Committee**:*

- How could the performance results be used to further the goal of high-quality, efficient healthcare?
- Do the benefits of the measure outweigh any potential unintended consequences?

Preliminary rating for Usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments:

Criteria 4: Usability and Use

4a1. Use - Accountability and Transparency: How is the measure being publicly reported? Are the performance results disclosed and available outside of the organizations or practices whose performance is measured? For maintenance measures - which accountability applications are the measure being used for? For new measures - if not in use at the time of initial endorsement, is a credible plan for implementation provided?4a2. Use - Feedback on the measure: Have those being measured been given performance results or data, as well as assistance with interpreting the measure results and data? Have those been measured, or other users been given an opportunity to provide feedback on the measure performance or implementation? Has this feedback has been considered when changes are incorporated into the measure?

- "Developer also indicates that the measure is intended to be used in quality improvement benchmarking" - remains as an issue because I do not see that the case has been made, or is provable, that more time in the home is, in of itself, directly proportional to or responsible for better healthcare outcomes other than cost.
- Measure will be used in pay-for-performance programs and benchmarking in quality improvement efforts, thus providing feedback to ACOs.
- This is a new measure and has not been publicly reported.
- Use does not match setting--valid for ACO--but what about other non ACO patients.

- Measure is not being publicly reported. Not currently in use.
- In accountability program—Pass
- It's all a promise. I would be more inclined to support this measure if feedback was already obtained.
- New measure

4b1. Usability – Improvement: How can the performance results be used to further the goal of high-quality, efficient healthcare? If not in use for performance improvement at the time of initial endorsement, is a credible rationale provided that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations? 4b2. Usability – Benefits vs. harms: Describe any actual unintended consequences and note how you think the benefits of the measure outweigh them.

- 4b1 - proof of concept has not been successfully made. 4b2 - unintended consequence would be to try to keep the patient at home when this might not be the medically safest place to be.
- The measure could be used as a "barometer" of use of care outside of the home, perhaps helping identify care partners in a catchment area that have prolonged length of stay. The complexity of the measure will make if difficult in most situations to identify single processes or structural issues that could improve days at home, and most likely need a suite of interventions for improvement.
- The data appears to be usable. There are several potential actions that a practice could implement to decrease hospitalization and ED utilization such as availability of urgent care and improved access to health care providers. The potential harm of this measure is that decreased hospital utilization could result in decreased quality of care and appropriate treatment. The developers have tried to account for this by risk stratification, but this still is of concern.
- Moderate
- New measure. There seemed to be some concerns about disparities with use of the measure.
- From the providers perspective do they believe the risk is equal for each of them at the start or it is only from the ACO's level that the risk is equal or fair?
- I am not clear how this will result in the right care being provided. Shouldn't our goal be to provide care at the least intrusive level possible? Could this measure add to the pressure for premature discharge? How are we assured harms will be monitored?
- Developer indicates that the measure is intended to enter use on a pay-for-performance basis in both the Primary Care First and Direct Contracting models for payment determinations beginning in 2023. Potential unanticipated harms could include adverse events or greater caregiver burden because individuals were kept at home. This would need to be monitored.

Criterion 5: Related and Competing Measures

Related or competing measures

The following measure is identified as related:

- NQF 2888 – Risk-Standardized Acute Admission Rates for Patients with Multiple Chronic Conditions

Harmonization

- Developer explains that the Days at Home measure expands on the UAMCC measure priorities of improving care coordination and home-based care while discouraging the use of preventable acute hospital visits by considering the total days spent in care (rather than just total number of admissions), considering care in a broader range of settings, and additionally accounting for mortality and transitions to residential nursing homes to mitigate potential unintended consequences.
- While the eligible cohorts overlap for both measures, they are not identical, with Days at Home including patients younger than 65 as well as patients with different illness or complex conditions;

while the UAMCC cohort includes patients who have two or more specific qualifying conditions, the Days at Home cohort includes patients with complex, chronic illness as defined by an HCC risk score greater than 2.0 (which may be attained through various combinations of risk factors). This is consistent with the cohort of the Primary Care First and Direct Contracting models (which include patients aged 18 and older), and with those models' objective to emphasize care of patients meeting a broad definition of serious illness or complex chronic disease.

- The developer states that both measures are claims-based and there would be no impact on data collection burden for providers reporting either or both measures.

Committee Pre-evaluation Comments: Criterion 5: Related and Competing Measures

5. Related and Competing: Are there any related and competing measures? If so, are any specifications that are not harmonized? Are there any additional steps needed for the measures to be harmonized?

- Not to my knowledge.
- The competing measure (UAMCC) have overlapping but different populations, and this measure looks not at event counts, but on actual days in care settings outside the home
- There appear to be several competing measures and the developers describe strategies to harmonize this measure with others.
- NA
- No related or competing measures.
- Since there is a related measure (2888: Risk-Standardized Acute Admission Rates for Patients with Multiple Chronic Conditions} could these data be utilized for validity analysis?
- Don't think so.
- Days at home is a different/broader concept then NQF 2888 – Risk-Standardized Acute Admission Rates for Patients with Multiple Chronic Conditions

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: 01/12/22

- **Of the one NQF member who has submitted a support/non-support choice:**
  - **Zero support the measure**
  - **One does not support the measure**
- Comment by: **American Medical Association**

The American Medical Association (AMA) appreciates the opportunity to comment on #3667, Days at Home for Patients with Complex, Chronic Conditions. We note that while the submission form indicates that the measure is intended to be used at the Accountable Care Organization (ACO) level, the wording, "provider groups," is used frequently throughout the submission. We request clarification on whether the measure is intended to be used for ACO reporting only or if it would also be applied to other levels such as clinician groups. Based on the specifications and testing completed, we do not believe that it would be appropriate to be applied to any other level, but the submission is not clear on its intent. In addition, The AMA strongly supports the inclusion of individuals with dual eligibility status in the risk model but remains concerned that CMS continues to test social risk factors after the assessment of clinical and demographic risk factors, and it is unclear why this multi-step approach is preferable. On review of the Evaluation of the NQF Trial Period for Risk Adjustment for Social Risk Factors report, it is clear that the approaches to testing these data should be revised to strategies such as multi-level models or testing of social factors prior to clinical factors and that as access to new data becomes available, it may elucidate more differences that are unrelated to factors within an entity's control. Additional testing that evaluates clinical and social risk factors at the same time or social prior to clinical variables rather than the current approach with clinical factors prioritized should be completed. References: National Quality Forum. Evaluation of

the NQF Trial period for Risk Adjustment for Social Risk Factors. Final report. July 18, 2017. Available at: http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=85635. Last accessed January 8, 2022.

- Comment by: **Jake Miller on behalf of Yale CORE**

Yale/CORE clarifications to the methods panel evaluation summary (1 of 2):
Specifications:

In their preliminary analyses, a few SMP members found the specifications confusing and occasionally arbitrary. Some members expressed concerns about the potential misalignment of concept presentations within the submission and noted the denominator statement appeared to lack an explanation of the target population, conditions, settings, and other pertinent measure constructs information. They were also concerned that several concepts included in the submission were not documented as exclusions in the specifications, which both threatens the measure's validity and may incentivize under-treatment of conditions potentially outside the locus of control of the accountable entity, including very low outliers that can never reach the expected performance gains, permanent nursing home residents

NQF Clarification: Please note that the issues noted here were raised by some but not all SMP members and that the summary should clearly reflect these as individual opinions, not the consensus of the entire SMP. Over 60% of subgroup members voted to support this measure on both reliability and validity in the preliminary analysis, indicating they were able to follow the information we provided in the submission. In the final vote after the SMP discussion of these issues, 4 of 10 SMP members still supported the measure validity and voted to pass the measure. It is important not to base this summary solely on the views of a few individual SMP members.

Clarification: The Days at Home measure is population-based and intended to capture performance broadly across eligible beneficiaries. The target population is patients with complex, chronic conditions (who have higher risk for needing complex care) as defined by the inclusion criteria. **This is clearly documented in the submission and should not be noted as lacking**. There are intentionally no denominator exclusions – all beneficiaries meeting the inclusion criteria are included in the denominator because conceptually all are at risk for days in care, and any further exclusions would lack face validity. Some members of the committee may have been confusing the cohort (included beneficiaries) and outcome (days in care that count in the model). We clarify the outcome below. However, it is not accurate to present the measure as "not documenting exclusions."

Clarification: The description of the SMP evaluation seems to reference comments related both to the cohort of included patients (as addressed above) and in the outcome definition of days in care (as clarified here). The measure uses a broad definition of "days in care" consistent with feedback from the Technical Expert Panel (TEP) and aligned with previous work by the Medicare Payment Advisory Commission (MedPAC), reflecting that patients tend to view any time in settings such as inpatient hospitals and facilities as disruptive to their daily life. The consensus recommendation of the TEP was to maintain a broad conception of days in care, so that no types of hospital admission were counted as "days at home." Such a broad definition is not intended to suggest every admission is avoidable, but instead to represent a patient-centered outcome definition which allows for flexibility in improvement strategies. The goal is not to achieve zero days in care, but to reduce the total days in care compared to expectation for a given case mix.

Clarification: It is not accurate to say that "very low outliers" or "permanent nursing home residents" are categorically "outside the locus of control of the accountable entity." Clinical groups and ACOs do have capacity to impact days in acute care for these populations (for example, through more proactive preventive care and improved care coordination to avoid preventable admissions) as confirmed by the TEP.

> • The SMP also questioned whether the consideration of exclusions included (i.e., patients treated in emergency departments, admitted to acute care settings, and days after a death occurs), indicated low-quality care. Another SMP member expressed concerns with adjusting for transitions to the nursing home, which purports that moving from home to a nursing home, is always negative. Other concerning date elements included permanent nursing home admissions requiring skilled nursing care, which may include personal and community resources that are not be modifiable by the accountable entity.

Clarification: As noted above, the Days at Home measure does not conceptually assume that *all* days in an included setting indicate low-quality care, and the goal is not to achieve zero days in these settings. Rather, the goal is to encourage providers to explore home-based options or other feasible means so that their patients can spend *fewer* days in these settings. Moreover, days after a death occurs are not counted as either days at home or days of acute care, but rather as unmeasured days.

Clarification: The goal of adjusting for nursing home transitions is to encourage providers to explore care options, such as providing home-and community-based care, preventive care services, or improved care coordination, which relieve some of the burden on their patients (and family/caregivers) while allowing patients to remain in their home and community longer. While in some cases a transition to nursing home is the best outcome for a patient, the TEP and CMS agreed this outcome is more often less desirable than remaining in the community setting and that the measure should not have the unintended consequence of rewarding providers who are quicker to transition patients to nursing homes. The adjustment is designed to have a modest effect on measure scores in those cases where there are much higher rates of transition than expected given the case-mix of patients. The current approach was developed as a compromise between counting days in a nursing home as "acute care days" and counting them as "days at home," both options that include notable drawbacks as discussed by the TEP.

Clarification: While most long-term nursing home residence days are considered "days at home," days in which **skilled** nursing care is utilized *do* count as "days in care."

• SMP members also noted that the unit of analysis reported in the measure vacillates between accountable care organizations (ACOs) and provider group.

Clarification: The measure is intended for use in different settings in which accountable entities comprise groups of individual providers, including provider groups and ACOs; the specifications have used the general term "provider group" to capture these different organizations. The term "ACO" is used only in documentation pertinent to the *testing* of the measure, which used a dataset of 2017-2018 Shared Savings Programs ACOs and aligned beneficiaries.

• One SMP member questioned whether this measure, which combines multiple risk models calculations into a single overall score, should be considered a cost composite measure.

Clarification: Days at Home is not a composite measure; it measures a single outcome. The mortality and nursing home transition component models are not standalone measures, nor are they intended to capture different outcomes. These component models are included as a means of safeguarding against potential adverse consequences for the measure that were identified in conversation with CMS, the TEP, and other experts. The only outcome is days at home, which is adjusted for multiple risk factors, as well as for unexpectedly high mortality or nursing home transition rates. This is demonstrated empirically in test results as noted in the additional comments in the final measure submission; the quality signal of the measure is dominated by the Days in Care component and the additional adjustments result in modest changes for a small number of ACOs.

Validity

The developer conducted construct validity with Pearson correlations to six other ACO-level measures hypothesizing that quality conceptually relates to excess days in care (EDIC) for patients with complex chronic diseases.

- Pearson's correlations did not correlate well, ranging between -0.549 and +0.048 resulting in a high inverse correlation for unplanned admissions (expected), moderate correlation with other measures, no correlation with fall risk, and an unexpected inverse correlation with patient experience.
- The developer explained that this is possibly due to endogeny of the hospital admissions and readmissions measures. The developer also reported the poor correlations may result from testing against measures using smaller sample sizes and which were not risk adjusted for clinical variables.

Clarification: This summary does not accurately reflect the developer's explanation. We documented the expected modest correlations in a direction that was pre-specified. The measures with significant correlation in the expected direction have key and notable differences in cohort (the patients included and the time period for measurement) and outcome (the settings included and the outcome metric) from Days at Home, despite some overlap. These measures were intended to assess construct validity because they measure similar aspects of quality in distinctly different way. These results do not undermine the validity of the measure as we would expect similar results across providers between similar measures.

- Comment by: **Jake Miller on behalf of Yale CORE**

Yale/CORE clarifications to the methods panel evaluation summary (2 of 2): Risk-Adjustment *The SMP members had concerns with the model construction, which they agreed lacked vital adjustment and consideration for many variables without theoretical or empirical justifications and used arbitrary measure weighting. The developers acknowledge these were not empirically assessed, but rather are subjective and based solely on TEP recommendation.*

Clarification: The Days in Care statistical count model includes an offset for days alive, so that "mortality days" are not counted in either the numerator nor denominator of the main measure component ("excess days in care" or EDIC). The Days in Care measure does incorporate an adjustment to EDIC for the excess mortality risk of the measured provider groups, as well as the excess risk of transition to nursing home. These adjustments are made by multiplying the EDIC by a standard mortality ratio (SMR) and by 0.5 times a standard nursing home transition ratio (SNHR). The SNHR is scaled to have the same distribution as SMR and then given a relative weight of 0.5, to accommodate feedback received from the TEP that nursing home transition as an outcome is less severe than death but should still be reflected in performance scores. Both the SMR and SNHR adjustments have a minor impact on the overall score, except in the case of extreme differences from the average provider group risk of mortality or nursing home transition.

Clarification: The "nursing home start date of January 1" refers to the classification of beneficiaries; those already in a nursing home on January 1 are not considered for a nursing home transition during the measurement period. This start date aligns exactly with the specified performance period for the measure of January 1 to December 31 (the calendar year). o A few SMP members discussed the effect of specific chronic conditions on the risk model, such as cancer, dementia, and congestive heart failure that increase EDIC by nature of the disease states.

Clarification: The measure includes risk adjustment to account for differences in case mix between providers, including for these stated factors. While these conditions may result in more observed (unadjusted) days in care for patients, risk adjustment accounts for this increased risk and these patients will not necessarily have more excess days in care. o The greatest concern for the risk adjustment model expressed from the SMP members was the development approach for days at home, and the mortality and nursing models. The SMP noted that formulas in the approach may include doubling the EDIC estimates for enrolled ACOs and negative impacts to the penalty schematic

Clarification: It is unclear what "faulty formulas" are being referenced here, what "doubling" is described, or how the specifications compromise the validity of the measure. The formulas used were endorsed by the TEP, which included members with expertise in measure development who had reviewed the approach and results in great detail. Performance on the measure is driven by the Days in Care model, which is a conventional risk adjustment model. The score is then modified such that only provider groups with both outlying performance in Days in Care and nursing home transitions and/or mortality are noticeably impacted. It is not true that this results in "doubling the estimates" for some providers. It is also not clear what "negative impacts to the penalty schematic" means in this statement or what "fault" in the specifications is proposed to give rise to that. Without more detail, it is difficult to further address the challenges being put forward.

Exclusions

- The SMP questioned the process-outcome pathway that resulted in increased, rather than decreased, days in care, and the lack of exclusions for long-term nursing home residents prior to a measurement period, who have no chance of "at home" days defined in the specifications.

Clarification: This is not an accurate description of the methodology. **Patients who reside in long-term nursing homes are considered "at home" for purposes of the Days in Care model.** For example, a nursing home resident on January 1 with no other care use during the year would be considered "at home" for the full 365 days. Similarly, for patients who transition to a nursing home during the performance year, all subsequent days in the nursing home with no other care use are counted as "days at home." o SMP members indicated the discrimination and calibration were generally acceptable but had concerns related to the low outliers. The developer described this as an unintended consequence of the measure construct as the measure attempts to balance days at home with other unintended consequences.

Clarification: The measure does not have a strict definition of outliers, nor is it proposed to report outliers. In clarification of results the SMP may be referring to, certain ACOs observed in testing with scores much lower than average did not arise as a result of "attempting to balance days at home with other intended consequences." These ACOs in the test dataset already had substantially more days in care than expected, based on the Days in Care model results even before accounting for nursing home transitions and mortality, and their low performance is unrelated to the additional adjustments. The nursing home and mortality adjustments simply have the greatest potential impact for provider groups that are already outliers (either high or low) in Excess Days in Care. The measure was designed to ensure that it is extremely difficult for a provider group with near-average Excess Days in Care to become a very high or very low performer due solely to outlying performance in the nursing home or mortality models.

Meaningful Differences

A few SMP members questioned the presence of meaningful differences in performance and the use of the measure for quality improvement purposes, and whether the measure could be used to identify differences in patient function or health-related quality of life.

Clarification: While scores are reported as "days at home" to align with the conceptual focus of the measure, differences in performance should be considered relative to days in care which are the basis of the main Days in Care model. As noted in the measure submission, the interquartile range of 3.0 days at home (329.1 – 332.1) reflects those patients of a provider at the 25th percentile of performance can each expect to spend on average 3.0 days more in care than they could expect at a provider at the 75th percentile of performance. As the average patient in the cohort spends 12.8 days in care, 3 days more or fewer represents a meaningful amount of time for each patient who, as noted above, strongly prefer to minimize time in these care settings when possible.

Combined Methods Panel Scientific Acceptability Evaluation

RELIABILITY: SPECIFICATIONS

1. **Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented? ☒ Yes ☒ No**

   Submission document: Items sp.01-sp.30

   ***NOTE**: NQF staff will conduct a separate, more technical, check of eCQM specifications, value sets, logic, and feasibility, so no need to consider these in your evaluation.*

2. **Briefly summarize any concerns about the measure specifications.**

   For example: Which data elements, if any, are not clearly defined? Which codes with descriptors, if any, are not provided? Which steps, if any, in the logic or calculation algorithm or other specifications (e.g., risk/case-mix adjustment, survey/sampling instructions) are not clear? What concerns do you have about the likelihood that this measure can be consistently implemented?

   Reviewer 2: The method for combining the information from the 3-risk adjusted statistical models, "adjusted day in care", mortality risk and risk of transfer to a nursing home, for stratification is confusing.
   Reviewer 3: The denominator specifications does include several exclusions/exceptions. It may be best to include those under the exclusion list (currently stated that no exclusions are listed). Have the developers considered other exclusions related to specific reasons for being accepted into acute care/ED, which might not indicate low quality of the accountable entity? For this measure to be valid, it is important that such cases are excluded to avoid incentivizing under-treatment of such conditions that may be outside of the control of the accountable entity. Otherwise, there could be a serious threat to the measure's validity.
   Reviewer 5: No concerns.
   Reviewer 7: No concerns over implement ability.
   Reviewer 8: My primary concern is the adjustment for transition to nursing home. The days at home measure has an adjustment for death, which makes sense (to some extent), but also bakes in a logic that moving from home to a nursing home is always negative. This philosophical decision should be made more explicit in the specifications. Also, the down-weighting of nursing home conversion seems arbitrary.
   Reviewer 11: The measure title says that the measure is about days at home for patients with complex chronic medical conditions, but there seems to be no specific identification of these conditions, nor any denominator definition linked specifically to any chronic conditions. There is mention of an HCC value >2, but no discussion of whether this is a marker for "complex, chronic conditions", and if so, why. In addition, the measure is presumably an ACO-level measure and was tested at the ACO level, but the text at several points refers to medical groups as the possible unit of analysis. ACO and medical group aren't the same thing.
   Reviewer 12: None.

Submission document: Questions 2a.01-09

3. **Reliability testing level**

   For example: for some types of measures, if patient/encounter level validity is demonstrated, additional reliability testing is not required. Please review table above.

   ☒ **Accountable-Entity Level** ☒ **Patient/Encounter Level** ☐ **Neither**

4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure**

   NOTE: "level of analysis" reflects which entity is being assessed or held accountable by the measure. For example: If a measure is specified for a clinician level of analysis, but facility-level testing is provided, then testing does NOT match level of analysis. Or, if two levels of analysis are specified (e.g., clinician and facility) but testing is conducted for only one, then testing does NOT match level of analysis. Or, if claims data are selected as a data source, but testing data doesn't include claims data, then testing does NOT match data source.
   Also, check "NO" if only descriptive statistics are provided or submitter only describes process for data management/cleaning/computer programming.

   ☒ **Yes** ☐ **No**

5. If accountable-entity level and/or patient/encounter level reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical VALIDITY testing** of **patient-level data** conducted?

   According to current guidance patient/encounter level validity testing can be used for patient/encounter level reliability testing. Answer ONLY if you responded "Neither" on question #3 and/or "No" to question #4. Note that for some types of measures, additional reliability testing is not required IF patient/encounter level validity is demonstrated.

   ☐ **Yes** ☐ **No**

6. **Assess the method(s) used for reliability testing**

   Submission document: Question 2a.10

   For example: Is the method(s) appropriate? If not, please explain (and offer potential alternatives if possible). Does the testing conform to NQF criteria and guidance? Was testing was conducted with the data source and level of analysis indicated for this measure? Address each level of testing provided, and each analysis under each method.

   Reviewer 2: Split-half reliability was performed for data elements on risk factor variables and on scores on the outcome measure. This method was within NQF reliability guidance.
   Reviewer 3: No concerns.
   Reviewer 4: Split half icc.
   Reviewer 5: The measure score testing conducted was appropriate for this type of measure. A split-half methodology was used. An intraclass correlation coefficient (ICC) test was performed.
   Reviewer 7: Method appropriate.
   Reviewer 8: Split-half methodology in one year of data - would be better to repeat the analysis in multiple years of data, particularly because ACO assignment rules are adjusted annually.
   Reviewer 9: Split half methodology.
   Reviewer 10: Split half testing of the measure score was performed with an ICC of 0.8326.
   Reviewer 11: The developers used a basic split-half method with some form of ICC used to assess agreement between the two split halves. The form of ICC is not specified.

Reviewer 12: No concerns. The developers estimated ICC using a random split-half methodology. This is a common approach to estimating reliability of risk-adjusted outcomes. The method involved randomly splitting each provider's data in half, estimating provider-specific outcomes separately in each half, and then using the 2 estimates per provider to estimate ICC. The reported ICC may under-estimate true reliability because scores are estimated using only half of each provider's data.
Reviewer 13: Intraclass coefficient.

7. **Assess the results of reliability testing**

   Submission document: Question 2a.11

   For example: Is the test sample adequate to generalize for widespread implementation? Is there high or moderate confidence that the measure results and/or the data used in the measure are reliable? Address each level of testing provided, and each analysis under each method.

   Reviewer 2: Data element and score level reliability results appear adequate.
   Reviewer 4: ICC 0.8326.
   Reviewer 5: The intraclass correlation coefficient (ICC) test result was 0.83, which is high / good. The result indicates a given provider rating is reliability & unlikely due to random chance.
   Reviewer 7: Split sample shows good correlation.
   Reviewer 8: The ICC is 0.83 comparing two samples in the same year of data. It would have been more convincing to test this measure on a broader sample of delivery systems and not just ACOs. I also wonder about reliability for smaller ACOs/provider groups?
   Reviewer 9: ICC = 0.83 - good score.
   Reviewer 10: Data elements were claims and enrollment data so not formally tested.
   Reviewer 11: ICC results are strong and acceptable, but the form of ICC is not specified, so interpretation of the ICC result is not as clear as it could be.
   Reviewer 12: The estimated ICC was 0.83. This is very good reliability for a risk-adjusted outcome.
   Reviewer 13: ICC = 0.83.

8. Was the method described and appropriate for assessing the proportion of variability due to real differences among measured entities? **NOTE:** If multiple methods used, at least one must be appropriate.

   Submission document: Question 2a.10-12
   For example: Appropriate signal-to-noise analysis; random split-half correlation; other accepted method with description of how it assesses reliability of the performance score.

   ☒ **Yes**
   ☐ **No**
   ☐ **Not applicable**

9. Was the method described and appropriate for assessing the reliability of ALL critical data elements?

   Submission document: Question 2a.10-12
   For example: inter-abstractor agreement (ICC, Kappa); other accepted method with description of how it assesses reliability of the data elements
   Answer NO if: only assessed percent agreement; did not assess separately for all critical data elements (or at minimum, for numerator, denominator, exclusions)

   ☒ **Yes**
   ☐ **No**
   ☒ **Not applicable** (patient/encounter level testing was not performed)

10. **OVERALL RATING OF RELIABILITY** (taking into account precision of specifications and all testing results):

    ☒ **High** (NOTE: Can be HIGH **only if** accountable-entity level testing has been conducted)

    ☒ **Moderate** (NOTE: Moderate is the highest eligible rating if accountable-entity level testing has not been conducted)

    ☐ **Low** (NOTE: Should rate **LOW** if you believe specifications are NOT precise, unambiguous, and complete or if testing methods/results are not adequate)

    ☐ **Insufficient** (NOTE: Should rate **INSUFFICIENT** if you believe you do not have the information you need to make a rating decision)

11. **Briefly explain rationale for the rating of OVERALL RATING OF RELIABILITY and any concerns you may have with the approach to demonstrating reliability.**

    Reviewer 2: Although the magnitude of between vs. within ACO variation was not assessed, split-half reliability results are adequate and within NQF guidance.
    Reviewer 4: This is a complex measure and approach. My hesitation to score as HIGH is based on my own ignorance here.
    Reviewer 5: Response to Q7: The intraclass correlation coefficient (ICC) test result was 0.83, which is high / good. The result indicates a given provider rating is reliability & unlikely due to random chance.
    Reviewer 7: Split sample method produced surprisingly good correlation.
    Reviewer 8: I think the evidence is helpful, but not sufficient. As indicated above, more testing in different years would be helpful. Also, would like to see results that show variation by ACO/provider organization size.
    Reviewer 9: ICC is high according to most standards.
    Reviewer 10: ICC for measure score was moderate.
    Reviewer 11: The ICC value for split-half reliability testing looks acceptable, but the specific form of ICC test used should have been specified.

VALIDITY: TESTING

12. **Validity testing level (check all that apply):**

    ☒ **Accountable-Entity Level** ☒ **Patient or Encounter-Level** ☐ **Both**

13. **Was the method described and appropriate for assessing the accuracy of ALL critical data elements?** **NOTE** that data element validation from the literature is acceptable.

    Submission document*:* Questions 2b.01-02.
    For example: Data validity/accuracy as compared to authoritative source- sensitivity, specificity, PPV, NPV; other accepted method with description of how it assesses validity of the data elements.
    Answer NO if: only assessed percent agreement; did not assess separately for all critical data elements (or at minimum, for numerator, denominator, exclusions)

    ☒ **Yes**

    ☒ **No**

    ☒ **Not applicable** (patient/encounter level testing was not performed)

14. **Method of establishing validity at the accountable-entity level:**

NOTE: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.

Submission document: Questions 2b.01-02

☒ **Face validity**

☒ **Empirical validity testing at the accountable-entity level**

☐ **N/A (accountable-entity level testing not conducted)**

15. **Was the method described and appropriate for assessing conceptually and theoretically sound hypothesized relationships?**

Submission document: Question 2b.02

For example: Correlation of the accountable-entity level on this measure and other performance measures; differences in performance scores between groups known to differ on quality; other accepted method with description of how it assesses validity of the performance score

☒ **Yes**

☐ **No**

☐ **Not applicable** (accountable-entity level testing was not performed)

16. **Assess the method(s) for establishing validity**

Submission document: Question 2b.02

For example:

- *If face validity the only testing conducted: Was it accomplished through a systematic and transparent process, by identified experts, explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality, and the degree of consensus and any areas of disagreement provided/discussed?*
- *If a maintenance measure, but no empirical testing conducted, was justification provided?*
- *If construct validation conducted, was the hypothesized relationship (including strength and direction) described and does it seem reasonable?*

Reviewer 2: Face validity was assessed with 21 TEP members who reviewed the measure for appropriateness for quality assessment at the ACO level. Construct validity was assessed using correlations with other ACO-level measures of quality conceptually related to excess days of care for patients with complex chronic diseases. Both are acceptable methods.
Reviewer 3: No concerns.
Reviewer 4: Inverse association with both unplanned readmissions and timely care.
Reviewer 5: The testing is appropriate for the given measure. More specifically, a summary of the testing follows: Empirical testing: Examined correlation ACO performance in the Days at Home measure & 6 other measures. Face validity testing: Review of the measure by a TEP in regard to technical specifications. After the review, the statement posed to the TEP for their reaction was: "The Days at Home measure, as specified, can be used to distinguish between better or worse performance at ACOs or provider groups."
Reviewer 7: Correlation with other measures is established approach.
Reviewer 8: The authors look at correlation between their measure score and other ACO measures scores. Would like to see some predictive validity testing.
Reviewer 9: Face validity via a TEP and correlation with other ACO outcome measures.

Reviewer 10: Measure score was validated by TEP and testing for correlation with other utilized quality measures, such as ACO-8, ACO-38, ACO-13, ACO-43, ACO-35, and ACO-1 with Pearson values between -0.549 and +0.048, the latter being falls.
Reviewer 11: Methods were reasonable - agreement with a set of potentially related process or outcome measures, and face validity assessment by a TEP.
Reviewer 12: Developers used Pearson correlations to compare provider-specific estimates for the proposed measure to various established measures that were hypothesized to capture related aspects of quality.
Reviewer 13: Face validity with a TEP. Correlation with other measures.

17. **Assess the results(s) for establishing validity**

Submission document: Questions 2b.03-04

For example: Is the test sample adequate to generalize for widespread implementation? Do the results demonstrate sufficient validity so that conclusions about quality can be made? Do you agree that the score from this measure as specified is an indicator of quality?

Reviewer 2: Results from the 19 respondents from the TEP indicate support for the face validity of this measure. The correlations of the days at home ACO measure with other ACO quality measures were generally significant and in the hypothesized direction with the exception of the "screening for fall risk" measure (NS) and the survey-based timely care, appointments and information measure which was significant but in the inverse from hypothesized direction. The latter was assessed on a smaller sample size and not risk adjusted for clinical variables but warrants further explanation.
Reviewer 3: Results were satisfactory.
Reviewer 4: Above. Overall fine.
Reviewer 5: The empirical testing results indicate the measure is valid. In four of the six comparisons of measure results, the directionality of the correlation was as expected. Regarding face validity testing, all panelists agreed (to varying degrees) with the statement that the measure can be used to discern provider performance. More specifically, testing results follows: Empirical testing: The following are the expected & actual correlation between the measure & other measures: ACO-8: Expected: Moderate inverse correlation Actual: Inverse correlation ACO-38: Expected: Low inverse correlation Actual: Inverse correlation ACO-13: Expected: Low inverse correlation Actual: Inverse correlation ACO-43: Expected: Low inverse correlation Actual: Inverse correlation ACO-35: Expected: Low inverse correlation Actual: No sig. correlation ACO-1: Expected: Low positive correlation Actual: Inverse correlation Face validity testing: The statement posed to the TEP for their reaction was: “The Days at Home measure, as specified, can be used to distinguish between better or worse performance at ACOs or provider groups.” The response was: 2 members indicated “strongly agree,” 15 indicated “agree,” and 2 indicated “somewhat agree.”
Reviewer 7: This measure is basically an excess day’s measure with adjustment for SMR and rSNHR. Correlation with excess days measure is relatively high, but not truly independent. My main concern about threats to validity is the failure to justify key measure construction approaches: --Method for adjusting for SMR and rSNHR is not justified theoretically or empirically. The weight given these measures in adjustment is arbitrary, and thus the value or validity over the excess days of care measures not established. --Elective and patient desired inpatient care not addressed. E.g., joint replacement that cannot be done on an outpatient basis. Pregnancy-related care is excluded but is not the only care that might be sought by patients**.**
Reviewer 8: The two strongest correlations were measures about admissions to an inpatient stay. This seems inappropriate since the measure in question is basically 'days out of the hospital/nursing home'. The remaining measure correlations where relatively weak with no theoretical reason some correlations should be stronger than others.
Reviewer 9: Significant correlation with 4/6 measures showing construct validity. TEP - 17/19 members agreed or strongly agreed that the measure was useful to distinguish performance.

Reviewer 10: Analysis for differences in performance consisted of ranges of differences in the scores by quartile percentages. Pearson correlation coefficient with other measures resulted in a high inverse for unplanned admissions (expected), moderate with other measures, no correlation with fall risk, and an unexpected inverse correlation with patient experience. This was attributed to the range of focus for the measures compared.
Reviewer 11: Validity is acceptable, as empirical results and face validity results were both at least up to prevailing standards and expectations.
Reviewer 12: Pearson correlations differed across comparison measures. I know that NQF requires correlation analyses, but I don't attach much importance to them personally.
Reviewer 13: Strong support from TEP. Predictable correlation with other utilization measures, including necessarily negative correlation with unplanned admissions.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

18. **Please describe any concerns you have with measure exclusions.**

Submission document: Questions 2b.15-18.

For example: Are there exclusions? If so, are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)? Are any patients or patient groups inappropriately excluded from the measure? If patient preference (e.g., informed decision-making) is a basis for exclusion, does it impact performance and if yes, is the measure specified so that the information about patient preference and the effect on the measure is transparent? If you have concerns based on a clinical rationale, please note here as well as in question #29.

Reviewer 2: There are no denominator exclusions.
Reviewer 3: See comments above regarding the need to comprehensively define all exclusions related to days in acute care that may be outside of the accountable entity's control. Orthopedic trauma comes to mind as an example.
Reviewer 5: No concerns as there are no exclusions.
Reviewer 7: No concerns with exclusions.
Reviewer 8: None.
Reviewer 10: The testing did not use claims data with missing data.
Reviewer 11: No concerns, although the choice to have a statistical model for risk of death rather than using data on death directly seems a bit curious. Patients who die during the measurement year can be neither "at home" nor in an institution after the death date, so it would have seemed reasonable to exclude patients from counts of both numerator and denominator days after death occurs.
Reviewer 12: None.

19. **Risk Adjustment**

Submission Document: Questions 2b.19-32
Applies to all outcome, cost, and resource use measures. Please answer all checkbox questions (19a -19d), then elaborate on your answers in your response to 19e.

19a. **Risk-adjustment method**

☐ None ☒ Statistical model ☐ Stratification

☐ Other method assessing risk factors (please specify)

19b. **If not risk-adjusted, is this supported by either a conceptual rationale or empirical analyses?**

☒ Yes ☐ No ☒ Not applicable

19c. **Social risk adjustment:**

19c.1 Are social risk factors included in risk model? ☒ Yes ☒ No ☐ Not applicable

19c.2 Conceptual rationale for social risk factors included? ☒ Yes ☒ No

19c.3 Is there a conceptual relationship between potential social risk factor variables and the measure focus? ☒ Yes ☐ No

19d.**Risk adjustment summary:**

19d.1 All of the risk-adjustment variables present at the start of care? ☒ Yes ☐ No

19d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion? ☒ Yes ☐ No

19d.3 Is the risk adjustment approach appropriately developed and assessed? ☒ Yes ☒ No

19d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration) ☒ Yes ☒ No

19d.5.Appropriate risk-adjustment strategy included in the measure? ☒ Yes ☒ No

19e**. Assess the risk-adjustment approach**

For example: If measure is risk adjusted*:*

- *If the developer asserts there is no conceptual basis for adjusting this measure for social risk factors, do you agree with the rationale?*
- *How well do social risk factor variables that were available and analyzed align with the conceptual description provided?*
- *Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented?*
- *Are all of the risk adjustment variables present at the start of care (if not, do you agree with the rationale)?*
- *If social risk factors are not included in the risk-adjustment approach, do you agree with the developer's decision?*
- *Is an appropriate risk-adjustment strategy included in the measure (e.g., adequate model discrimination and calibration)?*
- *Are all statistical model specifications included, including a "clinical model only" if social risk factors are included in the final model?*

If measure is NOT risk-adjusted*:*

- *Is a justification for not risk adjusting provided (conceptual and/or empirical)?*
- *Is there any evidence that contradicts the developer's rationale and analysis for not risk-adjusting?*

Reviewer 2: The decision not to include social risk factors (SRFs) in the final risk model was based largely on the change in R-square for the model with only clinical variables vs. the addition of SRFs, the "weak empirical support" (despite significant OR's for e.g. dual-eligibility), the difficulty in data collection from external sources, and the lack of control/mitigation of variables' impact on days at home by ACOs. The argument is not very persuasive, especially for dual eligibility. Discrimination and calibration data appear acceptable except for the highest days in care decile, raising issues related, e.g., to outliers, etc.

Reviewer 5: The risk adjustment approach is sound and appropriate for the nature of the measure. The testing to assess how the risk model performs is sufficient.

Reviewer 7: Avoidance of institutional care is a function of factors not included in the risk adjustment model including presence of spouse, other informal care givers, formal home health services. Some of these are not available in the data, and area level proxies are inadequate. Deviance r-square low.

Reviewer 8: RA is a strength of this measure - well thought through and tested. I particularly like the split half method of developing the data in one sample and testing it in another. I'm not totally clear if they mixed two years of data to do this, but that's probably fine either way for 2017-2018.

Reviewer 9: Three components are risk adjusted using different methods - unclear is the composite risk adjusted value is the correct approach - would like to discuss with group.

Reviewer 10: 53 risk factors were utilized in the model and tested. An adjusted rate ratio was derived and compared to actual performance. Social risk factors included dual eligibility.

Reviewer 11: The approach is very complex compared to almost all other measures that come to NQF, as there are three adjustment models and not just one. It wasn't entirely clear whether the model results get combined at some point to yield just one measure, or whether several forms of the measure can be used depending on which specific adjustments are applied.
Reviewer 12: The measure incorporates a detailed case mix adjustment using over 50 covariates. Three different models contribute to the measure: excess days model, mortality model, nursing home transition model. I don't have concerns about the adequacy of case mix adjustment, but I do have some concerns about results from excess days, mortality, and nursing home transitions are combined together. This isn't an issue with "risk adjustment" per se so I won't go into detail here.
Reviewer 13: C-statistic of components models in 0.7-0.8 range.

20. **Please describe any concerns you have regarding the ability to identify meaningful differences in performance.**

    Submission document: Questions 2b.05-07

    For cost/resource use measures, does this measure identify meaningful differences about cost and resource use between the measured entities?

    Reviewer 2: Despite the fact that the adjusted mean days at home was 330.4 (3.7), the interquartile range was only 3 days. Further, although the developer argues that differences of 3 days should not be considered trivial from a cost perspective, it is not clear whether this magnitude reflects a difference in quality of care, manifested for example, in differences in patient function or health-related quality of life. The developer also notes substantial within ACO variance in days spent at home but appears not to have tested between vs. within ACO variance adjusted for risk factors.
    Reviewer 5: The testing performed looked at difference in provider performance at several levels, e.g., interquartile range. While this is adequate, it would be preferable to see the degree to which providers' results are significantly "better" and "worse."
    Reviewer 7: Need more discussion of mean days in care and not in care and distribution among patients to assess whether reported interquartile ranges reflect meaningful differences.
    Reviewer 8: None.
    Reviewer 11: The concept of "meaningful" here is clearly in the eye of the beholder. The developers and CMS may feel that a handful of days more or less out of a possible 365 are "meaningful', and they attempt to make that argument. It is also quite possible, though, that differences in that range reflect effects of variables not included in the adjustment models, or in residual effects not fully adjusted for, so that the observed differences after adjustment are not really about performance. The differences that are discussed are at the extremes of the distribution - it's not clear at all that differences in the middle of the distribution are meaningful.
    Reviewer 12: None.

21. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**
    Submission document: Questions 2b.11-14.

    ***Note**: This item is directed to measures that are risk-adjusted (with or without social risk factors) **OR** to measures **with more than one set of specifications/instructions**. It does **not apply** to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator). Comparability is not required when comparing performance scores with and without social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.*

Note if not applicable. Note if applicable but not addressed. If multiple sets of specification (e.g., due to different data sources or methods of data collection): Do analyses indicate they produce comparable results?

Reviewer 2: None.
Reviewer 5: No concerns as there was only 1 data source employed.
Reviewer 7: N/A
Reviewer 8: N/A
Reviewer 11: N/A
Reviewer 12: None.

22. **Please describe any concerns you have regarding missing data.**

    Submission document: Questions 2b.08-10.

    For example: Are there any sources of missing data not considered? Is it clear how missing data are handled? Is missing data more of a problem for some providers or patients than others? Does the extent of missing data impact the validity of the measure?

    Reviewer 2: None.
    Reviewer 5: No concerns as claims data was used in the measure.
    Reviewer 7: N/A
    Reviewer 8: None.
    Reviewer 10: None.
    Reviewer 11: No concerns.
    Reviewer 12: None.

For cost/resource use measures ONLY:

If not cost/resource use measure, please skip to question 25.

23. **Are the specifications in alignment with the stated measure intent?**

    Consider these specific aspects of the measure specifications: attribution, cost categories, target population.

    ☐ **Yes** ☐ **Somewhat** ☐ **No (If "Somewhat" or "No", please explain)**

24. **Describe any concerns of threats to validity related to attribution, the costing approach, carve outs, or truncation (approach to outliers):**

    Attribution: Does the accountable entity have reasonable control over the costs/resources measured? Is this approach aspirational (intending to drive change) or was it developed based on current state?
    Costing Approach: Do the cost categories selected align with the measure intent, target population and care settings? Is the approach for assigning dollars to resources
    Carve Outs: Has the developer addressed how carve outs in the data source are handled (or should be handled for other users)? For example, if pharmacy data is carved out (missing) from the data set, can a measure that focuses on cost of care for asthmatics still be valid?
    Truncation (approach to outliers): What is the threshold for outliers (i.e., extremely high cost or low-cost cases) and how are they handled?

25. **OVERALL RATING OF VALIDITY taking into account the results and scope of all testing and analysis of potential threats.**

    ☒ **High** (NOTE: Can be HIGH only if accountable-entity level testing has been conducted)

    ☒ **Moderate** (NOTE: Moderate is the highest eligible rating if accountable-entity level testing has NOT been conducted)

☒ **Low** (NOTE: Should rate LOW if you believe that there **are** threats to validity and/or relevant threats to validity were **not assessed OR** if testing methods/results are not adequate)

☐ **Insufficient** (NOTE: For instrument-based measures and some composite measures, testing at both the accountable-entity level and the patient/encounter level **is required**; if not conducted, should rate as INSUFFICIENT.)

26. **Briefly explain rationale for rating of OVERALL RATING OF VALIDITY and any concerns you may have with the developers' approach to demonstrating validity.**

Reviewer 2: Face and construct validity appear to be adequate. Failure to include significant social risk factors is of concern, therefore a moderate score seems appropriate.
Reviewer 5: Response excerpted from Q17: The empirical testing results indicate the measure is valid. In four of the six comparisons of measure results, the directionality of the correlation was as expected. Regarding face validity testing, all panelists agreed (to varying degrees) with the statement that the measure can be used to discern provider performance.
Reviewer 7: With lack of theoretical or empirical basis for weight given SMR and NHSR, interpretation and validity of score cannot be assessed.
Reviewer 8: I think the validity testing is under-developed for this measure, but the strong risk adjustment offers ACOs a fair degree of protection from arbitrary results. As a results, I gave the measure a moderate rather than a low.
Reviewer 10: Calculation of days at home involves standardized ratios for mortality and for likely transition to nursing home. This then translates to excess days and contributes to an adjusted days at home. The model is clear and has some assumptions such as above. C-statistic was 0.738 for the mortality model, 0.760 for nursing home transition. Deviance from R-squared was 0.170 for the days in care model. Spearman rank correlation was 0.346 for more days in care.
Reviewer 11: Validity is acceptable, as empirical results and face validity results were both at least up to prevailing standards and expectations.
Reviewer 12: The measure purports to estimate risk-adjusted days at home but in fact it seems to be estimating a hard-to-describe combination of days at home, mortality, and transitions to nursing home. Is this not a composite measure? The method of combining results from 3 domains into a single number is somewhat unusual and the final result does not appear to have a simple or obvious interpretation. It's not clear how much each individual domain (excess days, mortality, nursing home) ends up contributing to the overall score.

For composite measures ONLY
If not composite, please skip this section.
Submission documents: Questions 2c.01-08
Examples of analyses:
1) If components are correlated - analyses based on shared variance (e.g., factor analysis, Cronbach's alpha, item-total correlation, mean inter-item correlation).
2) If components are not correlated - analyses demonstrating the contribution of each component to the composite score (e.g., change in a reliability statistic such as ICC, with and without the component measure; change in validity analyses with and without the component measure; magnitude of regression coefficient in multiple regression with composite score as dependent variable, or clinical justification (e.g., correlation of the individual component measures to a common outcome measure).
3) Ideally, sensitivity analyses of the effect of various considered aggregation and weighting rules and the rationale for the selected rules; at a minimum, a discussion of the pros and cons of the considered approaches and rationale for the selected rules.
4) Overall frequency of missing data and distribution across providers. Ideally, sensitivity analysis of the effect of various rules for handling missing data and the rationale for the selected rules; at a minimum, a discussion of the pros and cons of the considered approaches and rationale for the selected rules.

27. **What is the level of certainty or confidence that the empirical analysis demonstrates that the component measures add value to the composite and that the aggregation and weighting rules are consistent with the quality construct?**

For example: Do the component measures fit the quality construct and add value? Are the objectives of parsimony and simplicity achieved while supporting the quality construct? Do analyses demonstrate the aggregation and weighting rules fit the quality construct and rationale?

- ☐ **High**
- ☐ **Moderate**
- ☐ **Low**
- ☐ **Insufficient**

28. **Briefly explain rationale for rating of EMPIRICAL ANALYSES TO SUPPORT COMPOSITE CONSTRUCTION**

Additional Recommendations

29. **If you have listed any concerns in this form, do you believe these concerns warrant further discussion by the multi-stakeholder Standing Committee? If so, please list those concerns below.**

Reviewer 9: Would like to discuss risk adjustment approach with the group. Unclear if the method is appropriate.

Reviewer 10: Tricky model to follow, but if accepted, does have favorable statistics. Important metric to get right.

Reviewer 11: The Standing Committee should consider whether the observed differences are meaningful, particularly in the middle of the distribution of scores and not just at the outlier tales. They also should ask for clarification on whether this is to be considered as an ACO measure (the level at which all testing was done) or a physician group measure (in MIPS, for example) in which case none of the testing results are relevant.

Developer Submission

1. Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall, less-than-optimal performance. Measures must be judged to meet all sub criteria to pass this criterion and be evaluated against the remaining criteria

---

Please separate added or updated information from the most recent measure evaluation within each question response in the Importance to Measure and Report: Evidence section. For example:

2021 Submission:
Updated evidence information here.

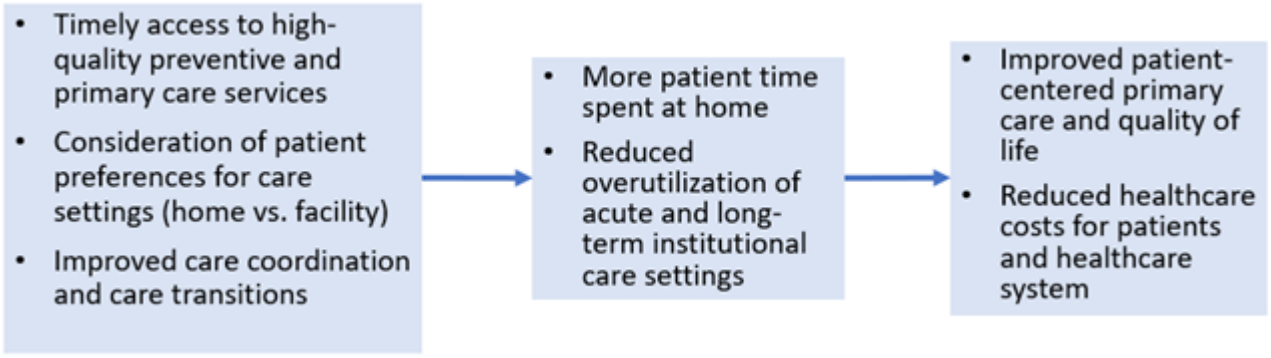
2018 Submission:
Evidence from the previous submission here.

1a. Evidence

1a.01. Provide a logic model.

Briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

[Response Begins]



[Response Ends]

1a.03. Provide empirical data demonstrating the relationship between the outcome (or PRO) and at least one healthcare structure, process, intervention, or service.

[Response Begins]
We conducted a literature review of relevant peer-reviewed publications by searching the PubMed Database. We performed a search that combines terms that broadly capture the concept of days at home: "days at home," "days in the home," "days spent at home," "time at home," "time in the home," "time spent at home," "community days," "home days," and "home time." Then we excluded articles for the following reasons:

- Published more than 10 years ago
- Non-English literature
- Pediatric literature
- Non-human study subjects
- Unrelated

Unrelated articles are those that use our search keywords as part of their protocol (for example, a clinical trial in which a subject spent three "days at home" in between drug administrations).

Most patients and families prefer spending time at home and in the community ("days at home") rather than in the hospital,[1-6] and more days at home are associated with both positive clinical outcomes and lower costs for patients and providers.[2,4,6-11] Poor care coordination can lead to unnecessary and preventable hospital visits for patients;[12,13] in contrast, improved care coordination and care transitions prevent unplanned hospital visits, leading to more days at home and high-quality timely care.[14,15] Given that patients with complex, chronic conditions often receive care from several clinicians and sites of care, this patient population may particularly benefit from improved care coordination.[6]

1. Barnato AE, Herndon MB, Anthony DL, et al. Are regional variations in end-of-life care intensity explained by patient preferences? A Study of the US Medicare Population. Medical care. 2007;45(5):386.
2. Fonarow GC, Liang L, Thomas L, et al. Assessment of Home-Time After Acute Ischemic Stroke in Medicare Beneficiaries. *Stroke.* 2016;47(3):836-842.
3. Higginson IJ, Sen-Gupta GJ. Place of care in advanced cancer: a qualitative systematic literature review of patient preferences. *J Palliat Med.* 2000;3(3):287-300.
4. Lee H, Shi SM, Kim DH. Home Time as a Patient-Centered Outcome in Administrative Claims Data. *Journal of the American Geriatrics Society.* 2019;67(2):347-351.

5. Leff B, Burton L, Guido S, Greenough WB, Steinwachs D, Burton JR. Home hospital program: a pilot study. *Journal of the American Geriatrics Society.* 1999;47(6):697-702.
6. McDermid I, Barber M, Dennis M, et al. Home-Time Is a Feasible and Valid Stroke Outcome Measure in National Datasets. *Stroke.* 2019;50(5):1282-1285.
7. Yu AYX, Fang J, Kapral MK. One-Year Home-Time and Mortality After Thrombolysis Compared With Nontreated Patients in a Propensity-Matched Analysis. *Stroke.* 2019:Strokeaha119026922.
8. Stienen MN, Smoll NR, Fung C, et al. Home-Time as a Surrogate Marker for Functional Outcome After Aneurysmal Subarachnoid Hemorrhage. *Stroke.* 2018;49(12):3081-3084.
9. Mishra NK, Shuaib A, Lyden P, et al. Home time is extended in patients with ischemic stroke who receive thrombolytic therapy: a validation study of home time as an outcome measure. *Stroke.* 2011;42(4):1046-1050.
10. Dewilde S, Annemans L, Peeters A, et al. The relationship between Home-time, quality of life and costs after ischemic stroke: the impact of the need for mobility aids, home, and car modifications on Home-time. *Disabil Rehabil.* 2020;42(3):419-425.
11. McCaffrey N, Agar M, Harlum J, Karnon J, Currow D, Eckermann S. Is home-based palliative care cost-effective? An economic evaluation of the Palliative Care Extended Packages at Home (PEACH) pilot. *BMJ Support Palliat Care.* 2013;3(4):431-435.
12. Brooks EM, Winship JM, Kuzel AJ. A "Behind-the-Scenes" Look at Interprofessional Care Coordination: How Person-Centered Care in Safety-Net Health System Complex Care Clinics Produce Better Outcomes. *Int J Integr Care.* 2020;20(2):5.
13. Valentijn PP, Schepman SM, Opheij W, Bruijnzeels MA. Understanding integrated care: a comprehensive conceptual framework based on the integrative functions of primary care. *Int J Integr Care.* 2013;13:e010.
14. Harrison JD, Auerbach AD, Quinn K, Kynoch E, Mourad M. Assessing the impact of nurse post-discharge telephone calls on 30-day hospital readmission rates. *J Gen Intern Med.* 2014;29(11):1519-1525.
15. Hoyer EH, Brotman DJ, Apfel A, et al. Improving Outcomes After Hospitalization: A Prospective Observational Multicenter Evaluation of Care Coordination Strategies for Reducing 30-Day Readmissions to Maryland Hospitals. *J Gen Intern Med.* 2018;33(5):621-627.

[Response Ends]

1b. Performance Gap

1b.01. Briefly explain the rationale for this measure.

Explain how the measure will improve the quality of care and list the benefits or improvements in quality envisioned by use of this measure.

[Response Begins]
This measure will directly benefit Medicare patients by reducing unnecessary hospitalizations and incentivizing days at home or in the community, which often reflect patient preferences. There are no competing measures for Medicare beneficiaries. The measure may help to incentivize care coordination between healthcare providers. The measure developers have worked to reduce the risk of the measure having negative unintended consequences on patients.
[Response Ends]

1b.02. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.

Include mean, std dev, min, max, interquartile range, and scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.

[Response Begins]
Testing of this version of the Days at Home measure was completed using Medicare FFS claims from calendar years 2017 and 2018 for patients of Shared Savings Program (SSP) Accountable Care Organizations (ACOs), comprising 610 ACOs with

1,154,779 patients meeting the measure inclusion criteria.

Among the 610 ACOs in Calendar Year 2018, the average Adjusted Days at Home were 330.4 days (standard deviation 3.7 days), ranging from 291.0 to 345.9. Notably, because of the use of risk adjustment, this performance gap does not simply reflect differences in case mix between ACOs but instead illuminates variation that is likely due to differences in ACO quality.
[Response Ends]

1b.03. If no or limited performance data on the measure as specified is reported above, then provide a summary of data from the literature that indicates opportunity for improvement or overall, less than optimal performance on the specific focus of measurement. Include citations.

[Response Begins]
Several studies demonstrate that time spent at home differs substantially among older patients, which suggests that there is potential for improving the quality of care and resulting days at home for the elderly population.[1,2]

While the majority of patients spend all or most days at home, one study noted that patients aged 65 and older with multiple chronic conditions spend fewer days at home, with patients having three or more chronic conditions spending an average of 12.3 fewer days at home (mean 336.6 days, SD 3.0) in a one-year period than do all patients ages 65 and older (mean 348.9 days, SD 1.7).[1]

1. Burke LG, Orav EJ, Zheng J, Jha AK. Healthy Days at home: A novel population-based outcome measure. *Healthcare (Amsterdam, Netherlands).* 2019:100378.
2. YNHHSC/CORE. Condition-Specific Excess Days in Acute Care Measures Updates and Specifications Report. 2019.

[Response Ends]

1b.04. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.

Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included. Include mean, std dev, min, max, interquartile range, and scores by decile. For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.

[Response Begins]
In the development and testing dataset we saw significant and substantial effects due to age and Medicaid dual-eligible status, which are included in the risk model to avoid disincentivizing care of patients at higher risk. We observed statistically insignificant or comparatively minor effects for select indicators of social risk, including local density of primary care physicians or specialists, local socioeconomic status indicator, urban residence, local percent of residents unmarried or living alone, and local density of hospital and nursing home beds.
[Response Ends]

1b.05. If no or limited data on disparities from the measure as specified is reported above, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in above.

[Response Begins]
There is inconsistent evidence linking socioeconomic status with days at home for older patients. Some studies linking poverty, female sex, age, and/or dual-eligible status with fewer days at home;[1,2] however, another study found no significant difference based on age, sex, or race/ethnicity.[3]

1. Burke LG, Orav EJ, Zheng J, Jha AK. Healthy Days at home: A novel population-based outcome measure. Healthcare (Amsterdam, Netherlands). 2019:100378.

2. Medicare Payment Advisory Commission. Report to the Congress: Medicare and the Health Care Delivery System. 2018.
3. Gill TM, Gahbauer EA, Leo-Summers L, Murphy TE, Han L. Days Spent at Home in the Last Six Months of Life Among Community-Living Older Persons. The American journal of medicine. 2019;132(2):234-239.

[Response Ends]

2. Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. Measures must be judged to meet the sub criteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.

---

sp.01. Provide the measure title.

Measure titles should be concise yet convey who and what is being measured (see What Good Looks Like).

[Response Begins]
Days at Home for Patients with Complex, Chronic Conditions
[Response Ends]

sp.02. Provide a brief description of the measure.

Including type of score, measure focus, target population, timeframe, (e.g., Percentage of adult patients aged 18-75 years receiving one or more HbA1c tests per year).

[Response Begins]
This is a provider group-level measure of days at home or in community settings (that is, not in acute care such as inpatient hospital or emergent care settings or post-acute settings such as Skilled Nursing Facilities (SNFs)) among adult (age 18 years or older) Medicare FFS beneficiaries with complex, chronic conditions who are aligned to participating provider groups. The measure includes risk adjustment for differences in patient mix across provider groups, with an adjustment based on patients' risk of death. An additional adjustment that accounts for patients' risk of transitioning to a long-term nursing home is also applied to encourage home- and community-based care in alignment with CMS's policy goals. A higher risk-adjusted score indicates better performance.
[Response Ends]

sp.04. Check all the clinical condition/topic areas that apply to your measure, below.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Surgery: General*

[Response Begins]
Other (specify)
Non-specific chronic disease
[Response Ends]

sp.05. Check all the non-condition specific measure domain areas that apply to your measure, below.

[Response Begins]
Primary Prevention
[Response Ends]

sp.06. Select one or more target population categories.

Select only those target populations which can be stratified in the reporting of the measure's result.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Populations at Risk: Populations at Risk*

[Response Begins]
Adults (Age >= 18)
[Response Ends]

sp.07. Select the levels of analysis that apply to your measure.

Check ONLY the levels of analysis for which the measure is SPECIFIED and TESTED.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Clinician: Clinician*
- *Population: Population*

[Response Begins]
Accountable Care Organization
[Response Ends]

sp.08. Indicate the care settings that apply to your measure.

Check ONLY the settings for which the measure is SPECIFIED and TESTED.
[Response Begins]
Inpatient/Hospital
Post-Acute Care
[Response Ends]

sp.09. Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials.

Do not enter a URL linking to a home page or to general information. If no URL is available, indicate "none available".

[Response Begins]
None available
[Response Ends]

sp.11. Attach the data dictionary, code table, or value sets (and risk model codes and coefficients when applicable). Excel formats (.xlsx or .csv) are preferred.

Attach an excel or csv file; if this poses an issue, contact staff. Provide descriptors for any codes. Use one file with multiple worksheets, if needed.
[Response Begins]
Available in attached Excel or csv file
[Response Ends]

Attachment: DaysAtHome_CodeSet_073021.xlsx

For the question below: state the outcome being measured. Calculation of the risk-adjusted outcome should be described in sp.22.

sp.12. State the numerator.

Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome).

DO NOT include the rationale for the measure.

[Response Begins]
The outcome measured for each eligible beneficiary is days spent "at home," adjusted for clinical and social risk factors, risk of death, and risk of transitioning to a long-term nursing home.
[Response Ends]

For the question below: describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in sp.22.

sp.13. Provide details needed to calculate the numerator.

All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets.

Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.

[Response Begins]
Days at home are defined as those days when a beneficiary is alive and not in care.

A "day in care" is defined as any eligible patient day on which a patient receives care in one (or more) of the following specified care settings: inpatient acute and post-acute facilities (short-term acute care hospitals, critical access hospitals (CAHs), inpatient rehabilitation facilities (IRFs), inpatient psychiatric facilities (IPFs), long-term care hospitals (LTCHs), and SNFs); emergency departments (ED); and observation stays. There are two exceptions:

1. A patient is always considered "at home" if they are enrolled in hospice, even if they receive care in settings normally counted as "days in care" (that is, if a patient enrolled in hospice is receiving care in an inpatient setting that will not count as a day in care)

Rationale: to promote effective and appropriate care for terminally ill patients

2. Hospital admissions for childbirth, miscarriage, or termination are not counted as "days in care"

Rationale: obstetric admissions may not indicate care quality; counting these admissions may create an inappropriate incentive to keep pregnant patients out of the hospital.

A "day at home" is defined as any eligible day that is not considered a "day in care" based on the above definition. "Eligible days" are all days in the measurement year that the beneficiary is alive.

Care in settings not listed above (including outpatient visits and procedures, hospice, residential psychiatric and substance abuse facilities, assisted living facilities and group homes, and home health and telehealth services) are not considered "days in care" in this measure; rather, they are treated as "days at home."

Finally, days spent in a long-term or residential nursing home (except for SNF care) are not counted as "days in care" by

this definition. However, as discussed in the "Measure Scoring" section, this measure includes an adjustment that accounts for patients' risk of transitioning to a long-term nursing home, to encourage home- and community-based care in alignment with CMS's policy goals.
[Response Ends]

For the question below: state the target population for the outcome. Calculation of the risk-adjusted outcome should be described in sp.22.

sp.14. State the denominator.

Brief, narrative description of the target population being measured.

[Response Begins]
Eligible beneficiaries aligned to participating provider groups.
[Response Ends]

For the question below: describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in sp.22.

sp.15. Provide details needed to calculate the denominator.

All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets.

Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.

[Response Begins]
Eligible beneficiaries are:

- Adult (age 18 or older);
- Alive as of the first day of the performance year;
- Continuously enrolled in Medicare FFS parts A and B during the full performance year (up to date of death among patients who died) and one full year prior; and
- Have an average Hierarchical Condition Category (HCC) composite risk score >= 2.0 in the year prior to the performance year.

The measure includes eligible beneficiaries who are aligned to a participating provider group as determined by the relevant program or model.

[Response Ends]

sp.16. Describe the denominator exclusions.

Brief narrative description of exclusions from the target population.

[Response Begins]
Not applicable. There are currently no denominator exclusions or exceptions for the measure. All patients meeting the denominator inclusion criteria are included.
[Response Ends]

sp.17. Provide details needed to calculate the denominator exclusions.

All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.

[Response Begins]
Not applicable. There are no denominator exclusions.
[Response Ends]

sp.18. Provide all information required to stratify the measure results, if necessary.

Include the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically adjusted version of the measure when appropriate. Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format in the Data Dictionary field.

[Response Begins]
The measure uses risk adjustment and is not stratified.

The full list of covariates and associated codes may be found in the included "DaysAtHome_CodeSet_073021.xlsx" workbook.

The Days at Home measure combines information from three risk-adjusted statistical models: one for “adjusted days in care,” one for mortality risk, and another for risk of transition to nursing home during the measurement year.

To select risk variables for the Days at Home measure, we started with the set of risk factors used in a fully-developed related measure, the Risk-Standardized Unplanned Acute Admission Rate for Patients with Multiple Chronic Conditions (UAMCC). The measure is currently used in the Medicare Shared Savings Program (SSP) and will be added to the CMS Merit-based Incentive Payment System (MIPS) with only minor differences in specifications.

The UAMCC patient population is clinically similar to the Days at Home measure cohort of patients with complex, chronic conditions and has a comparable diversity of severity of illness, socioeconomic status, and geographical access to care. There is a difference in eligibility based on age for the measure cohorts: patients who are age 18 or older are eligible for Days at Home, but patients are required to be 65 or older for the UAMCC measure. The UAMCC measure assesses the number of admissions within the performance year and thus the risk variables selected and grouped by the UAMCC measure (such as patient frailty) were strongly predictive of patients’ likelihood of having unplanned hospital visits.

The Days at Home measure uses the candidate clinical risk variables of the UAMCC measure (defined by Condition Categories (CCs)) for each of the three component models. These consist of 37 clinical comorbidity variables, nine chronic conditions, six variables related to frailty or disability (for example, walking aids, durable medical equipment, and reason for current Medicare entitlement), and age. Two variables (RF 25, pancreatic disease - due to very low prevalence, and reason for Medicare entitlement - due to strong association with age) were excluded, leaving 51 clinical risk variables. These same clinical risk variables are used in each of the three component risk models of the Days at Home measure.

In addition, beneficiaries’ Medicaid dual-eligibility status is included as a risk variable in the Days in Care and Nursing Home Transition component risk models. Dual-eligible patients have structurally different means to pay for various services and may have fewer resources and social supports to remain safely at home compared to Medicare-only beneficiaries. Adjusting for dual-eligible status acknowledges that provider groups may have limited ability to address this risk factor and reduces any incentive for provider groups to select against dual-eligible patients. Dual-eligible status is not included in the Mortality risk model.

The final list of risk variables is shown in the table below:

Variable name	Risk variable definition
RF1	Dialysis status
RF2	Respiratory failure
RF3	Liver disease
RF4	Pneumonia
RF5	Septicemia/shock
RF6	Marked disability/frailty
RF7	Pleural effusion/pneumothorax
RF8	Hematologic/al diseases
RF9	Advanced cancer
RF10	Infectious and immune disorders
RF11	Severe cognitive impairment
RF12	Major organ transplant status
RF13	Pulmonary heart disease
RF14	Cardiomyopathy
RF15	Gastrointestinal disease
RF16	Bone/joint/muscle infections/necrosis
RF17	Iron deficiency anemia
RF18	Diabetes with complications
RF19	Ischemic heart disease except AMI
RF20	Other lung disorders
RF21	Vascular or circulatory disease
RF22	Other significant endocrine disorders
RF23	Other disabilities and paralysis
RF24	Substance abuse
RF26	Other neurologic disorders
RF27	Arrhythmia (except atrial fibrillation)
RF28	Hypertension
RF29	Hip or vertebral fracture
RF30	Lower-risk cardiovascular disease
RF31	Cerebrovascular disease
Variable name	**Risk variable definition**
RF32	Other malignancy
RF33	Morbid obesity
RF34	Urinary disorders
RF36	Psychiatric disorders other than depression
RF37	Age
RF38	Acute Myocardial Infarction (AMI)

Variable name	Risk variable definition
RF39	Alzheimer's & Related Disorders
RF40	Atrial Fibrillation
RF41	Chronic Kidney Disease
RF42	Chronic Obstructive Pulmonary Disease (COPD)/Asthma
RF43	Depression
RF44	Heart Failure
RF45	Stroke/Transient Ischemic Attack (TIA)
RF46	Other organ transplant
RF47	Precerebral arterial occlusion and transient cerebral ischemia
RF48	Diabetic retinopathy
RF49	Walking aids
RF50	Wheelchairs
RF51	Hospital bed
RF52	Lifts
RF53	Oxygen
RF54	Dual eligibility status (Days in Care and Nursing Home Transition models only)

[Response Ends]

sp.19. Select the risk adjustment type.

Select type. Provide specifications for risk stratification and/or risk models in the Scientific Acceptability section.
[Response Begins]
Statistical risk model
[Response Ends]

sp.20. Select the most relevant type of score.

Attachment: If available, please provide a sample report.
[Response Begins]
Continuous variable, e.g., average
[Response Ends]

sp.21. Select the appropriate interpretation of the measure score.

Classifies interpretation of score according to whether better quality or resource use is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score
[Response Begins]

Better quality = Higher score
[Response Ends]

sp.22. Diagram or describe the calculation of the measure score as an ordered sequence of steps.

Identify the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period of data, aggregating data; risk adjustment; etc.

[Response Begins]
The final score (that is, the "provider group-level adjusted days at home") is the risk factor-adjusted, mortality-adjusted, nursing home transition-adjusted days at home, averaged over all patients within a provider group.

At a high level, the measure result (adjusted days at home) is calculated based on three risk-adjusted statistical models. We use the first model to calculate "excess days in care" for each patient, which represents the risk-adjusted days in acute care settings or SNFs among days alive in the year. Two additional risk-adjusted models are used to calculate the risk of mortality and risk of transition to nursing home for each patient. Finally, "excess days in care" are updated based on risk of death and risk of transition to nursing home care and then averaged across each provider group to produce the final measure scores. The details of each of these steps are described below.

First, "excess days in care" for each patient are modeled using a hierarchical negative binomial regression with an offset for days alive. "Excess days in care" is defined as predicted minus expected days in care, where "predicted" includes clinical and social risk adjustment, survival offset, and a provider group-specific effect, and "expected" includes only clinical and social risk adjustment and survival offset. More "excess days in care" indicate a patient spent more time in care than expected due to their provider group's performance.

Second, mortality is modeled using a hierarchical logistic regression model with adjustment for the patient case-mix, to calculate a standardized mortality ratio (SMR) at the patient level. SMR is defined as predicted divided by expected risk of death, where "predicted" includes clinical risk adjustment and a provider group-specific effect, and "expected" includes only clinical risk adjustment. A high SMR indicates a patient at greater-than-expected risk of death due to their provider group's performance.

Third, a patient's risk of transitioning to a residential nursing home is modeled using a hierarchical logistic regression model with adjustment for patient case-mix and Medicaid dual-eligibility status, to calculate a standardized "nursing home ratio" (NHR) which is scaled to have the same mean and standard deviation as the SMR. NHR is defined as predicted divided by expected risk of transitioning to a nursing home during the performance year, where "predicted" includes clinical and social risk adjustment and a provider group-specific effect, and "expected" includes only clinical and social risk adjustment. A higher NHR indicates a patient at greater-than-expected risk of transitioning to a nursing home due to their provider group's performance.

For the mortality adjustment for each patient, "excess days in care" is multiplied by SMR (if excess days >= 0) or divided by SMR (if excess days < 0), such that a greater SMR results in an absolute increase of "excess days in care" (that is, provider groups are rewarded for lower mortality than expected and penalized for greater mortality than expected given their case mix). Similarly, for the nursing home adjustment for each patient, "excess days in care" is multiplied by [0.5*NHR] (if excess days < 0) or divided by [0.5*NHR] (if excess days >= 0) so that provider groups are rewarded for lower rates of transition to the nursing home than expected given their case mix.

The SMR and NHR adjustments are combined additively to give a "mortality- and nursing home transition risk-adjusted excess days in care," which is subtracted from the patient-level national average of days alive, resulting in a risk-, mortality-, and nursing home transition-adjusted measure of "days at home" for that patient.

Finally, the adjusted days at home are averaged over all patients of each provider group to summarize the provider group's measure performance as the "provider group-level adjusted days at home."
[Response Ends]

sp.25. If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.

[Response Begins]
Not applicable.
[Response Ends]

sp.28. Select only the data sources for which the measure is specified.

[Response Begins]
Claims
[Response Ends]

sp.29. Identify the specific data source or data collection instrument.

For example, provide the name of the database, clinical registry, collection instrument, etc., and describe how data are collected.

[Response Begins]
Medicare inpatient claims

Medicare outpatient claims

Medicare SNF claims

Medicare beneficiary enrollment data

provider group-Beneficiary alignment/attribution file
[Response Ends]

sp.30. Provide the data collection instrument.

[Response Begins]
No data collection instrument provided
[Response Ends]

Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate fields in the Scientific Acceptability sections of the Measure Submission Form.

• Measures must be tested for all the data sources and levels of analyses that are specified. If there is more than one set of data specifications or more than one level of analysis, contact NQF staff about how to present all the testing information in one form.
• All required sections must be completed.
• For composites with outcome and resource use measures, Questions 2b.23-2b.37 (Risk Adjustment) also must be completed.
• If specified for multiple data sources/sets of specifications (e.g., claims and EHRs), Questions 2b.11-2b.13 also must be completed.
• An appendix for supplemental materials may be submitted (see Question 1 in the Additional section), but there is no guarantee it will be reviewed.
• Contact NQF staff with any questions. Check for resources at the Submitting Standards webpage.
• For information on the most updated guidance on how to address social risk factors variables and testing in this form refer to the release notes for the 2021 Measure Evaluation Criteria and Guidance.

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.
2a. Reliability testing demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For instrument-based measures (including PRO-PMs) and composite performance measures, reliability should be demonstrated for the computed performance score.
2b1. Validity testing demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For instrument-based measures (including PRO-PMs) and composite performance measures, validity should be demonstrated for the computed performance score.
2b2. Exclusions are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure;
AND
If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).
2b3. For outcome measures and other measures when indicated (e.g., resource use):
• an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at start of care; 14,15 and has demonstrated adequate discrimination and calibration
OR
• rationale/data support no risk adjustment/ stratification.
2b4. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful 16 differences in performance;
OR
there is evidence of overall less-than-optimal performance.
2b5. If multiple data sources/methods are specified, there is demonstration they produce comparable results.
2b6. Analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias.
2c. For composite performance measures, empirical analyses support the composite construction approach and demonstrate that:
2c1. the component measures fit the quality construct and add value to the overall composite while achieving the related objective of parsimony to the extent possible; and
2c2. the aggregation and weighting rules are consistent with the quality construct and rationale while achieving the related objective of simplicity to the extent possible.
(if not conducted or results not adequate, justification must be submitted and accepted)

Definitions
Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).
Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. The degree of consensus and any areas of disagreement must be provided/discussed.

Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.
Risk factors that influence outcomes should not be specified as exclusions.
With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v.$5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

Please separate added or updated information from the most recent measure evaluation within each question response in the Importance to Scientific Acceptability sections. For example:

2021 Submission:
Updated testing information here.

2018 Submission:
Testing from the previous submission here.

2a. Reliability

2a.01. Select only the data sources for which the measure is tested.

[Response Begins]
Claims
[Response Ends]

2a.02. If an existing dataset was used, identify the specific dataset.

The dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured, e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry).

[Response Begins]
- Medicare inpatient claims (2017-18)
- Medicare outpatient claims (2017-18)
- Medicare skilled nursing claims (2017-18)
- Medicare beneficiary enrollment data (2017-18)
- Shared Savings Program (SSP) Accountable Care Organization (ACO) assignment file (2017)
- Nursing Home Compare provider info file (2018)
- US Census Bureau American Community Survey 5-year estimates (2018)
- USDA Economic Research Service (2013)
- HRSA Area Health Resources File (AHRF) (2018)

[Response Ends]

2a.03. Provide the dates of the data used in testing.

Use the following format: "MM-DD-YYYY - MM-DD-YYYY"

[Response Begins]

01-01-2017 - 12-31-2018
[Response Ends]

2a.04. Select the levels of analysis for which the measure is tested.

Testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Clinician: Clinician*
- *Population: Population*

[Response Begins]
Accountable Care Organization
[Response Ends]

2a.05. List the measured entities included in the testing and analysis (by level of analysis and data source).

Identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample.

[Response Begins]
The development and testing dataset consisted of 610 Medicare Shared Savings Program ACOs, restricted to patients meeting the measure inclusion criteria. The average ACO included 1,893.4 eligible patients, ranging from 56 to 13,426; the interquartile range was 793 to 2,255 patients.
[Response Ends]

2a.06. Identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis), separated by level of analysis and data source; if a sample was used, describe how patients were selected for inclusion in the sample.

If there is a minimum case count used for testing, that minimum must be reflected in the specifications.

[Response Begins]
The development dataset includes 1,154,779 patients meeting the measure inclusion criteria.
Characteristics of patients with complex, chronic conditions attributed to ACOs

Characteristic	Total (n)	Percent (%)
Total Patients	1,154,779	100.00
Age Distribution	--	--
18 to <55	74,680	6.47
55 to <65	109,618	9.49
65 to <75	359,836	31.16
75 to <85	373,301	32.33
85 to <95	218,912	18.96
>=95	18,432	1.60
Female sex	621,937	53.86
Race Distribution	--	--
White	980,363	84.90
Black	112,383	9.73
Asian	16,162	1.40

Characteristic	Total (n)	Percent (%)
Hispanic	21,255	1.84
Other	24,616	2.13
Patients with Average HCC Risk Score ≥3.0	468,173	40.54
Long-Term Institution (LTI) Status (nursing home residence for ≥90 days) in Calendar Year (CY) 2017	52,403	4.54
Any Dual Eligibility in CY 2017	263,114	22.78
Skilled Nursing Facility Care in CY 2017	149,737	12.97
Hospice Care in CY 2017	17,306	1.50

[Response Ends]

2a.07. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing.

[Response Begins]
Not applicable.
[Response Ends]

2a.08. List the social risk factors that were available and analyzed.

For example, patient-reported data (e.g., income, education, language), proxy variables when social risk data are not collected from each patient (e.g., census tract), or patient community characteristics (e.g., percent vacant housing, crime rate) which do not have to be a proxy for patient-level data.

[Response Begins]
Medicaid dual-eligibility status; local Agency for Healthcare Research and Quality (AHRQ) Socioeconomic Status (SES) Index; urban residence; local primary care provider density; local physician specialist density; local density of hospital beds; local density of nursing home beds; local percent of residents never married; and local percent of residents living alone.
[Response Ends]

Note: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a.07 check patient or encounter-level data; in 2a.08 enter "see validity testing section of data elements"; and enter "N/A" for 2a.09 and 2a.10.

2a. Reliability

2a.09. Select the level of reliability testing conducted.

Choose one or both levels.
[Response Begins]
Accountable Entity Level (e.g., signal-to-noise analysis)
[Response Ends]

2a.10. For each level of reliability testing checked above, describe the method of reliability testing and what it tests.

Describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used.

[Response Begins]

We tested the reliability of the Days at Home measure using a split-half methodology: We randomly split the cohort in half, calculated the measure separately for each half, and compared the results between ACO scores in each half. We specifically calculated the level intraclass correlation coefficient (ICC) for the final Days at Home outcome metric.[1]

1. Shrout PE, Fleiss JL. Intraclass correlations uses in assessing rater reliability. Psychol Bull 1979;86:420-8.

[Response Ends]

2a.11. For each level of reliability testing checked above, what were the statistical results from reliability testing?

For example, provide the percent agreement and kappa for the critical data elements, or distribution of reliability statistics from a signal-to-noise analysis. For score-level reliability testing, when using a signal-to-noise analysis, more than just one overall statistic should be reported (i.e., to demonstrate variation in reliability across providers). If a particular method yields only one statistic, this should be explained. In addition, reporting of results stratified by sample size is preferred (pg. 18, NQF Measure Evaluation Criteria).

[Response Begins]
Using the method, we calculated an ICC of 0.8326 for ACO measure scores.
[Response Ends]

2a.12. Interpret the results, in terms of how they demonstrate reliability.

(In other words, what do the results mean and what are the norms for the test conducted?)

[Response Begins]
Using this method, we calculated an ICC of 0.8326, demonstrating very good agreement in ACO score between the two samples and indicating that the measure is consistent and not greatly sensitive to chance variations in the underlying data.
[Response Ends]

2b. Validity

2b.01. Select the level of validity testing that was conducted.

[Response Begins]
Empirical validity testing
Systematic assessment of face validity of performance measure score as an indicator of quality or resource use (i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance)
[Response Ends]

2b.02. For each level of testing checked above, describe the method of validity testing and what it tests.

Describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used.

[Response Begins]
Empirical Validity Testing:

To empirically evaluate the measure's construct validity, we correlated performance on the Days at Home measure among 610 SSP ACOs in CY 2018 to their performance on select other quality measures used by the SSP program in the same period. We identified the candidate measures as those that might capture quality constructs related to the Days at Home measure including care coordination, avoidance of acute care, and prevention of health decline. The measures and our expectation for their relationship to the Days at Home measure are shown in Table 1 below (with inverse correlations expected for measures in which a low score indicates better performance):

Table 1: Construct Validity Comparison Measures & Expected Relationship

ID	Description	SSP Domain	Source	Expected Relationship
ACO-8	Risk-standardized all condition readmission	Care coordination/ patient safety	Claims	Moderate inverse correlation
ACO-38	All-cause unplanned admissions for patients w/ multiple chronic conditions	Care coordination/ patient safety	Claims	Low inverse correlation
ACO-13	Falls: screening for future fall risk (CARE-2)	Care coordination/ patient safety	Web input	Low inverse correlation
ACO-43	Ambulatory sensitive condition acute composite (AHRQ PQI-91)	Care coordination/ patient safety	Claims	Low inverse correlation
ACO-35	SNF 30-day All-cause Readmission	n/a	Claims	Low inverse correlation
ACO-1	Timely care, appointments, and info	Patient/caregiver experience	Survey	Low positive correlation

Face Validity Assessment:

CORE assessed the measure's face validity through engagement with a large multi-disciplinary Technical Expert Panel (TEP) representing a variety of stakeholders. CORE met with the TEP on three occasions through the measure development process (in June 2020, September 2020, and April 2021) to discuss the measure specifications and gain feedback. Following the third meeting, CORE asked members of the TEP to respond to a survey assessing the face validity of the Days at Home measure specifications. Each member was asked if they "strongly agree", "agree," "somewhat agree," "somewhat disagree," "disagree," or "strongly disagree" with the statement, "The Days at Home measure, as specified, can be used to distinguish between better or worse performance at ACOs or provider groups," and to provide a rationale for their response.

[Response Ends]

2b.03. Provide the statistical results from validity testing.

Examples may include correlations or t-test results.

[Response Begins]
Empirical Validity Testing
The correlation of ACO performance between Days at Home and each comparison measure are shown in Table 2 below:

Table 2: Construct validity comparison results

ID	Description	Pearson r (p-value)	Conclusion
ACO-38	All-cause unplanned admissions for patients w/ multiple chronic conditions	-0.549 (<.0001)	• Inverse correlation
ACO-8	Risk-standardized all condition readmission	-0.182 (<.0001)	• Inverse correlation
ACO-35	SNF 30-day All-cause Readmission	-0.106 (.013)	• Inverse correlation
ACO-43	Ambulatory sensitive condition acute composite (AHRQ PQI-91)	-0.376 (<.0001)	• Inverse correlation
ACO-13	Falls: screening for future fall risk (CARE-2)	0.048 (.27)	• No sig. correlation

ID	Description	Pearson r (p-value)	Conclusion
ACO-1	Timely care, appointments, and info	-0.126 (0.003)	• Inverse correlation

Face Validity Assessment:

Of the 21 TEP members, 19 responded to the survey, all of whom agreed with the given statement. Specifically, 2 members indicated "strongly agree," 15 indicated "agree," and 2 indicated "somewhat agree." No TEP members disagreed that the specified Days at Home measure can be used to distinguish performance among provider groups.

Several TEP members agreed that the measure reflects an important outcome and would incentivize care coordination and home-based care, noting this is currently a gap in measurement. They noted that to perform well on the measure, providers will have to implement practices and interventions across multiple aspects or systems of care to prevent acute care use and support days at home, which will improve the patient experience as well as reduce overall spending.

Several TEP members also expressed support for the measure methodology (particularly the definition to count "days at home") and risk adjustment approach, suggesting that the measure fairly captures the outcome of interest.

The two TEP members who "somewhat agreed" both felt the measure did not adequately capture time spent in nursing homes or long-term institutions and that the measure should more strongly reflect those patients' loss of community living. However, they appreciated the extra dimension of this measure in capturing duration of care (over simply counting admissions) and reflected that meaningful variation over the population of interest would make the measure generally useful.

[Response Ends]

2b.04. Provide your interpretation of the results in terms of demonstrating validity. (i.e., what do the results mean and what are the norms for the test conducted?)

[Response Begins]
Empirical Validity Testing:

We observed a high inverse correlation with ACO-38 (All-cause unplanned admissions for patients with multiple chronic conditions), which is consistent with that measure's conceptually related cohort and outcome. We observed modest but statistically significant correlations with ACO-8 (Risk-standardized all condition readmission), ACO-35 (SNF 30-day all-cause readmission) and ACO-43 (Ambulatory sensitive condition acute composite), which are conceptually related but have different focuses in cohort and outcome definition.

We saw no significant correlation with ACO-13 (screening for future fall risk), and unexpected inverse correlation with ACO-1 (patient experience – timely care appointments and information); it is notable that ACO-1 is a patient experience survey measure with a smaller sample size and no clinical risk adjustment. Specifically, we note that a narrowly defined process or experience measure that only represents a fraction of included patients is unlikely to have a substantial relationship with broader measures of outcomes and utilization.

Face Validity Assessment

The members of the TEP broadly supported Days at Home as a valid measure of provider performance following a rigorous and structured process of engagement.
[Response Ends]

2b.05. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified.

Describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided in Importance to Measure and Report: Gap in Care/Disparities.

[Response Begins]
We did not conduct a statistical test for significant differences in performance.

To assess clinically and practically meaningful differences, we examined the range in distribution of ACO-level scores, including the interquartile range, the 5th-95th percentile range, and the total range among the 610 SSP ACOs.
[Response Ends]

2b.06. Describe the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities.

Examples may include number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined.

[Response Begins]
We found an interquartile range of 3.0 adjusted days at home (329.1 – 332.1), a range from 5th to 95th percentile of 8.8 adjusted days at home (325.7 – 334.5), and a total range from the lowest to highest ACO performance of 54.9 adjusted days at home (291.0 – 345.9).
[Response Ends]

2b.07. Provide your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities.

In other words, what do the results mean in terms of statistical and meaningful differences?

[Response Begins]
Each ACO's adjusted days at home indicates the average across all its patients, and so differences between two ACOs' scores should be interpreted as the aggregated difference in performance across all patients. It should be noted that within each ACO's patient population, there is often substantial variation in days spent at home, with some spending all or nearly all days at home and others with relatively few days at home.

We found substantial differences in performance between high and low performers on the measure. In comparing the extreme ends of the ACO performance spectrum (very low to very high performing ACOs), there is a stark contrast: patients of a 5th percentile ACO can expect to spend 8.8 fewer days at home than those at a 95th percentile ACO, even accounting for the differential case mix.

However, even small differences in Days at Home between ACOs within the interquartile range can be considered clinically meaningful. To illustrate, a typical patient of an ACO at the first quartile of performance can expect to have 3.0 fewer days at home (that is, 3.0 more days in care) than with an ACO at the third quartile of performance. Since the average patient across all ACOs spends approximately 11 days in care, a difference of 3 days should not be considered trivial. With ACOs ranging in size from several hundred to several thousand patients, these per-patient differences amount to a large difference in overall adjusted days at home. This is important from a total cost perspective, while also being meaningful to patients.
[Response Ends]

2b.08. Describe the method of testing conducted to identify the extent and distribution of missing data (or non-response) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders). Include how the specified handling of missing data minimizes bias.

Describe the steps—do not just name a method; what statistical analysis was used.

[Response Begins]
The Days at Home measure uses routinely submitted claims data to identify the measure's cohort, risk-adjustment variables, and outcome. We utilized only those data elements from the claims that have both face validity and reliability. To ensure that we use data elements that are reliable, we avoid the use of fields that are not coded consistently across hospitals or providers. Additionally, CMS has in place several auditing programs used to assess overall claims code accuracy, to ensure appropriate billing, and for overpayment recoupment. CMS routinely conducts data analysis to identify potential problem areas and detect fraud, and audits important data fields used in our measures. Using claims data imposes no costs on providers and eliminates provider burden, which is important since providers have limited time to dedicate to reporting. Prior research has demonstrated that administrative claims can be used to assess the quality of care delivered by individual or small clinician groups (for example, use of claims-based Hospital-Wide Readmission Measure in the Value Modifier).[1] These models have demonstrated consistent performance across years of claims data. No additional analyses were conducted as part of measure development.

1. YNHHSC/CORE. Methodology Report, Measure Testing Report, and Risk Adjustment Report: Clinician and Clinician Group Risk-Standardized Hospital Admission Rates for Patients with Multiple Chronic Conditions. 2019.

[Response Ends]

2b.09. Provide the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data.

For example, provide results of sensitivity analysis of the effect of various rules for missing data/non-response. If no empirical sensitivity analysis was conducted, identify the approaches for handling missing data that were considered and benefits and drawbacks of each).

[Response Begins]
Not applicable.
[Response Ends]

2b.10. Provide your interpretation of the results, in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and non-responders), and how the specified handling of missing data minimizes bias.

In other words, what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis was conducted, justify the selected approach for missing data.

[Response Begins]
Not applicable.
[Response Ends]

Note: This item is directed to measures that are risk-adjusted (with or without social risk factors) OR to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eCQMs). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator). Comparability is not required when comparing performance scores with and without social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.

2b.11. Indicate whether there is more than one set of specifications for this measure.

[Response Begins]
No, there is only one set of specifications for this measure

[Response Ends]

2b.12. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications.

Describe the steps—do not just name a method. Indicate what statistical analysis was used.

[Response Begins]
[Response Ends]

2b.13. Provide the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications.

Examples may include correlation, and/or rank order.

[Response Begins]
[Response Ends]

2b.14. Provide your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications.

In other words, what do the results mean and what are the norms for the test conducted.

[Response Begins]
[Response Ends]

2b.15. Indicate whether the measure uses exclusions.

[Response Begins]
N/A or no exclusions
[Response Ends]

2b.16. Describe the method of testing exclusions and what was tested.

Describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used?

[Response Begins]
Not applicable
[Response Ends]

2b.17. Provide the statistical results from testing exclusions.

Include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores.

[Response Begins]
Not applicable
[Response Ends]

2b.18. Provide your interpretation of the results, in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results.

In other words, the value outweighs the burden of increased data collection and analysis. Note: If patient preference is an exclusion, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion.

[Response Begins]
Not applicable
[Response Ends]

2b.19. Check all methods used to address risk factors.

[Response Begins]
Statistical risk model with risk factors (specify number of risk factors)
Statistical model with 52 risk factors
[Response Ends]

2b.20. If using statistical risk models, provide detailed risk model specifications, including the risk model method, risk factors, risk factor data sources, coefficients, equations, codes with descriptors, and definitions.

[Response Begins]
Three models are used to construct the final risk adjusted days at home measure: a count model for days at home, a logistic regression model for mortality, and a logistic regression model for transition to a nursing home. Each model includes a hierarchical "random effect" term to reflect the case mix of patients within each provider group (such as an ACO). Results from the three models are used to calculate a final risk-, mortality-, and nursing home transition-adjusted, or more simply "adjusted" days at home measure.

Days in Care Model

We model days in care rather than days at home because days in care is distributed as a typical count variable. To model days in care, we use a hierarchical negative binomial regression model. The model includes adjustment of the risk factors to account for patient case-mix and provider group-specific random effects to account for the patient mix within provider groups. It also includes an offset for the number of days the patients survived in the performance year for adjustment.

Specifically, we let Y_ij denote the number of days in care in the year experienced by i-th patient enrolled at the j-th provider group with risk factors

$$X_{ij,1}, \dots, X_{ij,p}$$

and the exposure-time s*ij (that is, the number of days alive from 1 up to 365 if the patient died or set to 365 if patient did not die during the performance year), p is the number of risk factors in the model. The days in care Y*ij is modeled as negative binomial distributed with mean

$$\mu_{ij}$$

and variance

$$\mu_{ij} + k\mu_{ij}^2$$

where k is the scale parameter. The hierarchical negative binomial regression model equation is

$$\log(\mu_{ij}) = \beta_0 + \beta_1 X_{ij,1} + \cdots + \beta_p X_{ij,p} + \log(s_{ij}) + z_j$$

where z_j is the provider group-specific random effect that is normally distributed with mean 0 and variance

$$\sigma_z^2$$

For each patient, the predicted days in care is calculated as:

$$P_{ij} = \exp\left(\beta_0 + \beta_1 X_{ij,1} + \cdots + \beta_p X_{ij,p} + \log(s_{ij}) + z_j\right)$$

the expected number of days in care is calculated as:

$$E_{ij} = \exp\left(\beta_0 + \beta_1 X_{ij,1} + \cdots + \beta_p X_{ij,p} + log(s_{ij})\right)$$

and the excess days in care (EDIC) is the difference of "predicted" minus "expected" days in care, calculated as:

$$EDIC_{ij} = P_{ij} - E_{ij}$$

$$EDIC_{ij} > 0$$

indicates that the patient spent more days in care due to their provider group's performance than expected at a provider group of average quality, while

$$EDIC_{ij} < 0$$

indicates the patient spent fewer days in care due to their provider group's performance than expected.

Mortality Model

For mortality model, we used a hierarchical logistic regression model.

We let M_ij denote whether the i-th patient enrolled at the j-th provider group died during the performance year with risk factors

$$X_{ij,1}, \ldots, X_{ij,p}$$

, and p is the number of risk factors in the model. The model includes adjustment of the risk factors and provider group-specific random effects to account the within-group variation for mortality. The hierarchical logistic regression model equation is:

$$\log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \alpha_0 + \alpha_1 X_{ij,1} + \cdots + \alpha_p X_{ij,p} + w_j$$

where p_ij is the Bernoulli distributed event probability of the mortality outcome and w_j is the provider group-specific random effect that is normally distributed with mean 0 and variance

$$\sigma_w^2$$

For each patient, the predicted mortality is calculated as:

$$Q_{ij} = \frac{1}{1 + \exp\left(-\left(\alpha_0 + \alpha_1 X_{ij,1} + \cdots + \alpha_p X_{ij,p} + w_j\right)\right)}$$

And the expected mortality is calculated as:

$$F_{ij} = \frac{1}{1 + \exp\left(-\left(\alpha_0 + \alpha_1 X_{ij,1} + \cdots + \alpha_p X_{ij,p}\right)\right)}$$

The standardized mortality ratio for the patient is calculated as the ratio of the predicted and expected 1-year mortality

$$SMR_{ij} = Q_{ij}/F_{ij}$$

SMR_ij is interpreted as the patient's risk of death due to their provider group's performance relative to their risk at a provider group of average quality; if

$$SMR_{ij} > 1$$

the patient is at higher risk of death due to their provider group's performance while if

$$SMR_{ij} < 1$$

the patient is at lower risk of death.

Nursing Home Transition Model

To model the transitioning to a nursing home, we use a hierarchical logistic regression model with specifications similar to that used for mortality (above). Similarly, the SNHR is given by

$$SNHR_{ij} = Q_{ij}/F_{ij}$$

, where Q_ij is the predicted risk of transition to a nursing home and F_ij is the expected risk of nursing home transition.

SNHR_ij is interpreted as the patient's risk of transitioning to a nursing home due to their provider group's performance relative to their risk at a provider group of average quality; if

$$SNHR_{ij} > 1$$

the patient is at higher risk of transitioning due to their provider group's performance while if

$$SNHR_{ij} > 1$$

the patient is at lower risk of transitioning.

We finally rescale

$$SNHR_{ij}$$

to have the same mean and standard deviation (SD) over all patients as the SMR, using the equation below; this rescaling ensures that the two values have similar impact when used to adjust the days in care.

$$rSNHR_{ij} = \exp\left(\frac{\left(\log\left(SNHR_{ij}\right) - \overline{\log\left(SNHR_{\iota j}\right)}\right)}{SD\left(\log\left(NHR_{ij}\right)\right)} SD\left(\log\left(SMR_{ij}\right)\right) + \overline{\log\left(SMR_{\iota j}\right)}\right.$$

Adjusted Days in Care

We then use each patient's SMR and SNHR to construct a corresponding adjustment factor.

For the mortality adjustment factor, each patient's EDIC is multiplied by SMR if EDIC ≥ 0 or divided by SMR if EDIC < 0; the patient's EDIC is then subtracted from the result to produce the number of "extra" excess days in care for that patient due to their provider group's performance on mortality.

Similarly, a nursing home transition adjustment factor is constructed by multiplying each patient's EDIC by rSNHR if EDIC ≥ 0 or divided by rSNHR if EDIC < 0; the patient's EDIC is then subtracted from the result to produce the number of "extra" excess days in care for that patient due to their provider group's performance on nursing home transitions. The nursing home transition adjustment factor is multiplied by 0.5. The adjustment factor, in combination with the rescaling of SNHR to have the same mean and standard deviation as SMR, is intended to address feedback from

stakeholders, experts, and patients that death is a more serious outcome than nursing home use by making the overall impact of the SMR adjustment greater. (We found that without these rescaling factors, in the test dataset the distribution of SNHR was broader than that of SMR across patients and so had a much greater impact on a provider group's Days at Home score.)

For each patient, the two adjustment factor representing "extra" excess days are added to each patient's original EDIC to get an "adjusted EDIC" for each patient:

$$Adjusted\ EDIC_{ij} = \begin{cases} EDIC_{ij} + \left[\left(SMR_{ij}EDIC_{ij} - EDIC_{ij}\right) + 0.5\left(rSNHR_{ij}EDIC_{ij} - EDIC_{ij}\right)\right] if\ EDIC_{ij} \geq 0 \\ EDIC_{ij} + \left[\left(\frac{EDIC_{ij}}{SMR_{ij}} - EDIC_{ij}\right) + 0.5\left(\frac{EDIC_{ij}}{rSNHR_{ij}} - EDIC_{ij}\right)\right] if\ EDIC_{ij} < 0 \end{cases}$$

$$= \begin{cases} EDIC_{ij}\left[SMR_{ij} + 0.5\left(rSNHR_{ij} - 1\right)\right] if\ EDIC\ \geq 0 \\ EDIC_{ij}\left[\frac{1}{SMR_{ij}} + 0.5\left(\frac{1}{rSNHR_{ij}} - 1\right)\right] if\ EDIC_{ij} < 0 \end{cases}$$

This risk-, mortality- and nursing home transition adjusted days in care is used to construct provider group-level days at home by subtracting from the cohort mean survival days and averaging over each provider group. That is, the risk-, mortality-, and nursing home transition-adjusted days at home ("Adjusted Days at Home") for provider group j

is the average adjusted days at home of all patients in the provider group, calculated as

$$Adjusted\ DAH_j = \sum_{i=1}^{n_j} \frac{M - adjustedEDIC_{ij}}{n_j}$$

where M is the mean number of days at home of all patients and the sum is over all patients, and n_j is the number of patients in provider group j.

.

[Response Ends]

Attachment: DaysAtHome_MethodReport_021221.pdf

2b.21. If an outcome or resource use measure is not risk-adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (i.e., case mix) is not needed to achieve fair comparisons across measured entities.

[Response Begins]
Not applicable.
[Response Ends]

2b.22. Select all applicable resources and methods used to develop the conceptual model of how social risk impacts this outcome.

[Response Begins]
Internal data analysis
[Response Ends]

2b.23. Describe the conceptual and statistical methods and criteria used to test and select patient-level risk factors (e.g., clinical factors, social risk factors) used in the statistical risk model or for stratification by risk.

Please be sure to address the following: potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10 or other statistical tests; correlation of x or higher. Patient factors should be present at the start of care, if applicable. Also discuss any "ordering" of risk factor inclusion; note whether social risk factors are added after all clinical factors. Discuss any considerations regarding data sources (e.g., availability, specificity).

[Response Begins]
We identified a set of 53 candidate risk variables (including age) used in the Risk-Adjusted Admission Rate for Patients with Multiple Chronic Conditions (MCC) measure. We evaluated the overall prevalence of each risk factor among eligible patient, and the bivariate correlation of each pair of risk factors. We examined bivariate (unadjusted) associations of each risk factor with the measure outcomes, as well as adjusted analyses including all remaining variables. We considered findings of these analyses, along with clinical considerations for the conceptual justification for each risk factor. All risk factors are defined using claims from the year prior to the performance year.

After finalizing the clinical risk model, we investigated the addition of potential social risk factors (SRFs) into the days at home model. We considered the granularity and type of data, we found that only dual-eligible status had a very strong correlation with the outcome, while the other candidate risk factors were either not statistically significant or had only a modest association with days in care. Furthermore, while dual-eligible status is a patient-level indicator, the remaining candidate variables are area-level indicators that may not represent the actual circumstance of a given patient. Finally, some of the other risk factors may be within the control of a provider group to mitigate in the care of its specific patients. Further consideration would be required to justify inclusion of these or other social risk factors.

[Response Ends]

2b.24. Detail the statistical results of the analyses used to test and select risk factors for inclusion in or exclusion from the risk model/stratification.

[Response Begins]
We excluded one variable (pancreatic disease) due to very low prevalence in the cohort, and one variable (reason for Medicare entitlement) due to very high collinearity with age. The statistical models for Days in Care and Nursing Home Transition models included 50 clinical risk variables, age, and dual-eligible status (52 variables total). The Mortality model included 50 clinical risk variables and age (but not dual-eligible status, for 51 variables total). The adjusted Days in Care rate ratio and Mortality and Nursing Home Transition odds ratios (with 95% confidence interval) are shown below.

Variable	Prevalence (%)	Days in Care Rate Ratio (95% CI)	Mortality Odds Ratio (95% CI)	Nursing Home Transition Odds Ratio (95% CI)
Dialysis status	15,361 (1.3%)	1.15 (1.11, 1.19)	1.27 (1.21, 1.33)	1.05 (0.96, 1.14)
Respiratory failure	235,365 (20.4%)	1.19 (1.18, 1.20)	1.28 (1.26, 1.31)	1.10 (1.07, 1.13)
Advanced liver disease	60,544 (5.2%)	1.20 (1.18, 1.22)	1.43 (1.39, 1.47)	1.07 (1.02, 1.12)
Pneumonia	228,942 (19.8%)	1.18 (1.16, 1.19)	1.21 (1.20, 1.23)	1.01 (0.98, 1.03)
Septicemia/shock	110,320 (9.6%)	1.16 (1.14, 1.18)	1.03 (1.01, 1.05)	1.10 (1.07, 1.13)
Marked disability/frailty	238,338 (20.6%)	1.42 (1.40, 1.43)	1.57 (1.55, 1.59)	1.31 (1.28, 1.34)
Pleural effusion/pneumothorax	137,690 (11.9%)	1.18 (1.17, 1.20)	1.44 (1.41, 1.46)	1.04 (1.01, 1.07)
Hematological diseases	204,466 (17.7%)	1.08 (1.07, 1.09)	1.10 (1.08, 1.12)	1.01 (0.99, 1.04)
Advanced cancer	219,594 (19.0%)	1.09 (1.08, 1.11)	2.04 (2.01, 2.07)	0.95 (0.92, 0.98)
Infectious and immunologic diseases	148,704 (12.9%)	1.05 (1.04, 1.07)	1.17 (1.15, 1.19)	0.96 (0.94, 1.00)
Severe cognitive impairment	110,993 (9.6%)	1.24 (1.22, 1.26)	1.29 (1.27, 1.32)	1.29 (1.26, 1.33)
Major organ transplant status	45,961 (4.0%)	0.90 (0.88, 0.92)	0.73 (0.70, 0.76)	0.65 (0.60, 0.71)
Pulmonary heart disease	72,134 (6.2%)	1.17 (1.15, 1.19)	1.27 (1.24, 1.30)	1.09 (1.05, 1.13)
Cardiomyopathy	137,189 (11.9%)	1.00 (0.99, 1.01)	1.13 (1.11, 1.15)	0.97 (0.94, 1.00)
Gastrointestinal disease	323,104 (28.0%)	1.05 (1.05, 1.06)	1.01 (1.00, 1.03)	0.95 (0.93, 0.97)
Bone/joint/muscle infections/necrosis	46,386 (4.0%)	1.19 (1.16, 1.21)	0.93 (0.91, 0.96)	1.12 (1.07, 1.17)

Variable	Prevalence (%)	Days in Care Rate Ratio (95% CI)	Mortality Odds Ratio (95% CI)	Nursing Home Transition Odds Ratio (95% CI)
Iron deficiency anemia	593,522 (51.4%)	1.20 (1.19, 1.21)	1.20 (1.18, 1.21)	1.05 (1.02, 1.07)
Ischemic heart disease except AMI	640,912 (55.5%)	1.04 (1.03, 1.05)	1.02 (1.00, 1.03)	0.96 (0.94, 0.98)
Other lung disorders	544,619 (47.2%)	0.97 (0.96, 0.98)	0.88 (0.87, 0.90)	0.92 (0.90, 0.94)
Vascular or circulatory disease	678,978 (58.8%)	1.15 (1.14, 1.16)	1.11 (1.09, 1.12)	1.16 (1.13, 1.18)
Other significant endocrine disorders	96,762 (8.4%)	1.03 (1.01, 1.04)	0.99 (0.97, 1.01)	0.97 (0.94, 1.00)
Other disability and paralysis	102,794 (8.9%)	1.16 (1.14, 1.17)	1.10 (1.07, 1.12)	1.19 (1.15, 1.22)
Substance abuse	207,690 (18.0%)	1.15 (1.14, 1.16)	1.09 (1.07, 1.11)	1.08 (1.05, 1.11)
Other neurologic disorders	480,678 (41.6%)	1.10 (1.09, 1.11)	0.94 (0.93, 0.95)	1.14 (1.12, 1.17)
Arrhythmia (except atrial fibrillation)	368,340 (31.9%)	1.05 (1.04, 1.06)	0.97 (0.96, 0.99)	1.04 (1.01, 1.06)
Hypertension	998,262 (86.4%)	1.02 (1.01, 1.03)	0.84 (0.82, 0.85)	0.97 (0.94, 1.00)
Hip or vertebral fracture	74,523 (6.5%)	1.26 (1.24, 1.28)	1.06 (1.04, 1.08)	1.28 (1.25, 1.32)
Lower-risk cardiovascular disease	331,724 (28.7%)	1.03 (1.02, 1.04)	0.99 (0.98, 1.01)	1.02 (1.00, 1.04)
Cerebrovascular disease	62,208 (5.4%)	1.07 (1.05, 1.09)	1.00 (0.97, 1.02)	1.09 (1.05, 1.13)
Other malignancy	239,688 (20.8%)	0.98 (0.97, 0.99)	1.23 (1.21, 1.25)	0.99 (0.96, 1.01)
Morbid obesity	190,568 (16.5%)	1.05 (1.04, 1.06)	0.75 (0.74, 0.77)	1.14 (1.11, 1.17)
Urinary disorders	390,470 (33.8%)	1.08 (1.07, 1.09)	1.02 (1.01, 1.03)	1.03 (1.01, 1.05)
Psychiatric disorders other than depression	360,851 (31.2%)	1.10 (1.09, 1.11)	1.01 (1.00, 1.02)	1.08 (1.06, 1.10)
AMI	20,715 (1.8%)	1.12 (1.09, 1.15)	1.26 (1.21, 1.30)	1.06 (1.00, 1.13)
Alzheimer's disease and related disorders or senile dementia	263,438 (22.8%)	1.33 (1.32, 1.35)	1.63 (1.61, 1.66)	1.73 (1.69, 1.76)
Atrial fibrillation	287,448 (24.9%)	1.10 (1.09, 1.11)	1.06 (1.05, 1.08)	1.05 (1.03, 1.07)
CKD	694,137 (60.1%)	1.20 (1.19, 1.21)	1.19 (1.17, 1.20)	1.06 (1.03, 1.08)
COPD and asthma	442,930 (38.4%)	1.10 (1.09, 1.11)	1.05 (1.04, 1.07)	0.97 (0.95, 0.99)
Depression	387,165 (33.5%)	1.16 (1.15, 1.17)	1.02 (1.00, 1.03)	1.24 (1.22, 1.27)
Diabetes	570,837 (49.4%)	1.07 (1.06, 1.07)	1.01 (0.99, 1.02)	0.93 (0.91, 0.94)
Heart failure	562,151 (48.7%)	1.23 (1.22, 1.24)	1.32 (1.30, 1.34)	1.11 (1.09, 1.14)
Stroke and TIA	107,238 (9.3%)	1.07 (1.05, 1.08)	1.01 (0.99, 1.03)	1.07 (1.04, 1.10)
Other organ transplant	116,641 (10.1%)	1.08 (1.07, 1.09)	1.04 (1.02, 1.06)	1.06 (1.03, 1.09)
Precerebral arterial occlusion and transient cerebral ischemia	146,291 (12.7%)	0.95 (0.93, 0.96)	0.89 (0.87, 0.91)	0.94 (0.92, 0.97)
Diabetic retinopathy	77,808 (6.7%)	1.01 (1.00, 1.03)	0.93 (0.91, 0.96)	0.95 (0.92, 0.98)
Walking aids	55,466 (4.8%)	1.07 (1.05, 1.09)	0.92 (0.89, 0.94)	1.12 (1.08, 1.16)
Wheelchairs	81,605 (7.1%)	1.14 (1.12, 1.16)	1.14 (1.12, 1.17)	1.25 (1.21, 1.28)
Hospital bed	32,422 (2.8%)	1.06 (1.03, 1.08)	1.24 (1.20, 1.28)	1.10 (1.05, 1.15)
Lifts	7,525 (0.7%)	0.87 (0.83, 0.91)	1.11 (1.05, 1.19)	0.91 (0.84, 0.99)
Oxygen	158,476 (13.7%)	1.14 (1.12, 1.15)	1.45 (1.42, 1.48)	0.99 (0.96, 1.02)
Age 85 and older (Referent)	*237,344 (20.55%)*	*1 (Ref.)*	*1 (Ref.)*	*1 (Ref.)*
Age 18-54	74,680 (6.5%)	0.56 (0.55, 0.57)	0.23 (0.22, 0.24)	0.17 (0.16, 0.17)
Age 55-64	109,618 (9.5%)	0.57 (0.56, 0.58)	0.32 (0.32, 0.33)	0.26 (0.24, 0.27)
Age 65-74	359,836 (31.2%)	0.62 (0.61, 0.63)	0.41 (0.40, 0.41)	0.34 (0.33, 0.35)
Age 75-84	373,301 (32.3%)	0.77 (0.76, 0.78)	0.56 (0.55, 0.57)	0.55 (0.54, 0.57)

Variable	Prevalence (%)	Days in Care Rate Ratio (95% CI)	Mortality Odds Ratio (95% CI)	Nursing Home Transition Odds Ratio (95% CI)
Dual-eligible	271,506 (23.5%)	1.49 (1.47, 1.50)	n/a	2.62 (2.57, 2.68)

[Response Ends]

2b.25. Describe the analyses and interpretation resulting in the decision to select or not select social risk factors.

Examples may include prevalence of the factor across measured entities, availability of the data source, empirical association with the outcome, contribution of unique variation in the outcome, or assessment of between-unit effects and within-unit effects. Also describe the impact of adjusting for risk (or making no adjustment) on providers at high or low extremes of risk.

[Response Begins]
In the table below, we report the adjusted rate ratio for days in care of each candidate SRF, also controlling for the 51 final clinical and demographic variables. The most notable effect was dual eligibility, with a risk ratio of 1.248. After adjustment, the effect for most other variables was either not significant (urban residence, hospital bed density, and percent living alone), fairly small in magnitude (SES index, PCP density, nursing home density, and percent never married). Inclusion of all SRFs resulted in only a minor improvement in predictive value of the model above the clinical and demographic factors, with a deviance r-squared of 0.01789 (compared to 0.01727 using clinical risk factors only).

Variable	Adjusted Risk Ratio (95% CI)	Interpretation
Dual-eligible	1.248 (1.236, 1.261)	Dual-eligible beneficiaries have more days in care than Medicare-only
AHRQ SES index	0.988 (0.985, 0.991)	Fewer days in care among higher quintiles of SES index (that is, for patients with higher SES and lower social risk)
Urban	0.993 (0.980, 1.261)	Residence in an urban county not significantly associated with days in care
Specialist density	1.109 (1.076, 1.143)	Patients in counties with zero (0) specialists have more days in care than those in counties with one (1) or more
PCP density	1.022 (1.014, 1.030)	More days in care among higher quintiles of PCP density (that is, for patients in counties with more PCPs per 100,000)
Hospital beds	0.997 (0.994, 1.000)	Higher density of hospital beds not significantly associated with days in care
Certified nursing home beds	1.071 (1.067, 1.076)	More days in care among higher quintiles of nursing home beds (that is, for patients in counties with more beds per 100,000)
Never married	1.025 (1.021, 1.029)	More days in care among higher quintiles of unmarried density (that is, for patients in counties with higher percentage of individuals never married)
Living alone	0.997 (0.994, 1.000)	Percentage households within a county that are single occupant not significantly associated with days in care

Based on these results, we included dual-eligible status as a risk factor in the Days in Care and Nursing Home Transition models, though not in the Mortality model. Dual-eligible status is available at the patient level, bears a strong conceptual and empirical relationship to the outcomes of interest (as both an indicator of financial hardship that may affect patterns of care and a structural difference in patients' ability to pay for select services), and is readily available in the Medicare enrollment file. By incorporating this adjustment, DCEs would not be incentivized to preferentially treat Medicare-only patients or penalized for treating more dual-eligible patients.

We did not include dual-eligible status in the Mortality component model. This decision aligns with other existing measures of mortality, reflects the stronger relationship of mortality to clinical risk factors, and avoids creating different

standards for the care or survival of dual-eligible patients.

We did not include other SRFs for several reasons: they are area-level indicators that may not represent the actual circumstance of each patient, the empirical support for inclusion is weaker, and they require abstraction of data from external sources which could present a problem in measure implementation. In addition, some of these risk factors (particularly physician density and density of hospital or nursing home beds) can be directly addressed and mitigated by DCEs, in which case statistical adjustment would be inappropriate.

[Response Ends]

2b.26. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (describe the steps—do not just name a method; what statistical analysis was used). Provide the statistical results from testing the approach to control for differences in patient characteristics (i.e., case mix) below. If stratified ONLY, enter "N/A" for questions about the statistical risk model discrimination and calibration statistics.

Validation testing should be conducted in a data set that is separate from the one used to develop the model.

[Response Begins]
We computed the c-statistic (area under the ROC curve) to evaluate the discrimination of the Mortality and Nursing Home Transition logistic regression models.

We evaluated the goodness-of-fit of the Days in Care negative binomial count model using the deviance R-squared and the Spearman rank correlation coefficient.

We also computed the c-statistic of a separate logistic regression model using the same risk factors as Days in Care for which the outcome was "at least one day in care" vs. "zero days in care" as an alternative way to assess the predictive ability of the risk factors used in the main count model.

We computed deviance R^2 to assess the performance of the Days in Care count model, and the c-statistic (area under the ROC curve) for the Mortality and Nursing Home Transition models.
[Response Ends]

2b.27. Provide risk model discrimination statistics.

For example, provide c-statistics or R-squared values.

[Response Begins]
- Mortality logistic model: C-statistic = 0.738
- Nursing home transition logistic model: C-statistic = 0.760
- Days in Care NB model: Deviance R-squared = 0.01698
- Days in Care NB model: Spearman R = 0.346 ($p < 0.0001$)
- "Any Day in Care" logistic model: C-statistic = 0.688

[Response Ends]

2b.28. Provide the statistical risk model calibration statistics (e.g., Hosmer-Lemeshow statistic).

[Response Begins]
We assessed the calibration of the Days at Home component models using a split-half methodology. We split the cohort randomly in half, fit each model to the first half of data (the development sample), and then used those coefficients in the second half (the validation sample) to confirm the models are generalizable and well-calibrated.
Testing and calibration results, Mortality model (dataset: CY 2018 development dataset)

Characteristic	Development Sample	Validation Sample
Number of patients	577,544	577,235
Number of eligible ACOs	610	610
Number of deaths	68,668	68,546
Unadjusted outcome rate (ACO level)	11.55%	11.60%
Calibration (γ0, γ1)	(0,1)	(-0.0083, 0.9954)
Discrimination - predictive ability (lowest decile %, highest decile %)	(2.79%, 35.13%)	(2.80%, 34.95%)
Discrimination – C-statistic	0.739	0.737

Testing and calibration results (Nursing Home Transition model) (dataset: CY 2018 development dataset)

Characteristic	Development Sample	Validation Sample
Number of patients	577,544	577,235
Number of eligible ACOs	610	610
Number of nursing home transitions	26,619	26,450
Unadjusted outcome rate	4.44%	4.39%
Calibration (γ0, γ1)	(0,1)	(-0.0117, 0.9977)
Discrimination - predictive ability (lowest decile %, highest decile %)	(1.20%, 14.32%)	(1.20%, 14.18%)
Discrimination – C-statistic	0.734	0.733

Testing and calibration results (Days in Care model) (dataset: CY 2018 development dataset)

Characteristic	Development Sample	Validation Sample
Number of patients	577,544	577,235
Number of eligible ACOs	610	610
Unadjusted mean Days in Care (ACO level)	12.72	12.71
Predictive ability (lowest decile of predicted Days in Care, highest decile)	3.90, 50.70	3.92, 50.65
Model fit (deviance R-squared)	0.01033	0.01029

[Response Ends]

2b.29. Provide the risk decile plots or calibration curves used in calibrating the statistical risk model.

The preferred file format is .png, but most image formats are acceptable.

[Response Begins]
Figure 1: Days in Care Calibration Deciles (Development Sample)

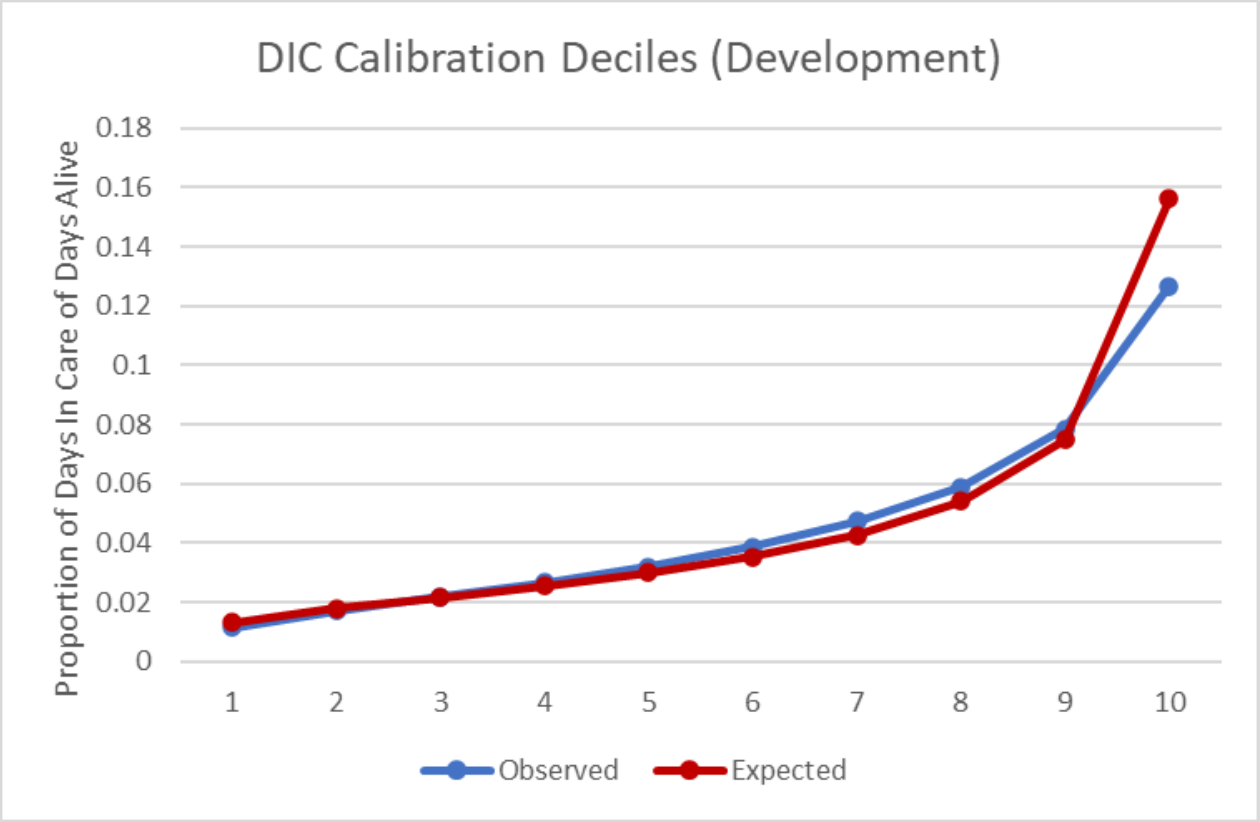
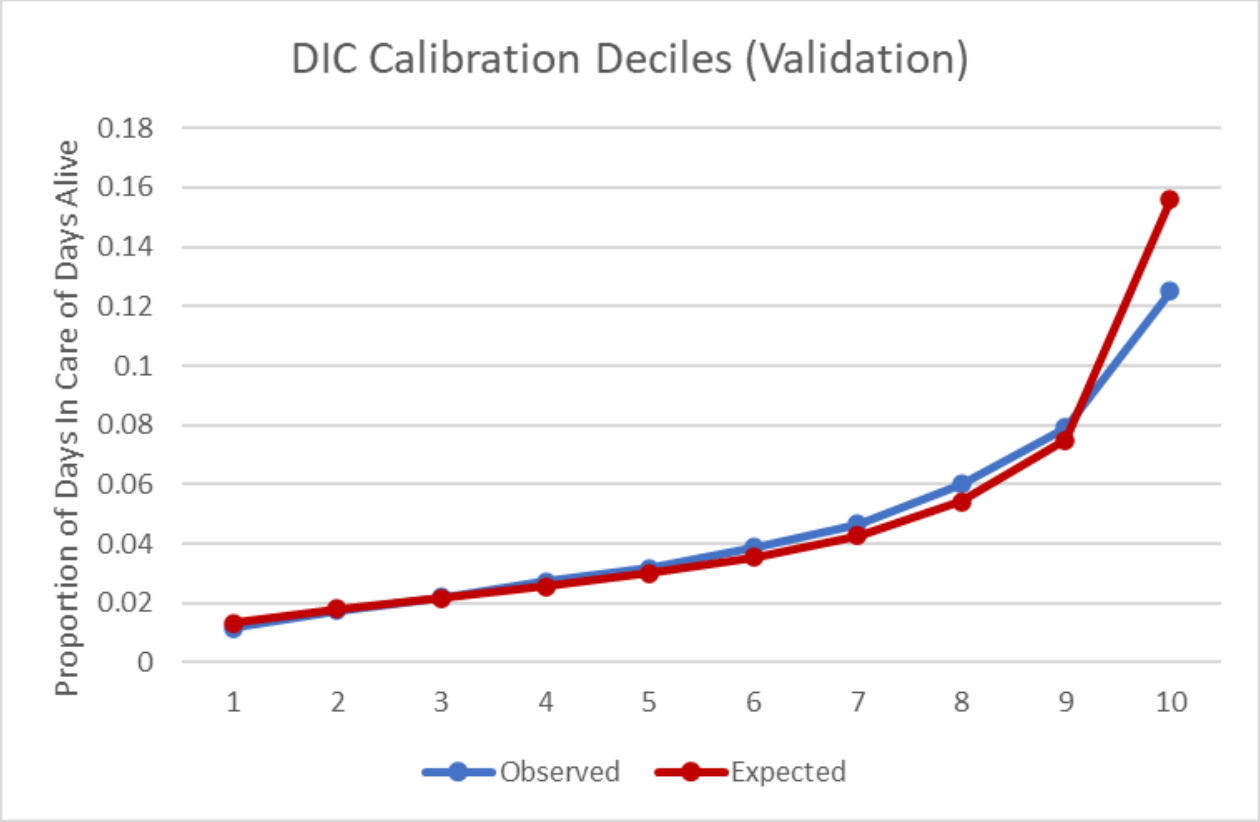


Figure 2: Days in Care Calibration Deciles (Validation Sample)



[Response Ends]

2b.30. Provide the results of the risk stratification analysis.

[Response Begins]
Not applicable.
[Response Ends]

2b.31. Provide your interpretation of the results, in terms of demonstrating adequacy of controlling for differences in patient characteristics (i.e., case mix).

In other words, what do the results mean and what are the norms for the test conducted?

[Response Begins]
Using the SSP ACO test dataset, we evaluated the discrimination and goodness-of-fit of the Days at Home component models to assess their ability to differentiate outcomes.

- We found a C-statistic of 0.738 for the Mortality model, which indicates good classification of observed deaths using model predictions. We found a C-statistic of 0.760 for the Nursing Home Transition model, which indicates similarly good model performance. Potential values of the C-statistic range from 0, indicating perfectly poor classification to 0.5, meaning the model predictions are no better than chance, to 1.0, indicating perfect classification; "perfect classification" here implies that patients' outcomes can be predicted by the specified risk factors alone and that no other factors, including performance of healthcare providers, play a role in patients' outcomes.
- We evaluated the goodness-of-fit of the Days in Care count model using several statistics. We first computed the deviance R-squared to be 0.0170. It is important to note that the deviance R-squared is distinct from the "standard" R-squared definition that is often reported for linear regression models and should not be interpreted in the same way; deviance R-squared is a separate measure of fit applicable to count models estimated using maximum-likelihood methods (equal to one minus the ratio of the log-likelihoods of the final model and the null model).1 Deviance R-squared values are typically lower than standard R-squared values; the deviance R-squared of 0.0170 for the Days in Care model is comparable to that observed in other similar count models used in CMS measures (specifically values of 0.060, 0.028, and 0.038 for the count models of the 30-day Excess Days in Acute Care measures for Acute Myocardial Infarction, Heart Failure, and Pneumonia respectively).2
- We also computed the patient-level Spearman rank correlation for predicted days in care (as a proportion of days alive) from the Days in Care model and the observed days in care; we found a Spearman r of 0.346 ($p < 0.0001$), indicating that the model tends to predict more days in care for patients who actually did spend more days in care.
- Finally, to assess the predictive value of the risk factors in the Days in Care model, we computed the C-statistic of a separate logistic regression model using the same risk factors in which we treat "at least one day in care" as a binary outcome and predict, for each patient, the probability that they spend at least one day in care. We compared these predicted probabilities to observed and found a C-statistic of 0.688, indicating an adequate prediction of patient experience a day in care.

Using the split-half development and validation sample results, we computed discrimination and calibration metrics for each of the component models (Mortality, Nursing Home Transition, Days in Care). Overall, the models fitted in the development sample show very similar performance and results when applied to data in the validation sample.

- Of particular note, the overfitting indices ($\gamma 0$, $\gamma 1$) in the validation sample of the mortality model (-0.0083, 0.9954) and nursing home transition model (-0.0117, 0.9977) are very close to the development sample (by definition: 0, 1), indicating that those models are generalizable and have high predictive ability.
- Furthermore, the C-statistics based on the validation sample (0.737 for Mortality, 0.733 for Nursing Home Transitions) are approximately equal to those based on the development sample (0.739 and 0.734), indicating that the model discrimination is maintained when applied to new data.
- Similarly, the deviance R-squared of the Days in Care model is similar between the development and validation sample results (0.01033 and 0.01029 respectively), which indicates similar goodness-of-fit in both samples.

1. Cameron AC, Windmeijer F. R-Squared Measures for Count Data Regression Models with Applications to Health-care Utilization. Journal of Business & Economic Statistics. 1996;14(2):209-220.

2. YNHHSC/CORE. Condition-Specific Excess Days in Acute Care Measures Annual Updates and Specifications Report. 2021.

[Response Ends]

2b.32. Describe any additional testing conducted to justify the risk adjustment approach used in specifying the measure.

Not required but would provide additional support of adequacy of the risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed.

[Response Begins]
Not applicable
[Response Ends]

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3.01. Check all methods below that are used to generate the data elements needed to compute the measure score.

[Response Begins]
Coded by someone other than person obtaining original information (e.g., DRG, ICD-10 codes on claims)
[Response Ends]

3.02. Detail to what extent the specified data elements are available electronically in defined fields.

In other words, indicate whether data elements that are needed to compute the performance measure score are in defined, computer-readable fields.
[Response Begins]
ALL data elements are in defined fields in electronic claims
[Response Ends]

3.03. If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using data elements not from electronic sources.

[Response Begins]
Not applicable.
[Response Ends]

3.04. Describe any efforts to develop an eCQM.

[Response Begins]
Not applicable.
[Response Ends]

3.06. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.

[Response Begins]
Not applicable.
[Response Ends]

Consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

3.07. Detail any fees, licensing, or other requirements to use any aspect of the measure as specified (e.g., value/code set, risk model, programming code, algorithm),

Attach the fee schedule here, if applicable.

[Response Begins]
There are no fees, licensing, or other requirements to use this measure as specified.
[Response Ends]

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making.

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement, in addition to demonstrating performance improvement.

4a.01.

Check all current uses. For each current use checked, please provide:

Name of program and sponsor

URL

Purpose

Geographic area and number and percentage of accountable entities and patients included

Level of measurement and setting

[Response Begins]
Payment Program
Quality Improvement with Benchmarking (external benchmarking to multiple organizations)
The measure is currently being implemented on a pay-for-reporting basis in the Direct Contracting model, using claims data collected beginning January 1, 2021.

- Purpose: The Direct Contracting model is a new alternative payment model that offers innovative payment structures to support the delivery of advanced primary care among participating ACOs
- Geographic area: All geographic areas in the United States are eligible to participate
- Number and percentage of accountable entities and patients included: To be determined
- Level of measurement: ACO-level
- Setting: Outpatient Rehabilitation, Behavioral Health: Inpatient, Behavioral Health: Outpatient, Home Health
- Direct Contracting Model URL: https://innovation.cms.gov/initiatives/direct-contracting-model-options/

The measure is intended to enter use on a pay-for-performance basis in both the Primary Care First and Direct Contracting models for payment determinations beginning in 2023.

The measure is currently being implemented on a pay-for-reporting basis in the Direct Contracting model, using claims data collected beginning January 1, 2021.

- Purpose: The Direct Contracting model is a new alternative payment model that offers innovative payment structures to support the delivery of advanced primary care among participating ACOs
- Geographic area: All geographic areas in the United States are eligible to participate
- Number and percentage of accountable entities and patients included: To be determined
- Level of measurement: ACO-level
- Setting: Outpatient Rehabilitation, Behavioral Health: Inpatient, Behavioral Health: Outpatient, Home Health
- Direct Contracting Model URL: https://innovation.cms.gov/initiatives/direct-contracting-model-options/

The measure is intended to enter use on a pay-for-performance basis in both the Primary Care First and Direct Contracting models for payment determinations beginning in 2023.

[Response Ends]

4a.02. Check all planned uses.

[Response Begins]
Payment Program
Quality Improvement with Benchmarking (external benchmarking to multiple organizations)
[Response Ends]

4a.03. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing), explain why the measure is not in use.

For example, do policies or actions of the developer/steward or accountable entities restrict access to performance results or block implementation?

[Response Begins]
Not applicable.
[Response Ends]

4a.04. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes: used in any accountability application within 3 years, and publicly reported within 6 years of initial endorsement.

A credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.

[Response Begins]
Not applicable.
[Response Ends]

4a.05. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

Detail how many and which types of measured entities and/or others were included. If only a sample of measured entities were included, describe the full population and how the sample was selected.

[Response Begins]
This is a new measure, and data and performance results have not yet been provided to entities to be measured. CMS has made a Measure Information Form (MIF) outlining the measure specifications available to entities participating in the Direct Contracting model and will do similarly for the Primary Care First model. Entities will have opportunities to ask questions about the measure specifications and interpretation of results through each model's Question & Answer mechanism; feedback gained through this mechanism will be used by the measure steward to inform measure maintenance.
[Response Ends]

4a.06. Describe the process for providing measure results, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

[Response Begins]
Not applicable.
[Response Ends]

4a.07. Summarize the feedback on measure performance and implementation from the measured entities and others. Describe how feedback was obtained.

[Response Begins]
No feedback has been obtained from entities being measured or others.
[Response Ends]

4a.08. Summarize the feedback obtained from those being measured.

[Response Begins]
No feedback has been obtained from entities being measured.
[Response Ends]

4a.09. Summarize the feedback obtained from other users.

[Response Begins]
No feedback has been obtained other users.
[Response Ends]

4a.10. Describe how the feedback described has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.

[Response Begins]
Not applicable.
[Response Ends]

4b.01. You may refer to data provided in Importance to Measure and Report: Gap in Care/Disparities, but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included). If no improvement was demonstrated, provide an explanation. If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

[Response Begins]
This is a new measure. Data on performance improvement are not available.

Performance on the measure will provide a quality signal to providers if their patients are spending more time in select acute and post-acute settings and out of their home or community setting than expected compared to other providers in the same program.
[Response Ends]

4b.02. Explain any unexpected findings (positive or negative) during implementation of this measure, including unintended impacts on patients.

[Response Begins]
Not applicable.
[Response Ends]

4b.03. Explain any unexpected benefits realized from implementation of this measure.

[Response Begins]
Not applicable.

[Response Ends]

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

If you are updating a maintenance measure submission for the first time in MIMS, please note that the previous related and competing data appearing in question 5.03 may need to be entered in to 5.01 and 5.02 if the measures are NQF endorsed. Please review and update questions 5.01, 5.02, and 5.03 accordingly.

5.01. Search and select all NQF-endorsed related measures (conceptually, either same measure focus or target population).

(Can search and select measures.)
[Response Begins]
2888: Risk-Standardized Acute Admission Rates for Patients with Multiple Chronic Conditions
[Response Ends]

5.02. Search and select all NQF-endorsed competing measures (conceptually, the measures have both the same measure focus or target population).

(Can search and select measures.)
[Response Begins]
[Response Ends]

5.03. If there are related or competing measures to this measure, but they are not NQF-endorsed, please indicate the measure title and steward.

[Response Begins]
Not applicable.
[Response Ends]

5.04. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s), indicate whether the measure specifications are harmonized to the extent possible.

[Response Begins]
Yes
[Response Ends]

5.05. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.

[Response Begins]
The Days at Home measure expands on the UAMCC measure priorities of improving care coordination and home-based care while discouraging the use of preventable acute hospital visits by considering the total days spent in care (rather than just total number of admissions), considering care in a broader range of settings, and additionally accounting for mortality and transitions to residential nursing homes to mitigate potential unintended consequences. The Days at Home measure uses a similar set of risk factors to take advantage of the development of the UAMCC measure, with a few differences as discussed in the risk adjustment section.

While the eligible cohorts overlap for both measures, they are not identical, with Days at Home including patients younger than 65 as well as patients with different illness or complex conditions; while the UAMCC cohort includes patients who have two or more specific qualifying conditions, the Days at Home cohort includes patients with complex, chronic illness as defined by an HCC risk score greater than 2.0 (which may be attained through various combinations of

risk factors). This is consistent with the cohort of the Primary Care First and Direct Contracting models (which include patients age 18 and older), and with those models' objective to emphasize care of patients meeting a broad definition of serious illness or complex chronic disease.

Both measures are claims-based and there would be no impact on data collection burden for providers reporting either or both measures.
[Response Ends]

5.06. Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality). Alternatively, justify endorsing an additional measure.

Provide analyses when possible.

[Response Begins]
Not applicable. There are no competing measures.
[Response Ends]

Appendix

Supplemental materials may be provided in an appendix.: Available in attached file
Attachment: 3667_DaysAtHome3667_TEPSummaryReport.pdf

Contact Information

Measure Steward (Intellectual Property Owner): Centers for Medicare & Medicaid Services
Measure Steward Point of Contact: Winder-Wells, Teresa, teresa.winder-wells@cms.hhs.gov
Stearle, Carla, carla.stearle@cms.hhs.gov
Gutermuth, Leah, leah.gutermuth@cms.hhs.gov

Measure Developer if different from Measure Steward: Yale New Haven Health Services Corporation – Center for Outcomes Research and Evaluation (CORE)
Measure Developer Point(s) of Contact: Bernheim, Susannah, susannah.bernheim@yale.edu
Bagshaw, Kyle, kyle.bagshaw@yale.edu
Miller, Jake, jake.miller@yale.edu

Additional Information

1. Provide any supplemental materials, if needed, as an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be collated one file with a table of contents or bookmarks. If material pertains to a specific criterion, that should be indicated.

[Response Begins]
Available in attached file
[Response Ends]

Attachment: 3667_DaysAtHome3667_TEPSummaryReport.pdf

2. List the workgroup/panel members' names and organizations.

Describe the members' role in measure development.

[Response Begins]
CORE convened a multi-disciplinary Technical Expert Panel (TEP) composed of individuals with expertise in clinical practice, administration, quality measurement, performance improvement, healthcare disparities, and/or empirical methodologies, as well as patients, family caregivers, and advocates. The TEP has provided feedback and informed our decisions and recommendations for the technical specifications of the measure, including the cohort definition, outcome definition, risk factor selection, inclusion of mortality information, nursing home use and social risk factors, and the overall concept and impact of the measure from the perspective of both patients and providers. A detailed summary of the TEP's input may be found in the attached TEP Summary Report document.

- Sheila Antony, MD, MHCDS; Iora Health, Denver, CO
- Rosie Bartel, MA; Patient & Family Advocate Network, Chilton, WI
- Vipul Bhatia, MD, MBA; WellSpan Health, York, PA
- Stephanie Bruce, MD, FACP; Medstar House Call Program, Washington DC
- David Casarett, MD, MACP, FAAHPM, HMDC; Duke University/Duke Health/Duke Center for Palliative Care, Durham, NC
- Todd Cook, MBA, MSW, EdD; Sharp HealthCare, San Diego, CA
- Melody Danko-Holsomback, CRNP, MSN; Keystone Accountable Care Organization, Danville, PA
- Linda Delo, DO; Delo Medical Associates, Port Saint Lucie, FL
- Mark Friedberg, MD, MPP; Blue Cross Blue Shield, Boston, MA
- Deborah Hill, BS, FAHA; Gainesville, FL
- Karen Hyden, PhD, MSN, MEd, APN-BC, ACHPN; Compassus Hospice, Brentwood, TN
- Rebecca Kirch, JD; National Patient Advocate Foundation, Washington DC
- Bruce Kinosian, MD; University of Pennsylvania/Department of Veterans Affairs, Philadelphia, PA
- Bruce Leff, MD; Johns Hopkins School of Medicine, Baltimore, MD
- David Longnecker, MD, FRCA; Coalition to Transform Advanced Care, Washington DC
- Dana Lustbader, MD, FAAHPM; ProHEALTH Care, New York, NY
- James Mittelberger, MD, MPH, FACP, FAAHPM; Center for Elders' Independence, Oakland, CA
- Jennifer Ofelt, MHA, MSN, RN; UnityPoint at Home, Urbandale, IA
- Carol Raphael, MPA; Manatt Health Solutions, New York, NY
- Robert Rosati, PhD; Visiting Nurse Association Health Group, Holmdel, NJ
- Janelle Shearer, RN, BSN, MA, CPHQ; Stratis Health, Bloomington, MN

[Response Ends]

3. Indicate the year the measure was first released.

[Response Begins]
2021
[Response Ends]

4. Indicate the month and year of the most recent revision.

[Response Begins]
Not applicable.
[Response Ends]

5. Indicate the frequency of review, or an update schedule, for this measure.

[Response Begins]
To be determined.
[Response Ends]

6. Indicate the next scheduled update or review of this measure.

[Response Begins]
To be determined.
[Response Ends]

7. Provide a copyright statement, if applicable. Otherwise, indicate "N/A."

[Response Begins]
This quality measure was developed for CMS by Yale New Haven Health Services Corporation Center for Outcomes Research and Evaluation (CORE) in 2021.
[Response Ends]

8. State any disclaimers, if applicable. Otherwise, indicate "N/A."

[Response Begins]
N/A
[Response Ends]

9. Provide any additional information or comments, if applicable. Otherwise, indicate "N/A."

[Response Begins]
We would like to provide some additional comments pertaining to the measure validity following the discussion of the NQF Scientific Methods Panel (SMP) on October 27, 2021 – specifically, the validity of the adjustment for mortality and nursing home transitions in the measure methodology, and the results of our empiric validity testing.

Comment on Validity of Mortality & Nursing Home Adjustments

A few members of the SMP expressed concerns about the approach to incorporating mortality and nursing home transition information into the Days at Home measure score. The specific concern was that the adjustments for mortality and nursing home transition were arbitrary and thus a threat to validity. However, these adjustments were based on input from the Technical Expert Panel (TEP) engaged in the development process and were to enhance the measure validity and address concerns about potential unintended consequences of measures. As discussed below, we think that the testing results provided in our submission demonstrate the validity of the Days at Home measure.

The centerpiece of the Days at Home measure is the risk-adjusted model of "excess days in care" (EDIC) which is computed at the patient level; these excess days in care are then subtracted from the days alive to compute a final "days at home." Then, using the same cohort, a "standardized mortality ratio" (SMR) and "standardized nursing home ratio" (SNHR) is calculated for each patient. The EDIC for each patient is then multiplied by the SMR and SNHR to derive a days

in care "adjustment" which reflects the provider's performance on mortality and nursing home transitions. This (possibly positive or negative) adjustment is then added into the final EDIC for each patient. As shown, in most cases it has a very modest impact on the Excess Days in Care Score (which is ultimately converted to a Days at Home Score)

The motivation behind the adjustment approach is twofold. First, the adjustments account for the differing exposure periods experienced by patients who are at home for the entire year versus those patients who do not survive or who transfer to a nursing home. Though the EDIC model includes an offset for survival days, we felt that this did not fully address the adverse nature of mortality, in that days *not* at home are in most cases preferable to death. Similarly, though the EDIC model counts days at home for patients who are in a nursing home the same as those who are community residents, we felt that in many cases (but not all), the transition itself represents an adverse outcome for patients, who again would likely value days at home more if they were community-based.

The second motivation is to formally account for potential adverse consequences of measuring EDIC only – namely, to ensure the measure does not inadvertently reward providers who achieve more Days at Home in ways that put patients at greater risk for these other undesired outcomes (for example, if it is more difficult for patients to access medically necessary care, or if patients are transitioned to nursing homes without providers first fully exploring options for home- and community-based care). The adjustments for mortality and nursing home transition are not intended to counter-balance the measure of days in care, but rather to provide a formal adjustment so that providers are assured a fair comparison with other providers. CORE incorporated the adjustments made to the EDIC to address both concerns – that of patient perspective and competing risks - in a way that is both direct and modest.

CORE convened a 21-member, multi-disciplinary Technical Expert Panel (TEP) to provide input and feedback through the development process. The TEP consisted of individuals with expertise in quality measurement, technical methodologies, and performance improvement, along with clinical experts in multiple settings and several patients and patient advocates. CORE extensively discussed the methodology decisions with this group, which independently noted the importance of accounting for patient deaths and nursing home transitions to better reflect patient experience and to mitigate possible adverse consequences, and broadly supported CORE's final decision to do so in this way (including a unanimous vote of all 19 responding TEP members supporting the measure's face validity). More details on the TEP's discussions can be found in the attached TEP Summary Report.

Specifically, the TEP unanimously agreed that mortality is an adverse outcome that the measure should capture in some way, to ensure that providers are not somehow rewarded for putting patients at greater risk. While the Days in Care model includes an offset for days alive, this only has the effect of modeling days in care for each patient as a proportion of days alive but does not account for the actual risk of death due to their provider's performance.

In addition, the TEP broadly agreed that transitioning to a nursing home should not simply be considered as being "at home." At the same time, the TEP did not support counting days in a nursing home the same as days in care in acute, post-acute, or skilled inpatient care settings; they broadly felt that once patients have transitioned to a nursing home, their providers should still be accountable for providing quality care in place and reducing their need for care in those more serious settings.

Because the measure cohort includes patients with complex illness, the risk of death and nursing home use is high – the overall crude mortality rate in the SSP ACO testing cohort is 11.6% (ranging from 6.0% to 30.4% at the ACO level), and the overall crude nursing home transition rate is 4.4% (ranging from 0.4% to 15.8%). In other words, the risk of these outcomes for patients in the cohort is not trivial (which was also independently noted from the personal experience of our patient & advocate TEP members) and varies notably by provider.

In the context of the measure's intended use in the Primary Care First and Direct Contracting Innovation Models, the lack of a suitable balancing measure for these adverse outcomes in the models' quality strategies was a major consideration. (For example, while readmission measures in the Hospital Quality Reporting Program do not account for mortality, that program separately includes measures of mortality for very similar patient cohorts.)

Empirically (based on our SSP ACO test dataset), the SMR and SNHR adjustments have a small-to-modest impact on providers' Days at Home scores, with the greatest impact for providers with substantially above- or below-average performance in both EDIC and either SMR or SNHR. For most providers in the middle of the EDIC distribution, the effect of the SMR and SNHR adjustment is quite small, and even for providers with more extreme EDIC the dominating component

is the EDIC score. For example, the 10th and 90th percentile of absolute change in score due to these adjustments for ACOs is between -0.67 days and +0.49 days at home, whereas the range from 10th percentile to 90th percentile of ACO performance on the measure score is 327.0 to 333.5 days at home. In addition, the Spearman rank correlation between Days at Home (without mortality and nursing home adjustments) and Days at Home (with both adjustments) is 0.9979, showing that the adjustments have very minor impact on the performance of providers relative to each other. These results demonstrate that, given the validity of the underlying EDIC model itself, the validity of the final measure is not compromised by incorporating these adjustments. Overall, the adjustments are fulfilling the intended purpose of capturing the value to patients of avoiding both mortality and nursing home transitions without dominating the signal of the Days in Care component.

Comment on Empirical Validity Testing Results

A few SMP members had a concern regarding validity based on the results of several empiric validity analyses reported by CORE, which compared ACO-level scores on Days at Home to scores on several other SSP measures. (To note, only the TEP Face Validity assessment is required as Days at Home is a new measure.)

The Days at Home measure is different from other SSP measures (specifically Risk-Standardized All-Condition Readmission (ACO-8) and SNF 30-day All-Cause Readmission (ACO-35)) in two key aspects.

First, the cohort of the Days at Home measure is quite different. Days at Home is a population-based measure, including all adult (age 18+) patients with complex, chronic conditions aligned to a participating provider throughout the measurement period regardless of observed care use. By contrast, the cohorts for ACO-8 and ACO-35 consist of patients age 65+ who have had a qualifying index admission, for a follow-up period of 30 days following the index admission. In other words, a broader patient population is eligible for ACO-8 and ACO-35, but the cohort only includes those with a recent acute event.

Second, the outcome of the Days at Home measure is much broader, capturing care utilization in multiple different settings (including hospitals, skilled nursing facilities, and other inpatient facilities) throughout the measurement period rather than only hospital (for ACO-8) or SNF (for ACO-35) admissions within 30 days of an index event. Days at Home additionally captures the duration of admissions to these settings rather than a dichotomous outcome.

A few SMP members noted that the SSP measures of admission share some overlap in the outcome definitions and so might be expected to have a correlation; while there may be some overlap in the cohort and outcome between Days at Home and ACO-8 and ACO-35, there are notable differences in the specifications that should be taken into account when interpreting these results.

Finally, it is important to note that having an expected correlation with measures which share cohort or outcome characteristics may or may not support the validity of the Days at Home measure, it does not undermine the validity of the measure; we expect similar measures to have similar patterns across providers.

The method and results of this testing are reported in greater detail in the “Scientific Acceptability: Validity – Testing” section.
[Response Ends]