Content

Brief Measure Information

CBE #: 3742

Corresponding Measures:

Measure Title: ESRD Dialysis Patient Life Goals Survey (PaLS)

Measure Steward: Centers for Medicare & Medicaid Services

sp.02. Brief Description of Measure: The PaLS is a patient self-report survey that includes eight items related to dialysis facility care team discussions about patient life goals. Six of the items are Likert-type items that are used to generate a "quality of facility care team discussion" score (described below). The remaining two items on the PaLS are checklist items: (1) a list of patient-reported life goals; and (2) a patient-reported list of dialysis care team members that the patient reports has talked with them about their life goals. These items are not scored. Instead, these items serve to provide contextual information for both the patient and the facility to guide care team discussions.

The PaLS is used to generate a patient-level t-score that reflects patient-reported satisfaction with how well his/her/their facility is doing in discussing life goals with the patient as part of the treatment planning process. For each individual patient at a given facility, the calculated t-score (M=50; SD=10) represents a patient's perceptions of their satisfaction with their dialysis care team discussions about life goals. A t-score greater than 40 and less than 60 reflects a score that is within normal limits of existing practices and should not warrant further action, assuming typical existing practices for patient and care team discussions are deemed adequate. Assuming a normal distribution, scores that are ≤40 would warrant follow-up by the facility. Specifically, scores ≤40 (i.e., ≥1 SD below the mean) suggest patient perceptions of care discussions are worse than 84% of their peers, whereas scores ≤30 (i.e., ≥2 SDs below the mean) suggest patient perceptions of care discussions are worse than 98% of their peers; scores that are >40 would be within "normal limits" (Heaton et al., 2004). The t-score is based on the data collected for the instrument testing, as described in the section on scientific acceptability, but is currently not part of the process measure calculation.

The target population for the measure is patients on chronic dialysis who meet all of the following criteria:

- Are at least 18 years old
- Completed the PaLS survey at least once during the one-year reporting period

Content

Reference:
Heaton, R. K., Miller, S. W., Taylor, J. T., & Grant, I. (2004). Revised comprehensive norms for an expanded Halstead-Reitan Battery: Demographically adjusted neuropsychological norms for African American and Caucasian adults. Lutz, FL: Psychological Assessment Resources, Inc.

1b.01. Developer Rationale: For people that are on chronic dialysis, discussion of patient life goals with their dialysis facility care team can lead to better understanding of these goals at the facility-, provider- and patient-levels, facilitating decision making that explicitly takes these goals into consideration and enhances patient shared decision making about modality selection, vascular access, and other treatment options for their dialysis and ESRD care. Discussing personal life goals that are important to the patient also results in patient-centered care which is an outcome in its own right (Blake and Brown 2020). This measure is intended to provide information to users about one particular aspect of patient-centered care, identification and discussion of patient life goals with those delivering care.

This new measure, the PaLS, will provide a way for providers, payers, and others to assess how well a facility is doing in identifying and discussing life goals with ESRD chronic dialysis patients as part of the treatment planning process. For example, the PaLS can provide care teams with information about important life goals that they can use to ensure that dialysis treatment modality is synergistic with a patient's specific life goals. A discussion of life goals between a patient and his/her/their provider(s) can inform modality option selection, such as a home dialysis therapy (home hemodialysis, peritoneal dialysis) or kidney transplant, and help ensure that modality selection is aligned with the patient's life goals. Results from the PaLS can also provide payers and other stakeholders national level information about patients' experience with discussions of life goals (or absence of) with their care team, and whether patients feel their treatments are aligned with their life goals.

Reference:
Blake, PG., and Brown, EA. Person-centered peritoneal dialysis prescription and the role of shared decision-making. 2020. Peritoneal Dialysis Int, published ahead of print, 1-8.

sp.12. Numerator Statement: The numerator is the number of eligible patients from the denominator that completed at least one scorable item of the PaLS (i.e., at least one of the six Likert-type items).

sp.14. Denominator Statement: All prevalent adult chronic dialysis patients (≥18 y/o) treated by the facility (both In-Center and Home Dialysis) for greater than 90 days during the reporting period, who read and understand English*.

*At present, this instrument is available to patients who read and understand English. Generalizing the survey to other languages will require additional development work.

Content
sp.16. Denominator Exclusions: Exclusions are implicit based on eligibility criteria to complete the survey. These include: • Persons under age 18 • Persons who are kidney transplant recipients with a functioning allograft • Persons who had previously been on chronic dialysis but have recovered renal function, or are lost to follow up during the reporting period • Persons with duplicate surveys – we used either the first or the more complete survey • Persons that are unable to read and/or understand English (self-assessed and self-reported)* *At present, this instrument is available to patients who read and understand English. Generalizing the survey to other languages will require additional development work
Measure Type: Process **sp.28. Data Source:** Claims, Instrument-Based Data, Registry Data **sp.07. Level of Analysis:** Other (US Chronic Dialysis Population [patient-level]. The measure testing was performed on a sample that reflected the US chronic dialysis population at the patient-level.)
IF Endorsement Maintenance—Original Endorsement Date: N/A New Measure **Most Recent Endorsement Date:** N/A New Measure
IF this measure is included in a composite, Composite#/title: N/A **IF this measure is paired/grouped, CBE#/title:** N/A **sp.03. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?** N/A Not a paired measure.
Staff Assessment: New Measure
Criterion 1: Importance to Measure and Report

Content
1a. Evidence. The evidence requirements for a ***structure, process, or intermediate outcome*** measure are that it is based on a systematic review (SR) and grading of the body of empirical evidence in which the specific focus of the evidence matches what is being measured. For measures derived from a patient report, the evidence also should demonstrate that the target population values the measured process or structure and finds it meaningful.
The developer provides the following description for this measure: • This is a new process measure at the patient-level that uses a self-report survey to assess patient satisfaction with discussion about life goals within a dialysis facility. • The developer provides the following logic model: Identification of patient life goals (Patient life goal survey) → discussion of different treatment plans (e.g., dialysis or transplant modality; vascular access type) → shared decision making → alignment of treatment plan with life goals → patient-centered care
The developer provides the following evidence for this measure: • SR of the evidence specific to this measure? ☐ Yes ☒ No • Quality, Quantity, and Consistency of evidence provided? ☐ Yes ☒ No • Evidence graded? ☐ Yes ☒ No

Summary:
• The developer conducted a scoping review to identify peer-reviewed studies, reviews, and perspective or conceptual articles examining the relationship between life goals, provider knowledge of those goals, and shared decision-making related to treatment planning.
• The developer highlighted the following evidence in support of their measure:
  ○ CMS regulations (DHHS 2008) and clinical practice guidelines place patient life goals as the cornerstone of kidney replacement treatment decision-making and care planning.
  ○ The updated National Kidney Foundation's Kidney Disease Outcomes Quality Initiative (KDOQI) Guideline Statements for vascular access, which stresses the importance of life goals identification and on-going discussions as an integral part of ESRD treatment decisions.
  ○ Studies suggesting that around 30% of ESRD patients reported feeling they were not adequately informed about treatment modality options, or that they were not the ones that made the decision.

Questions for the Standing Committee: • *What is the relationship between this measure and patient outcomes?*

Content

- *How strong is the evidence for this relationship?*
- *Does the target population value the measured outcome and find it meaningful?*

Guidance From the Evidence Algorithm
Process measure with empirical evidence submitted but not systematically reviewed (Box 3) -> empirical evidence without systematic review/grading of evidence (Box 7) -> evidence summarized include all studies (box 8) -> high certainty that benefits outweigh undesirable effects (box 9) ->Moderate

The highest possible rating is moderate.

Preliminary rating for evidence: ☐ High ☒ Moderate ☐ Low ☐ Insufficient

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

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

- The developer shared the following:
  - The measure fills the gap in assessing and promoting patient-centered care in the context of chronic dialysis for individuals with ESRD.
  - The measure will provide a way for providers, payers, and others to assess how well a facility is doing in identifying and discussing life goals with ESRD chronic dialysis patients as part of the treatment planning process.
  - Results obtained from the PaLS can provide valuable insights at a national level, giving payers and stakeholders information about patients' experience with discussions of life goals (or absence of) with their care team, and whether patients feel their treatments are aligned with their life goals.
- The developer provided performance score data for both the calibration sample and the subsequent validation testing sample.
- Calibration sample distribution of t-scores. Dates of data collection: 06/03/2020 – 12/29/2020. Number of measured entities (participants): 517.
  - Mean t-score for the overall sample is 50.0 with a standard deviation of 9.5.
  - The minimum t-score observed is 25.9, while the maximum is 69.8.
  - The lower quartile is 43.2, the upper quartile is 55.8, and the interquartile range (IQR) is 12.7.
- Validation testing sample distribution of t-scores. Dates of data collection: 04/07/2021 – 11/06/2021. Number of measured entities (participants): 420.
  - Mean t-score for the overall sample is 50.1 with a standard deviation of 9.7.
  - The minimum t-score observed is 25.9, while the maximum is 69.8.
  - The lower quartile is 43.5, the upper quartile is 56.4, and the interquartile range (IQR) is 12.9.

Disparities

Content

- The developer presents a distribution of t-scores for the calibration sample by demographic factors: race, ethnicity, sex, and education.
    - The developer reports there were minor differences in mean scores within demographic groups, including race, ethnicity (non-Hispanic participant mean scores were 2 points lower than Hispanic participant scores), sex (males had mean scores that were 0.8 points lower than females) and level of education; there was no statistically significant difference in t-scores between groups.
- The developer presents a distribution of t-scores for the validation sample by demographic factors: race, ethnicity, sex, and education.
    - The developer reports there were minor differences in mean scores within demographic groups, including race, ethnicity (non-Hispanic participant mean scores were 0.6 points lower than Hispanic participant scores), and level of education; there was no statistically significant difference in t-scores between groups.
- The developer conducted score disparity analysis for dual eligibility and reports the following:
    - Calibration sample: there was no statistically significant difference in t-scores between participants in the no dual eligibility group compared to the dual eligibility group.
        - Among the participants who were not in the dual eligible group, 16.3% had a poor PaLS score compared to 13.7% of participants that were dual eligible with a poor PaLS score.
    - Validation testing sample: there was no statistically significant difference in t-scores between participants that were not dual eligible compared to participants that were dual eligible.
        - Among the participants who were not in the dual eligible group, 17.6% had a poor PaLS score compared to 12.8% of participants that were dual eligible with a poor PaLS score.
- The developer conducted score disparity analysis for level of education and reports the following:
    - Calibration sample: there was no statistically significant difference in t-scores for participants with a 4-year college degree or more vs. less than a 4-year college degree.
        - Among participants with a 4-year college degree or more, 13.5% had a poor PaLS score compared to 16.6% of participants with less than a 4-year college degree with a poor PaLS score.
    - Validation testing sample: there was no statistically significant difference in t-scores for participants with a 4-year college degree or more vs. less than a 4-year college degree.
        - Among participants with a 4-year college degree or more, 15.6% had a poor PaLS score compared to 16.0% of participants with less than a 4-year college degree with a poor PaLS score.
- The developer notes that the results of testing for score-based disparities on the PaLS did not indicate any significant disparities in life goals for individuals with or without dual eligibility or by level of education.

Questions for the Standing Committee:

- *Is there a gap in care that warrants a national performance measure?*

Preliminary rating for opportunity for improvement:

Content
☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**
Criteria 2: Scientific Acceptability of Measure Properties
Complex measure evaluated by the Scientific Methods Panel (SMP)? ☐ **Yes** ☒ **No**
Evaluators: Staff/Laura Aume
2a. Reliability: Specifications and Testing
2a1. Specifications require the measure, as specified, to produce consistent (i.e., reliable) and credible (i.e., valid) results about the quality of care when implemented.
2a2. Reliability testing demonstrates whether the measure data elements are repeatable and producing the same results a high proportion of the time when assessed in the same population in the same time period, and/or whether the measure score is precise enough to distinguish differences in performance across providers.
Specifications: • The developer indicated the score is a rate/proportion, however Battelle staff reviewed the measure as a t-score. • The numerator and denominator descriptions are clear and concise. However, the exclusion in the denominator, "Persons with duplicate surveys – we used either the first or the more complete survey" is unclear. The developer should clarify which survey should be excluded. • Measure specifications for the instrument-based measure also include the specific instrument; standard methods, modes, and languages of administration; whether (and how) proxy responses are allowed; standard sampling procedures; handling of missing data; and the calculation of response rates to be reported with the performance measure results.
Reliability Testing: • Reliability testing conducted at the Patient/Encounter Level: o Patient/Encounter level reliability testing of the six Likert-type items was done on both the calibration sample and the validation testing sample. Examples of results are a response pattern reliability of 91%, Cronbach's alpha of 84%-85%, and a marginal reliability of 90%-91%. o Testing shows no evidence that any of the six Likert-type items are redundant. • Reliability testing conducted at the Accountable Entity Level:

Content
○ Accountable Entity reliability testing included results such as an intraclass correlation coefficient (ICC) of 80%.
Questions for the Standing Committee regarding reliability: • *Do you have any concerns that the measure cannot be consistently implemented (i.e., are the measure specifications adequate)?*
Guidance From the Reliability Algorithm Precise specifications (Box 1) -> Empirical testing reliability conducted (Box 2) -> Reliability testing conducted with computer measure scores (Box 4) -> Appropriate methods (Box 5) -> Moderate (Box 6) The highest possible rating is high.
Preliminary rating for reliability: ☐ High ☒ Moderate ☐ Low ☐ Insufficient
2b. Validity: Validity Testing; Exclusions; Risk Adjustment; Meaningful Differences; Comparability; Missing Data
2b2. Validity testing should demonstrate that 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. **Validity Testing** • Validity testing conducted at the Patient/Encounter Level: ○ The developer conducted “known groups” validity testing (dialysis modality) to assess mean score differences. ○ The developer also conducted “known groups” validity testing (HRQOL) to assess mean score differences (validation sample only). ○ In both circumstances the effects were in the anticipated direction. ○ Difference in performance scores between groups known to differ in quality performed. Examples include: • Difference between dialysis modalities (in-center and peritoneal versus home) is statistically significant with 95% confidence for the calibration dataset ($p<0.0001$) but not for the validation dataset ($p=0.07$). • Differences in several PROMIS measures were all statistically significant with 95% confidence in the validation dataset. ○ Analysis of floor and ceiling effects show that only a small percentage of participants selected the very worst or very best responses. ○ Convergent and discriminant validity testing shows that, as expected, the measure correlates with PROMIS

Content
Meaning and Purpose scores and does not correlate with various other PROMIS scores.
Exclusions • The measure does not use exclusions.
Risk Adjustment • The measure is not risk-adjusted or stratified.
Meaningful Differences • The developer examined changes in scores from baseline to 3-months. • As hypothesized the scores changed over time. However, PROMIS measures scores were not predictive of change.
Missing Data • Missing data were reported to be infrequent. The percentage of patients who missed each of the six Likert-type items was calculated and considered to be negligible and should not impact the overall t-score.
Comparability • The measure only uses one set of specifications for this measure.
Questions for the Standing Committee regarding validity: • *Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk adjustment approach, etc.)?*
Guidance From the Validity Algorithm All threats assessed (Box 1) -> Empirical validity testing not conducted on the measure as specified (Box 2) -> Face validity testing not assessed to determine agreement on whether the computed measure score can be used to distinguish good and poor quality (Box 3) -> Insufficient. The information provided in the testing section appears to be related to testing the instrument rather than the computed performance measure score.
Preliminary rating for validity: ☐ **High** ☐ **Moderate** ☐ **Low** ☒ **Insufficient** **Rationale:** Although the developer provides testing results for the instrument there is little empirical demonstration for the performance measure based on the instrument.
Criterion 3. Feasibility
3. <u>Feasibility</u> 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. • Data elements are generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score).

Content

- Data elements needed to compute the performance measure score are from patient/family reported information (may be electronic or paper).

Questions for the Standing Committee:

- *Is the data collection strategy ready to be put into operational use?*

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

Criterion 4: Use and Usability

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

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

4a1. 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 they are not in use at the time of initial endorsement, then a credible plan for implementation within the specified time frames is provided.

Current uses of the measure

Publicly reported?	☐ Yes	☒ No	
Current use in an accountability program?	☐ Yes	☒ No	☐ UNCLEAR
Planned use in an accountability program?	☒ Yes	☐ No	☐ N/A

Accountability program details

- The developer noted that this measure is not yet in use because it is a new measure.
- The developer states that CMS will determine if/when to use this measure in a public reporting/payment program.
- The developer shared potential applications for the measure in the ESRD Quality Incentive Program (ESRD QIP) or the Dialysis Facility Care Compare program on Medicare.gov.

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; and (3) This feedback has been considered when changes are incorporated into the measure.

Content
Feedback on the measure provided by those being measured or others • The developer notes that the measure has not yet been implemented and results have not been disseminated to those being measured as part of the development process.
Questions for the Standing Committee: • *How have (or can) the performance results be used to further the goal of high quality, efficient healthcare?* • *How has the measure been vetted in real-world settings by those being measured or others?*
Preliminary rating for Use: ☒ Pass ☐ No Pass
4b. Usability (4b1. Improvement; 4b2. Benefits of measure)
4b. Usability evaluates the extent to which audiences (e.g., consumers, purchasers, providers, and policymakers) use or could use performance results for both accountability and performance improvement activities.
4b1 Improvement. Progress toward achieving the goal of high quality, efficient healthcare for individuals or populations is demonstrated. **Improvement results** • The developer notes the measure is not yet implemented in a public reporting program, so improvement could not be evaluated.
4b2. Benefits versus harms. The 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** • Not applicable; the measure has not yet been implemented. **Potential harms** • None identified, as the measure is not yet implemented.
Questions for the Standing Committee: • *How can 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?*

Content
Preliminary rating for Usability: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**
Criterion 5: Related and Competing Measures
Related/Competing Measures • None identified.
Harmonization • N/A

QUALITY MEASURE SUBMISSION FORM

Version: 1.0; Generated: 13 April 2023

Introduction

Thank you for your interest in submitting a measure to Battelle for possible endorsement.

What criteria are used to evaluate measures? Measures are evaluated on standardized criteria: importance to measure and report, scientific acceptability of measure properties, feasibility, usability and use, and related and competing measures. For your measure to be evaluated against these measure evaluation criteria, you must complete the measure submission form.

Why do I have to complete a form? Due to the volume and/or complexity of proposed measures, Battelle provides measure information to committee reviewers in a standardized format to facilitate their evaluation of whether the measure meets the measure evaluation criteria. This form allows the measure steward to present information demonstrating that the proposed measure meets endorsement criteria.

What is on the form? The information requested in this form is directly related to the measure evaluation criteria.

Can't I just submit our files for consideration? No. Measures must be submitted through the online form to be considered for the Spring 2023 cycle. Requested information should be entered directly into this form and as well as any necessary or required attachments.

Can I submit additional details and materials? Additional materials will be considered only as supplemental. Do NOT rely on material provided in an appendix to provide measure specifications or to demonstrate meeting the criteria. The core information needed to evaluate the measure should be provided in the appropriate submission form fields and required attachments. Please contact PQMsupport@battelle.org regarding questions about submitting supplemental materials.

What do I do first? If you have started a new submission by answering five qualifying questions, you may proceed to the “Previous Submission Information" tab to continue with your submission. The “Conditions” tab will list the conditions that must be met before your proposed measures may be considered and evaluated for suitability as endorsed voluntary consensus standards. You are asked to acknowledge reading and accepting the conditions.

Can I make changes to a form once I have submitted it? No. Once you submit your measure, you will NOT be able to return to this submission form to make further revisions. You will need to contact project staff.

What if I need additional help? Please contact the project staff at PQMsupport@battelle.org if you have questions regarding the information requested or submitting supplemental materials.

NOTE: All measure submissions should be 508-compliant. Refer to the Checklist for Developer 508 Guidelines (PDF) to ensure all guidelines apply to all parts of your submission, including all fields and attachments used within the measure submission form.

Please email us at PQMsupport@battelle.org if you experience technical difficulties using the online submission form.

Thank you for your interest in submitting measures to Battelle.

Previous Submission Information (1 – 4)

1) Select whether this measure was previously submitted to the prior consensus-based entity (the National Quality Forum [NQF]) and given an identifying number.

☐ Previously submitted to NQF
☒ New measure, never submitted.

2) Provide the measure number of the previously submitted measure.
N/A

3) If the measure has an electronic clinical quality measure (eCQM) version, provide the measure number of the previously submitted measure.
N/A

4) If this eCQM has a registry version, provide the measure numbers of the previously submitted measure.
N/A

Conditions (1 - 2)

Several conditions must be met before proposed measures may be considered and evaluated for suitability as voluntary consensus standards. If any of the conditions are not met, the measure will not be accepted for consideration.

A. A Measure Steward Agreement is signed or the steward is a government organization. (All non-government organizations must sign a Measure Steward Agreement.) For more information about completing a Measure Steward Agreement, please go to: Endorsement | Partnership for Quality Measurement (p4qm.org) and follow the instructions.

B. The measure owner/steward verifies there is an identified responsible entity and a process to maintain and update the measure on a schedule that is commensurate with the rate of clinical innovation, but at least every three years.

C. The intended use of the measure includes both accountability applications (including public reporting) and performance improvement to achieve high-quality, efficient healthcare.

D. The measure is fully specified and tested for reliability and validity.

E. The measure developer/steward attests that harmonization with related measures and issues with competing measures have been considered and addressed, as appropriate.

F. The requested measure submission information is complete and responsive to the questions so that all the information needed to evaluate all criteria is provided.

1) Check if either of the following apply.

☐ Proprietary measure or components (e.g., risk model, codes)
☐ Proprietary measure or components with fees
☒ None of the above

2) Check the box below to agree to the conditions listed above.

☒ I have read and accept the conditions as specified above

Specifications: Maintenance Update (spma.01 - spma.02)

spma.01) Indicate whether there are changes to the specifications since the last updates/submission. If yes, update the specifications in the Measure Specifications section of the Measure Submission Form, and explain your reasoning for the changes below.

☒ No
☐ Yes

spma.02) Briefly describe any important changes to the measure specifications since the last measure update and provide a rationale.

For annual updates, please explain how the change in specifications affects the measure results. If a material change in specification is identified, data from re-testing of the measure with the new specifications is required for early maintenance review.

For example, specifications may have been updated based on suggestions from a previous measure endorsement review.

Measure Specifications (sp.01 - sp.32)

sp.01) Provide the measure title.

Measure titles should be concise yet convey who and what is being measured.

ESRD Dialysis Patient Life Goals Survey (PaLS)

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).

The PaLS is a patient self-report survey that includes eight items related to dialysis facility care team discussions about patient life goals. Six of the items are Likert-type items that are used to generate a "quality of facility care team discussion" score (described below). The remaining two items on the PaLS are checklist items: (1) a list of patient-reported life goals; and (2) a patient-reported list of dialysis care team members that the patient reports has talked with them about their life goals. These items are not scored. Instead, these items serve to provide contextual information for both the patient and the facility to guide care team discussions.

The PaLS is used to generate a patient-level *t*-score that reflects patient-reported satisfaction with how well his/her/their facility is doing in discussing life goals with the patient as part of the treatment planning process. For each individual patient at a given facility, the calculated *t*-score (M=50; SD=10) represents a patient's perceptions of their satisfaction with their dialysis care team discussions about life goals. A *t*-score greater than 40 and less than 60 reflects a score that is within normal limits of existing practices and should not warrant further action, assuming typical existing practices for patient and care team discussions are deemed adequate. Assuming a normal distribution, scores that are ≤40 would warrant follow-up by the facility. Specifically, scores ≤40 (i.e., ≥1 *SD* below the mean) suggest patient perceptions of care discussions are worse than 84% of their peers, whereas scores ≤30 (i.e., ≥2 *SD*s below the mean) suggest patient perceptions of care discussions are worse than 98% of their peers; scores that are >40 would be within "normal limits" (Heaton et al., 2004). The *t*-score is based on the data collected for the instrument testing, as described in the section on scientific acceptability, but is currently not part of the process measure calculation.

The target population for the measure is patients on chronic dialysis who meet all of the following criteria:

- Are at least 18 years old
- Completed the PaLS survey at least once during the one-year reporting period

Reference:

Heaton, R. K., Miller, S. W., Taylor, J. T., & Grant, I. (2004). *Revised comprehensive norms for an expanded Halstead-Reitan Battery: Demographically adjusted neuropsychological norms for African American and Caucasian adults*. Lutz, FL: Psychological Assessment Resources, Inc.

sp.03) Provide a rationale for why this measure must be reported with other measures to appropriately interpret results.

N/A

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

☐ Behavioral Health
☐ Behavioral Health: Alcohol, Substance Use/Abuse
☐ Behavioral Health: Anxiety
☐ Behavioral Health: Attention Deficit Hyperactivity Disorder (ADHD)
☐ Behavioral Health: Bipolar Disorder
☐ Behavioral Health: Depression
☐ Behavioral Health: Domestic Violence
☐ Behavioral Health: Other Serious Mental Illness
☐ Behavioral Health: Post-Traumatic Stress Disorder (PTSD)
☐ Behavioral Health: Schizophrenia
☐ Behavioral Health: Suicide
☐ Cancer
☐ Cancer: Bladder
☐ Cancer: Breast
☐ Cancer: Colorectal
☐ Cancer: Gynecologic
☐ Cancer: Hematologic
☐ Cancer: Liver
☐ Cancer: Lung, Esophageal
☐ Cancer: Prostate
☐ Cancer: Renal
☐ Cancer: Skin
☐ Cancer: Thyroid
☐ Cardiovascular
☐ Cardiovascular: Arrythmia
☐ Cardiovascular: Congestive Heart Failure
☐ Cardiovascular: Coronary Artery Disease
☐ Cardiovascular: Coronary Artery Disease (AMI)
☐ Cardiovascular: Coronary Artery Disease (PCI)
☐ Cardiovascular: Hyperlipidemia
☐ Cardiovascular: Hypertension
☐ Cardiovascular: Secondary Prevention

- ☐ Critical Care
- ☐ Critical Care: Assisted Ventilation
- ☐ Critical Care: Intensive Monitoring
- ☐ Dental
- ☐ Dental: Caries
- ☐ Dental: Tooth Loss
- ☐ Ears, Nose, Throat (ENT)
- ☐ Ears, Nose, Throat (ENT): Ear Infection
- ☐ Ears, Nose, Throat (ENT): Hearing
- ☐ Ears, Nose, Throat (ENT): Pharyngitis
- ☐ Ears, Nose, Throat (ENT): Tonsilitis
- ☐ Endocrine
- ☐ Endocrine: Calcium and Metabolic Bone Disorders
- ☐ Endocrine: Diabetes
- ☐ Endocrine: Female and Male Endocrine Disorders
- ☐ Endocrine: Hypothalamic-Pituitary Disorders
- ☐ Endocrine: Thyroid Disorders
- ☐ Eye Care
- ☐ Eye Care: Age-related macular degeneration (AMD)
- ☐ Eye Care: Cataracts
- ☐ Eye Care: Diabetic retinopathy
- ☐ Eye Care: Glaucoma
- ☐ Gastrointestinal (GI)
- ☐ Gastrointestinal (GI): Constipation
- ☐ Gastrointestinal (GI): Gall Bladder Disease
- ☐ Gastrointestinal (GI): Gastroenteritis
- ☐ Gastrointestinal (GI): Gastro-Esophageal Reflux Disease (GERD)
- ☐ Gastrointestinal (GI): Hemorrhoids
- ☐ Gastrointestinal (GI): Hernia
- ☐ Gastrointestinal (GI): Inflammatory Bowel Disease
- ☐ Gastrointestinal (GI): Irritable Bowel Syndrome
- ☐ Gastrointestinal (GI): Peptic Ulcer
- ☐ Genitourinary (GU)
- ☐ Genitourinary (GU): Benign Prostatic Hyperplasia
- ☐ Genitourinary (GU): Erectile Dysfunction/Premature Ejaculation
- ☐ Genitourinary (GU): Incontinence/pelvic floor disorders
- ☐ Genitourinary (GU): Prostatitis
- ☐ Genitourinary (GU): Urinary Tract Injection (UTI)
- ☐ Gynecology (GYN)
- ☐ Gynecology (GYN): Abnormal bleeding

- ☐ Gynecology (GYN): Endometriosis
- ☐ Gynecology (GYN): Infections
- ☐ Gynecology (GYN): Menopause
- ☐ Gynecology (GYN): Pelvic Pain
- ☐ Gynecology (GYN): Uterine fibroids
- ☐ Infectious Diseases (ID)
- ☐ Infectious Diseases (ID): HIV/AIDS
- ☐ Infectious Diseases (ID): Influenza
- ☐ Infectious Diseases (ID): Lyme Disease
- ☐ Infectious Diseases (ID): Meningococcal Disease
- ☐ Infectious Diseases (ID): Pneumonia and respiratory infections
- ☐ Infectious Diseases (ID): Sepsis
- ☐ Infectious Diseases (ID): Sexually Transmitted
- ☐ Infectious Diseases (ID): Tuberculosis
- ☐ Liver
- ☐ Liver: Viral Hepatitis
- ☐ Musculoskeletal
- ☐ Musculoskeletal: Falls and Traumatic Injury
- ☐ Musculoskeletal: Gout
- ☐ Musculoskeletal: Joint Surgery
- ☐ Musculoskeletal: Low Back Pain
- ☐ Musculoskeletal: Osteoarthritis
- ☐ Musculoskeletal: Osteoporosis
- ☐ Musculoskeletal: Rheumatoid Arthritis
- ☐ Neurology
- ☐ Neurology: Alzheimer's Disease
- ☐ Neurology: Autism
- ☐ Neurology: Brain Injury
- ☐ Neurology: Epilepsy
- ☐ Neurology: Migraine
- ☐ Neurology: Parkinson's Disease
- ☐ Neurology: Spinal Cord Injury
- ☐ Neurology: Stroke/Transient Ischemic Attack (TIA)
- ☐ Other (please specify here:)
- ☐ Palliative Care and End-of-Life Care
- ☐ Palliative Care and End-of-Life Care: Advanced Directives
- ☐ Palliative Care and End-of-Life Care: Amyotrophic Lateral Sclerosis (ALS)
- ☐ Palliative Care and End-of-Life Care: Hospice Management
- ☐ Palliative Care and End-of-Life Care: Inappropriate use of acute care services
- ☐ Palliative Care and End-of-Life Care: Pain Management

☐ Perinatal Health
☐ Perinatal Health: Labor and Delivery
☐ Perinatal Health: Newborn Care
☐ Perinatal Health: Post-Partum Care
☐ Perinatal Health: Preconception Care
☐ Perinatal Health: Prenatal Care
☒ Renal
☐ Renal: Acute Kidney Injury
☐ Renal: Chronic Kidney Disease (CKD)
☒ Renal: End Stage Renal Disease (ESRD)
☐ Renal: Infections
☐ Reproductive Health
☐ Reproductive Health: Family planning and contraception
☐ Reproductive Health: Infertility
☐ Reproductive Health: Male reproductive health
☐ Respiratory
☐ Respiratory: Acute Bronchitis
☐ Respiratory: Allergy
☐ Respiratory: Asthma
☐ Respiratory: Chronic Obstructive Pulmonary Disease (COPD)
☐ Respiratory: Dyspnea
☐ Respiratory: Pneumonia
☐ Respiratory: Sleep Apnea
☐ Surgery
☐ Surgery: Cardiac Surgery
☐ Surgery: Colorectal
☐ Surgery: Neurosurgery / Spinal
☐ Surgery: Orthopedic
☐ Surgery: Orthopedic Hip/Pelvic Fractures
☐ Surgery: Pediatric
☐ Surgery: Perioperative and Anesthesia
☐ Surgery: Plastic
☐ Surgery: Thoracic Surgery
☐ Surgery: Trauma
☐ Surgery: Vascular Surgery

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

☐ Access to Care
☐ Care Coordination

☐ Care Coordination: Readmissions
☐ Care Coordination: Transitions of Care
☐ Disparities Sensitive
☐ Health and Functional Status
☐ Health and Functional Status: Change
☐ Health and Functional Status: Nutrition
☐ Health and Functional Status: Obesity
☐ Health and Functional Status: Physical Activity
☐ Health and Functional Status: Quality of Life
☐ Health and Functional Status: Total Health
☐ Immunization
☐ Other (please specify here:)
☒ Person-and Family-Centered Care: Person-and Family-Centered Care
☐ Person-and Family-Centered Care: Workforce
☐ Primary Prevention
☐ Primary Prevention: Nutrition
☐ Primary Prevention: Tobacco Use
☐ Safety
☐ Safety: Complications
☐ Safety: Healthcare Associated Infections
☐ Safety: Medication
☐ Safety: Overuse
☐ Screening

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.

☒ Adults (Age >= 18)
☐ Children (Age < 18)
☒ Elderly (Age >= 65)
☒ Populations at Risk: Dual eligible beneficiaries of Medicare and Medicaid
☒ Populations at Risk: Individuals with multiple chronic conditions
☐ Populations at Risk: Veterans
☐ Women

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.

☐ Accountable Care Organization

☐ Clinician: Group/Practice
☐ Clinician: Individual
☐ Facility
☐ Health Plan
☐ Integrated Delivery System
☒ Other (please specify here: US Chronic Dialysis Population (patient-level). The measure testing was performed on a sample that reflected the US chronic dialysis population at the patient-level.)
☐ Population: Community, County or City
☐ Population: Regional and State

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

Check ONLY the settings for which the measure is SPECIFIED and TESTED.
☐ Ambulatory Care
☐ Behavioral Health
☐ Home Care
☐ Inpatient/Hospital
☐ Other (please specify here:)
☒ Outpatient Services
☐ Post-Acute Care

sp.09) Provide a Uniform Resource Locator (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".

None available.

sp.10) Indicate whether Health Quality Measure Format (HQMF) specifications are attached.

Attach the zipped output from the measure authoring tool (MAT) for eCQMs - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications). HQMF specifications are attached.

☐ HQMF specifications are NOT attached (Please explain).

N/A

sp.11) Attach the simulated testing attachment.

All eCQMs require a simulated testing attachment to confirm that the HTML output from Bonnie testing (or testing of some other simulated data set) includes 100% coverage of measured patient population testing, with pass/fail test cases for each sub-population. This can be submitted in the form of a screenshot.

☐ Testing is attached
☐ Testing is NOT attached (please explain)

N/A

sp.12) 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 at PQMsupport@battelle.org. Provide descriptors for any codes. Use one file with multiple worksheets, if needed.

☐ Available in attached Excel or csv file
☒ No data dictionary/code table – all information provided in the submission form

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

sp.13) 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.

The numerator is the number of eligible patients from the denominator that completed at least one scorable item of the PaLS (i.e., at least one of the six Likert-type items).

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.14) 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.

We begin with the number of patients that took the PaLS survey and completed at least one of the six Likert-type scorable PaLS items that comprise the "quality of facility care team discussions" score. The response options for these six items are scored from 1 to 5. Higher scores indicate greater overall patient reported agreement that the care team is asking about and discussing life goals with the patient. IRT scores are initially estimated on the theta metric (M=0; SD=1). In order to enhance the clinical utility of our PaLS measure, we converted theta scores to standardized scores on the *t*-score metric (M=50; SD=10). The conversion from a theta score to a *t*-score can be made using the following linear transformation: *t*-score=(theta x 10)+50. This patient-level *t*-score represents a patient's perceptions about how well the facility is doing in discussing life goals as part of the treatment planning process.

Although missing PaLS responses are allowed, patients must answer at least one of the six Likert-type scorable PaLS items to receive a *t*-score.

The *t*-score is based on the data collected for the instrument testing, as described in the scientific acceptability, but is currently not part of the process measure calculation.

The numerator is comprised of the number of eligible patients from the denominator who completed at least one Likert-type scorable item of the PaLS.

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

sp.15) State the denominator.

Brief, narrative description of the target population being measured.

All prevalent adult chronic dialysis patients (≥18 y/o) treated by the facility (both In-Center and Home Dialysis) for greater than 90 days during the reporting period, who read and understand English*.

*At present, this instrument is available to patients who read and understand English. Generalizing the survey to other languages will require additional development work.

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

sp.16) 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.

To be in the denominator, chronic dialysis patients at the facility must be eligible to complete the PaLS; that is, they must be (a) at least 18 years of age; (b) receiving long-term dialysis in the United States or any U.S. Territory for greater than 90 days during the reporting period; and (c) able to read and understand English (self-assessed and reported). Receiving long-term dialysis in the 90 day period was selected in order to allow time for the patient to stabilize after beginning chronic dialysis, and for the dialysis care team to initiate discussions about patient life goals as part of the treatment planning process. This 90 day period also reduces facility-related burden. At present, this instrument is available to patients who read and understand English. Generalizing the survey to other languages will require additional development work.

To construct our denominator for testing, we used the following self-report data from survey participants: first name, last name, sex, birthdate, last four digits of their social security number (SSN), race, ethnicity, and level of education completed. The first four of these data elements were required; the last four elements participants could elect to not report. Using self-reported first name, last name, last four digits of SSN (if provided), and birthdate, participants were then matched to our ESRD database, which contains treatment history data on all U.S. ESRD patients. We used CMS administrative data to confirm dialysis modality for participants linked to the UM-KECC ESRD database (in-center hemodialysis, home hemodialysis, peritoneal dialysis, or kidney transplant). In some cases, we could not match participants to their data in the UM-KECC ESRD database (i.e., if self-reported first or last name, birthdate, sex, or last four SSN digits were either missing, illegible or incomplete). In these cases, participants were not included in the analysis using dialysis modality.

We implemented two different field-testing data collection efforts as part of our measurement development process, which we refer to hereafter as: 1) the calibration sample; and 2) the validation testing sample. For the calibration sample, 10.4% of participants were not able to be matched to the ESRD database. For the validation testing sample, 20.2% of participants were not able to be matched to the ESRD database.

sp.17) Describe the denominator exclusions.

Brief narrative description of exclusions from the target population.

Exclusions are implicit based on eligibility criteria to complete the survey. These include:

- Persons under age 18
- Persons who are kidney transplant recipients with a functioning allograft
- Persons who had previously been on chronic dialysis but have recovered renal function, or are lost to follow up during the reporting period
- Persons with duplicate surveys – we used either the first or the more complete survey

- Persons that are unable to read and/or understand English (self-assessed and self-reported)*

*At present, this instrument is available to patients who read and understand English. Generalizing the survey to other languages will require additional development work.

sp.18) 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.

To be in the denominator, chronic dialysis patients at the facility must be eligible to complete the PaLS; that is, they must be (a) at least 18 years of age; (b) receiving long-term dialysis in the United States or any U.S. Territory for greater than 90 days during the reporting period; (c) able to read and understand English (self-assessed and reported). Receiving long-term dialysis in the 90 day period was selected in order to allow time for the patient to stabilize after beginning chronic dialysis, and for the dialysis care team to initiate discussions about patient life goals as part of the treatment planning process. This 90 day period also reduces facility-related burden.

Again, at present, this instrument is available to patients who read and understand English. Generalizing the survey to other languages will require additional development work.

For our testing (see sp.15, above) we used CMS administrative data to confirm patients were ESRD and on a chronic dialysis modality.

Exclusions are implicit based on eligibility criteria to complete the survey. These include age less than 18; patient has a kidney transplant; patient with recovered renal function, or lost to follow up; and unable to read and/or understand English (whether self-assessed or self-reported). In our testing we also excluded duplicate patient surveys.

sp.19) 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.

N/A

sp.20) Is this measure adjusted for socioeconomic status (SES)?

☐ Yes
☒ No

sp.21) Select the risk adjustment type.

Select type. Provide specifications for risk stratification and/or risk models in the Scientific Acceptability section.

☒ No risk adjustment or risk stratification
☐ Statistical risk model
☐ Stratification by risk category/subgroup (specify number of risk factors)
☐ Other approach to address risk factors (please specify here:)

sp.22) Select the most relevant type of score.

Attachment: If available, please provide a sample report.

☐ Categorical, e.g., yes/no
☐ Continuous variable, e.g. average
☐ Count
☐ Frequency Distribution
☐ Non-weighted score/composite/scale
☐ Other (please specify here:)
☒ Rate/proportion
☐ Ratio
☐ Weighted score/composite scale

sp.23) 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.

☒ Better quality = Higher score
☐ Better quality = Lower score
☐ Better quality = Score within a defined interval
☐ Passing score defines better quality

sp.24) 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.

The response options for the six Likert-type scorable PaLS items range from “strongly agree” (5) to “strongly disagree” (1) for three items, and from “always” (5) to “never” (1) for the other three items. Response pattern scoring was applied to item responses, using the measure’s established item parameters.

See attached flowchart.

sp.25) Attach a copy of the instrument (e.g. survey, tool, questionnaire, scale) used as a data source for your measure, if available.

☒ Copy of instrument is attached.
☐ Copy of instrument is NOT attached (please explain).

sp.26) Indicate the responder for your instrument.

☒ Patient
☐ Family or other caregiver
☐ Clinician
☐ Other (specify)

sp.27) If measure testing is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.

Examples of samples used for testing:

• Testing may be conducted on a sample of the accountable entities (e.g., hospital, physician). The analytic unit specified for the particular measure (e.g., physician, hospital, home health agency) determines the sampling strategy for scientific acceptability testing.

• The sample should represent the variety of entities whose performance will be measured. The samples used for reliability and validity testing often have limited generalizability because measured entities volunteer to participate. Ideally, however, all types of entities whose performance will be measured should be included in reliability and validity testing.

• The sample should include adequate numbers of units of measurement and adequate numbers of patients to answer the specific reliability or validity question with the chosen statistical method.

• When possible, units of measurement and patients within units should be randomly selected.

Snowball sampling was used to obtain the calibration and validation testing samples. Participants received materials via web link or paper survey packets. Sample size requirements were based on IRT-related analyses (i.e., graded response model [GRM] analyses and differential item functioning [DIF] analyses). Sample size requirements for use of GRM analyses have been estimated to be between 200 and 1000, with larger sample sizes producing more stable parameter estimates (Muraki, 1990; Samejima, 1969; Samejima et al., 1996). For DIF analyses using lordif, a sample size should be at least n=200 participants per DIF factor sub-group of interest (Clauser & Hambleton, 1994).

References:

Clauser, B.E., & Hambleton, R.K. (1994). Review of differential item functioning. *Journal of Educational Measurement, 31*(1), 88-92.

Muraki, E. (1990). Fitting a polytomous item response model to Likert-type data. *Applied Psychological Measurement, 14*(1), 59-71.

Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores (Psychometric Monograph No. 17)*. Richmond, VA: Psychometric Society.

Samejima, F., van der Liden, W.J., & Hambleton, R. (1996). The graded response model. In: W.J. van der Liden (Ed.), *Handbook of modern item response theory (pp. 85-100)*. Springer.

sp.28) Identify whether and how proxy responses are allowed.

Proxy responses are not allowed.

sp.29) Survey/Patient-reported data.

Provide instructions for data collection and guidance on minimum response rate. Specify calculation of response rates to be reported with performance measure results.

ESRD patients needed to answer at least one of the six Likert-type scorable PaLS items to receive a patient-level score. We were not able to calculate a facility response rate, given that data collection and testing were performed at the patient-level. Prior to possible implementation at the dialysis facility level, the response rate will need to be calculated at the facility level. The facility-level response rate should be calculated by dividing the number of patients who answer at least one scorable PaLS item (i.e., one of the six Likert-type items from the PaLS) by the number of patients who are eligible for the survey and complete its required demographic items.

sp.30) Select only the data sources for which the measure is specified.

☐ Assessment Data
☒ Claims
☐ Electronic Health Data
☐ Electronic Health Records
☒ Instrument-Based Data

☐ Management Data
☐ Other (please specify here:)
☐ Paper Medical Records
☒ Registry Data

sp.31) 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.

To identify ESRD patients, we used the following self-report data from survey participants: first name, last name, sex, birthdate, last four digits of their social security number (SSN), race, ethnicity, and level of education completed. The first four data elements were required; the last four elements participants could elect to not report. Using self-reported first name, last name, last four digits of SSN (if provided), and birthdate, participants were then matched to our own ESRD database, which contains treatment history data on all U.S. ESRD patients. We used CMS administrative data to confirm dialysis modality for participants linked to our database (in-center hemodialysis, home hemodialysis, peritoneal dialysis, or kidney transplant). In some cases, we could not match participants to their data in the ESRD database (if self-reported first or last name, birthdate, sex, or last four SSN digits were missing, illegible or incomplete). In these cases, participants were not included in the analysis using dialysis modality.

ESRD data were used to confirm self-reported ESRD status and treatment information. These data were derived from the UM-KECC database, a national ESRD patient database that includes information from the Renal Management Information System (REMIS), CROWNWeb facility-reported clinical and administrative data (including CMS-2728 Medical Evidence Form, CMS-2746 Death Notification Form, and CMS-2744 Annual Facility Survey Form and patient tracking data), the Medicare Enrollment Database (EDB), and Medicare dialysis claims data (primarily outpatient). In addition, the UM-KECC database includes transplant data from the Scientific Registry of Transplant Recipients (SRTR), data from the Nursing Home Minimum Dataset, data from the Quality Improvement Evaluation System (QIES) Business Intelligence Center (QBIC; which includes Provider and Survey and Certification data from Automated Survey Processing Environment [ASPEN]), and data from the Dialysis Facility Care Compare (DFCC).

sp.32) Provide the data collection instrument.

☐ Available at measure-specific web page URL identified in sp.09
☒ Available in attached appendix in Question 1 of the Additional Section
☐ No data collection instrument provided

Importance to Measure and Report: Maintenance of Endorsement (1ma.01)

1ma.01) Indicate whether there is new evidence about the measure since the most recent maintenance evaluation. If yes, please briefly summarize the new evidence, and ensure you have updated entries in the Evidence section as needed.

☐ Yes
☐ No

New measure, never submitted.

Importance to Measure and Report: Evidence (Complete for Outcome Measures) (1a.01 - 1a.03)

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:

Current Submission:

Updated evidence information here.

Previous (Year) Submission:

Evidence from the previous submission here.

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.

1a.02) Provide evidence that the target population values the measured outcome, process, or structure and finds it meaningful.

Describe how and from whom input was obtained.

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

Importance to Measure and Report: Evidence (Complete for Process Measures) (1a.03 - 1a.16)

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:

Current Submission:

Updated evidence information here.

Previous (Year) Submission:

Evidence from the previous submission here.

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.

For people that are on chronic dialysis, discussion of patient life goals with their dialysis facility care team can lead to better understanding by facilities and providers of those life goals, facilitating decision making that explicitly takes these goals into consideration and enhances patient shared decision making about modality selection, vascular access, and other treatment options for their dialysis and ESRD care. Discussing personal life goals that are important to the patient also results in patient-centered care which is an outcome in its own right (Blake and Brown 2020).

This measure is intended to facilitate discussions about life goals between the patient and their dialysis care team in order to increase provider awareness of these goals and to support patient decisions about modality (dialysis and transplant), vascular access type, and other aspects of treatment that can be tailored to align with the patient's life goals.

Identification of patient life goals (Patient life goal survey) → discussion of different treatment plans (e.g., dialysis or transplant modality; vascular access type) → shared decision making → alignment of treatment plan with life goals → patient-centered care

Reference:
Blake, PG., and Brown, EA. Person-centered peritoneal dialysis prescription and the role of shared decision-making. 2020. Peritoneal Dialysis Int, published ahead of print, 1-8.

1a.02) Select the type of source for the systematic review of the body of evidence that supports the performance measure.

A systematic review is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies. It may include a quantitative synthesis (meta-analysis), depending on the available data.

☐ Clinical Practice Guideline recommendation (with evidence review)
☐ US Preventive Services Task Force Recommendation
☐ Other systematic review and grading of the body of evidence (e.g., Cochrane Collaboration, AHRQ Evidence Practice Center)
☒ Other (please specify here: literature review)

If the evidence is not based on a systematic review, skip to the end of the section and do not complete the repeatable question group below. If you wish to include more than one systematic review, you may add additional tables to the relevant sections. Please follow the 508 Checklist for tables.

Evidence - Systematic Reviews Table (Repeatable)

1a.03) Provide the title, author, date, citation (including page number) and URL for the systematic review.

1a.04) Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the systematic review.

1a.05) Provide the grade assigned to the evidence associated with the recommendation and include the definition of the grade.

1a.06) Provide all other grades and definitions from the evidence grading system.

1a.07) Provide the grade assigned to the recommendation, with definition of the grade.

1a.08) Provide all other grades and definitions from the recommendation grading system.

1a.09) Detail the quantity (how many studies) and quality (the type of studies) of the evidence.

1a.10) Provide the estimates of benefit, and consistency across studies.

1a.11) Indicate what, if any, harms were identified in the study.

1a.12) Identify any new studies conducted since the systematic review, and indicate whether the new studies change the conclusions from the systematic review.

Evidence

1a.13) If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, describe the evidence on which you are basing the performance measure.

We conducted a scoping review to identify peer-reviewed studies, reviews, and perspective or conceptual articles examining the relationship between life goals, provider knowledge of those goals, and shared decision-making related to treatment planning.

1a.14) Briefly synthesize the evidence that supports the measure.

In medicine, there is an ethical imperative to incorporate patient goals and personal values into therapeutic decisions. This is especially true in the care provided to people with end-stage kidney failure, a condition that has the potential to significantly impact quality life and survival. As part of facilitating high quality ethical treatment, clinicians must consider patient life goals in order to optimize the therapeutic effects of selected treatments. Identification and incorporation of patient life goals into treatment planning results in treatment that is truly patient-centered. CMS regulations (DHHS 2008) and clinical practice guidelines place patient life goals as the cornerstone of kidney replacement treatment decision-making and care planning (Lok et al., 2020; Chan et al., 2019). Life goals are specifically addressed in the updated National Kidney Foundation's Kidney Disease Outcomes Quality Initiative (KDOQI) Guideline Statements for vascular access: "*The ESKD Life-Plan is a strategy that should start in the pre-dialysis period and encompasses a continuum-of-care model for CKD to ESKD. It aims to maximize ESKD modality choices and utilization for a specific patient's foreseeable lifespan and specifically considers the patient's current medical situation,* ***current and future life goals****, preferences, social support, functional status, and logistics and other practical feasibilities*" (Lok et al., 2020):

> 1.1 KDOQI considers it reasonable that each patient with progressive CKD and/or with an eGFR 15-20 mL/min/1.73 m2 or already on kidney replacement therapy should have an individualized ESKD Life-Plan that is regularly reviewed, updated, and documented on their medical record. (Expert Opinion)
>
> 1.2 KDOQI considers it reasonable to conduct an annual review and update of each patient's individualized ESKD Life-Plan, together with their health care team. (Expert Opinion)

The specific Guideline statements are based on Expert Opinion. KDOQI's exposition of these statements highlights the importance of life goals identification and on-going discussions as an integral part of ESRD treatment decisions.

Despite this emphasis on the importance of life goals discussions to inform kidney replacement modality and other treatment decisions (Lu and Chai 2021; Dorough et al., 2020; Chan et al., 2019), these conversations are not always happening (University of Michigan-KECC 2017). Certain decisions about dialysis/renal replacement therapy rely on the identification and incorporation of patient life goals into those decisions in order to support optimal patient-centered and programmatic outcomes. For example, the low uptake of home dialysis and

access to transplant in the U.S. suggests the current approach to inform patient treatment decisions may be inadequate. This has resulted in a need for a paradigm shift in the delivery of pre-dialysis and dialysis care (Chan et al., 2019). Moreover, several studies include findings that suggest that around 30% of ESRD patients did not feel like they were provided adequate information about treatment modality options, or that they were not the ones that made the decision (Dahlerus 2016; Ladin 2016; Van Biesen 2014).

The evidence described here includes empirical studies in the ESRD population as well as other chronic disease populations like cancer, diabetes, chronic pain (e.g., Dorough 2020; de Vries 2017; Ladin 2016; Dahlerus 2016; Davison 2006; Huang 2005). In each, the focus is on life goals, patient preference or values and life goals identification as part of shared decision-making among treatment options, or the impact of treatment on the ability to purse one's personal goals. While these were mainly smaller studies using qualitative or mixed methods, the results across them are pretty consistent in identification of life goals as an important process of patient treatment planning (e.g., Karel 2016; Gardner 2015; Davison 2006).

More generally there are several systematic reviews (e.g., Basile 2019; Imbeault and Nadeau-Fredette 2019; Hullman 2016; Nair 2003) and perspective articles that cite the importance of including life goals in decisions about treatment. In the ESRD setting this includes: selection of a dialysis modality and consideration of transplant as well as selection of vascular access type for delivery of hemodialysis treatment (Basile 2019; Chan et al., 2019; Woo and Lok, 2016; Lok and Davidson 2012). For example, Blake and Brown (2020) outline the ways such decision making should inform selection of peritoneal dialysis modality or other treatment options for people approaching kidney failure.

1a.15) Detail the process used to identify the evidence.

We carried out a literature review to identify studies and clinical practice guidelines that identified discussion or importance of life goals as part of treatment decisions. The search strategy targeted studies within the CKD/ESKD populations, but we also considered studies in other chronic disease populations along with opinion pieces and reviews.

A PubMed search was initially conducted in June 2017 subsequent to the Patient Report Outcomes Technical Expert Panel recommendation for a new Patient Life Goals measure for people with ESRD. The search was limited to articles published in the English language with the following search criteria for "life goals": dialysis AND life goals (all fields); dialysis AND patient goals (all fields); kidney disease AND life goals (all fields); life goals AND patient centered (all fields). The literature review was updated in 2023 to identify additional evidence on the relationship among patient life goals discussion as part of treatment planning and patient decision making about treatment options (e.g., ESKD modality, vascular access type). All studies, reviews, and perspective pieced were reviewed for relevance to life goals and treatment decision-making. Those determined most relevant to the PaLS are reported here (sec 1a.16).

We also consulted the 2020 KDOQI Vascular Access Guideline Update that issued statements in support of the ESKD Life-Plan.

1a.16) Provide the citation(s) for the evidence.

Basile, C., Casino, F. G., & Aucella, F. (2019). **Incremental hemodialysis, a valuable option for the frail elderly patient.** *Journal of nephrology*, *32*(5), 741–750. https://doi.org/10.1007/s40620-019-00611-4

> Management of older people on dialysis requires focus on the wider aspects of aging as well as dialysis. Recognition and assessment of frailty is vital in changing our approach in elderly patients. Current guidelines in dialysis have a limited evidence base across all age group, but particularly the elderly. We need to focus on new priorities of care when we design guidelines "for people not diseases". Patient-centered goal-directed therapy, arising from shared decision-making between physician and patient, should allow adaption of the dialysis regime. Hemodialysis (HD) in the older age group can be complicated by intradialytic hypotension, prolonged time to recovery, and access-related problems. There is increasing evidence relating to the harm associated with the delivery of standard thrice-weekly HD. Incremental HD has a lower burden of treatment. There appears to be no adverse clinical effects during the first years of dialysis in presence of a significant residual kidney function. The advantages of incremental HD might be particularly important for elderly patients with short life expectancy. There is a need for more research into specific topics such as the assessment of the course of frailty with progression of chronic kidney disease and after dialysis initiation, the choice of dialysis modality impacting on the trajectory of frailty, the timing of dialysis initiation impacting on frailty or on other outcomes. In conclusion, understanding each individual's goals of care in the context of his or her life experience is particularly important in the elderly, when overall life expectancy is relatively short, and life experience or quality of life may be the priority.

Blake, PG., and Brown, EA. **Person-centered peritoneal dialysis prescription and the role of shared decision-making**. 2020, Peritoneal Dialysis Int, published ahead of print, 1-8.

> Person-centered care has become a dominant paradigm in modern health care. It needs to be applied to people with end-stage kidney disease considering the initiation of dialysis and to peritoneal dialysis (PD) prescription and care delivery. It is relevant to their decisions about goals of care, transplantation, palliative care, and discontinuation of dialysis. It is also relevant to decisions about how PD is delivered, including options such as incremental PD. Shared decision-making is the essence of this process and needs to become a standard principle of care. It requires engagement, education, and empowerment of patients. Patient-reported outcomes and patient-reported experience are also central to person-centered care in PD.

Chan, C., Blankestijn, P., Dember, L., Gallieni, M., Harris, D., Lok, C., Mehrotra, R., Stevens, P., Wang, A., Cheung, M., Wheeler, D., Winkelmayer, W., & Pollock, C. (2019) Dialysis initiation, modality choice, access, and prescription: conclusions from a Kidney Disease: Improving Global Outcomes (KDIGO) Controversies Conference. Kidney Int 96:37-47, 2019

> Globally, the number of patients undergoing maintenance dialysis is increasing, yet throughout the world there is significant variability in the practice of initiating dialysis. Factors such as availability of resources, reasons for starting dialysis, timing of dialysis initiation, patient education and preparedness, dialysis modality and access, as well as varied “country-specific” factors significantly affect patient experiences and outcomes. As the burden of end-stage kidney disease (ESKD) has increased globally, there has also been a growing recognition of the importance of patient involvement in determining

the goals of care and decisions regarding treatment. In January 2018, KDIGO (Kidney Disease: Improving Global Outcomes) convened a Controversies Conference focused on dialysis initiation, including modality choice, access, and prescription. Here we present a summary of the conference discussions, including identified knowledge gaps, areas of controversy, and priorities for research. A major novel theme represented during the conference was the need to move away from a “one-size-fits-all” approach to dialysis and provide more individualized care that incorporates patient goals and preferences while still maintaining best practices for quality and safety. Identifying and including patient-centered goals that can be validated as quality indicators in the context of diverse health care systems to achieve equity of outcomes will require alignment of goals and incentives between patients, providers, regulators, and payers that will vary across health care jurisdictions.

Childers, JW, AL Back, JA Tulsky and RM Arnold. **REMAP: A framework for goals of care conversations**. *Journal of Oncology Practice*. 2017, Epub date: 2017/04/27, doi: 10.1200/JOP.2016.018796.

Conversations regarding goals of care with patients who have advanced cancer still occur too late, and oncologists say they lack the training to have these conversations effectively. Experts recommend a number of strategies when having these discussions, including discussing prognosis, responding to patient emotion, exploring values, and often making a recommendation for medical treatments that fit those values. To help learners, from residents to attending oncologists, learn these complex conversational skills, we have developed a framework with a mnemonic, REMAP: Reframe, Expect emotion, Map out patient goals, Align with goals, and Propose a plan. In the reframe step, the oncologist provides a big picture "headline" that lets the patient know things are in a different place. This is followed by actively attending to the patient's emotional response (expect emotion). Then, to map the patient's goals, the oncologist asks open-ended questions that are designed to help the patient think about the values that should guide his or her treatment. The oncologist then aligns with those values by explicitly reflecting them back to the patient. If the patient gives permission, the oncologist will then use those values to propose a medical plan that matches patient values. The processes underlying REMAP encourage oncologists and other clinicians to seek to understand and remain flexible, adapting their recommendations to what they hear from the patient, with ongoing revision based on the shared decision-making process. This will lead to patient-centered decisions that promote better end-of-life care.

Dahlerus C, Quinn M, Messersmith E, Lachance L, Subramanian L, Perry E, Cole J, Zhao J, Lee C, McCall M, Paulson L, Tentori F. **Patient Perspectives on the Choice of Dialysis Modality: Results From the Empowering Patients on Choices for Renal Replacement Therapy (EPOCH-RRT) Study** Am J Kidney Dis. 2016 Dec;68(6):901-910. doi: 10.1053/j.ajkd.2016.05.010. Epub 2016 Jun 21.

BACKGROUND: Little is known about factors that are important to patients with advanced kidney disease and their perspectives at the time they choose a dialysis modality. EPOCH-RRT, a study supported in part by the Patient-Centered Outcomes Research Institute (PCORI), was designed to assist patients with this choice by identifying such factors and effectively provide relevant information. STUDY DESIGN: Cross-sectional study, designed and conducted in collaboration with a multistakeholder advisory panel that included patients, caregivers, and health care professionals. SETTING & PARTICIPANTS: 180 patients with advanced chronic kidney disease (CKD;

estimated glomerular filtration rate < 25mL/min/1.73m(2)), either non-dialysis-dependent (NDD-CKD; n=65) or on dialysis therapy (hemodialysis [HD], n=77; or peritoneal dialysis, n=38), recruited across the United States through social media and in-person contacts. METHODOLOGY: Semistructured telephone interviews including open- and closed-ended questions. ANALYTICAL APPROACH: Mixed methods, integrating quantitative and qualitative approaches; themes identified through content analysis of interview transcripts by 2 independent coders. RESULTS: Themes most often reported as important were keeping as much independence as possible, quality and quantity of life, and flexibility in daily schedule. Other factors (eg, concern about the way they look) differed across patient subgroups based on age, sex, and NDD-CKD/dialysis modality. Among patients who had initiated dialysis therapy, almost half (47%) the HD patients believed that the decision to be treated by HD had largely not been their choice; this was only reported by 3% of peritoneal dialysis patients. LIMITATIONS: Recruitment through social media and willingness to participate in lengthy telephone interviews resulted in a select sample that may not be representative of the broader advanced CKD population; therefore, generalizability of findings cannot be determined. CONCLUSIONS: Incorporation of patient priorities in care improves health outcomes. Given the perceived limited role in the choice of dialysis treatment, our findings support the need for interventions to improve shared decision making on dialysis treatment options, targeting both patients and clinicians.

Davison, SN and C Simpson. **Hope and advance care planning in patients with end stage renal disease: Qualitative interview study**. *British Medical Journal*. 2006 333(7574): 886.

OBJECTIVE: To understand hope in the context of advance care planning from the perspective of patients with end stage renal disease. DESIGN: Qualitative in-depth interview study. SETTING: Outpatient department of a university affiliated nephrology programme. PARTICIPANTS: 19 patients with end stage renal disease purposively selected from the renal insufficiency, haemodialysis, and peritoneal dialysis clinics. RESULTS: Patients' hopes were highly individualised and were shaped by personal values. They reflected a preoccupation with their daily lives. Participants identified hope as central to the process of advance care planning in that hope helped them to determine future goals of care and provided insight into the perceived benefits of advance care planning and their willingness to engage in end of life discussions. More information earlier in the course of the illness focusing on the impact on daily life, along with empowerment of the patient and enhancing professional and personal relationships, were key factors in sustaining patients' ability to hope. This helped them to imagine possibilities for a future that were consistent with their values and hopes. The reliance on health professionals to initiate end of life discussions and the daily focus of clinical care were seen as potential barriers to hope. CONCLUSIONS: Facilitated advance care planning through the provision of timely appropriate information can positively enhance rather than diminish patients' hope. Current practices concerning disclosure of prognosis are ethically and psychologically inadequate in that they do not meet the needs of patients.

Department of Health and Human Services, Centers for Medicare and Medicaid Services: The Centers for Medicare and Medicaid Programs. (2008**). Conditions for coverage for end-stage renal disease facilities, Final Rule**. 42 CFR Parts 405, 410, 413 Fed Regist 73(73), April 15, 2008, Accessed August 30, 2021 https://www.cms.gov/Regulations-and-Guidance/Legislation/CFCsAndCoPs/downloads/esrdfinalrule0415.pdf.

de Vries, AM, T Schulz, R Westerhuis, GJ Navis, J Niesing, AV Ranchor and MJ Schroevers. **Goal disturbance changes pre/post-renal transplantation are related to changes in distress**. *British Journal of Health Psychology*. 2017, Epub date: 2017/05/22, doi: 10.1111/bjhp.12243.

OBJECTIVE: Renal transplantation (RTx) is considered the treatment of choice for end-stage renal disease (ESRD) given its association with lower mortality, and improved overall quality of life and psychological functioning compared to dialysis. However, much less is known about which factors underlie these psychological improvements across RTx. Goal theory suggests that experienced disturbances in important goals are related to lower psychological functioning. This study aimed to (1) identify the most disturbed and most important goals for patients before RTx, (2) to examine changes in goal disturbance and goal importance pre/post-RTx, and (3) to examine whether changes in goal disturbance are associated with changes in psychological distress over time, and whether this relationship is mediated by changes in perceived control. METHODS: In this longitudinal study, 220 patients completed questionnaires before and after RTx, including questionnaires to assess goals (GOALS questionnaire), psychological distress (GHQ-12), and perceived control (Mastery scale). RESULTS: End-stage renal disease affected both general and disease-specific goals. Approximately 30% of the patients indicated to experience high or very high disturbance before transplantation. Goal disturbance generally decreased significantly pre- to post-RTx, whereas goal importance did not change significantly pre- to post-RTx. No mediation effect of perceived control was found. Instead, both changes in goal disturbance and perceived control showed independent effects on changes in distress. CONCLUSIONS: Intervention strategies targeting attainable and realistic goal setting, and perceived control in RTx recipients who do not benefit optimally from RTx, might enhance psychological functioning in this population. Statement of contribution What is already known on this subject? Kidney transplantation improves patients' psychological functioning. Experienced disturbances in important life goals are related to lower psychological functioning in chronic illness. What does this study add? Goal disturbance decreases after renal transplantation, and this is related to a decrease in distress over time. Perceived control does not mediate the relationship between goal disturbance and distress pre/post-transplantation. Changes in perceived control have an additional main effect on changes in distress.

Della Santina, C and RH Bernstein. **Whole-patient assessment, goal planning, and inflection points: Their role in achieving quality end-of-life care**. *Clinics in Geriatric Medicine*. 2004 20(4): 595-620, v.

This article provides a framework for performing whole-patient assessment and goal planning. These clinical tasks involve a multidisciplinary, multidimensional, patient-centered approach to care and a deep appreciation for the complex interplay between the physical, psychological, social, and spiritual aspects of the human experience of dying. This article stresses the iterative nature of whole-patient assessment and goal planning, both of which should be conducted at certain important junctures in a patient's progression to manage effectively the evolving challenges faced by terminally ill persons and their families. This article also provides suggestions on successfully managing the communication challenges in caring for patients near the end of life and their family.

Dorough, A., Forfang, D., Mold, J., Kshirsagar, A., DeWalt, D., Flythe, J. (2020). A person-centered interdisciplinary plan-of-care program for dialysis: Implementation and preliminary testing. *Kidney Med 3*(2), 193-205.

Background: Dialysis care often focuses on outcomes that are of lesser importance to patients than to clinicians. There is growing international interest in individualizing care based on patient priorities, but evidence-based approaches are lacking. The objective of this study was to develop a person-centered dialysis care planning program. To achieve this objective we performed qualitative interviews, responsively developed a novel care planning program and then assessed program content and burden. Methods: We conducted 25 concept elicitation interviews with US hemodialysis patients, care partners and care providers, using thematic analysis to analyze transcripts. Interview findings and interdisciplinary stakeholder panel input informed the development of a new care planning program, My Dialysis Plan. We then conducted 19 cognitive debriefing interviews with patients, care partners and care providers to assess the program's content and face validities, comprehensibility and burden. Results: We identified five themes in concept elicitation interviews: feeling boxed in by the system, navigating dual lives, acknowledging an evolving identity, respecting the individual as a whole person and increasing individualization to enhance care. We then developed a person-centered care planning program and supporting materials that underwent 32 stakeholder-informed iterations. Data from subsequent cognitive interviews led to program revisions intended to improve contextualization and understanding, decrease burden and facilitate implementation. Conclusions: My Dialysis Plan is a content-valid, person-centered dialysis care planning program that aims to promote care individualization. Investigation of the program's capacity to improve patient experiences and outcomes is needed.

Gardner, T, K Refshauge, J McAuley, S Goodall, M Hubscher and L Smith. **Patient led goal setting in chronic low back pain-what goals are important to the patient and are they aligned to what we measure?** *Patient Education and Counseling*. 2015 98(8): 1035-1038.

OBJECTIVE: To determine the extent of alignment between clinical outcome measures and patient-derived goals for the management of chronic low back pain (cLBP). METHODS: A customised, patient-led goal setting intervention was implemented facilitated by a physiotherapist, in which participants identified problem areas and developed strategies to address them. Patient goals were compared to the most commonly used outcome measures in cLBP as well as research outcomes recommended by the IMMPACT consortium. RESULTS: From 20 participants, a total of 27 unique goals were identified, the most common goal related to physical activity (49%). Comparison of participant goals to the most common measures used by physiotherapists found none of the goals could be aligned. Comparison of goals and domains with IMPACCT outcome domains found 76% of the goals were aligned with physical functioning and 16% with emotional functioning. CONCLUSION: This study has identified goals important to patients in cLBP, these were varied, and most did not correspond with current clinical measures. PRACTICE IMPLICATIONS: Clinical outcome measures may not be providing accurate information about the success of treatments that are meaningful to the patient. Clinicians should consider a collaborative approach with cLBP patients to determine treatment interventions that are driven by patient preference.

Goldwater, D., & Wenger, N. K. (2023). **Patient-centered care in geriatric cardiology.** Trends in cardiovascular medicine, 33(1), 13–20. https://doi.org/10.1016/j.tcm.2021.11.001

Geriatric cardiology involves providing cardiovascular care to older adults in relation to aging. Although cardiovascular diseases are the most common diseases faced by older adults, they often co-occur with numerous aging-related challenges, such as

multimorbidity, frailty, polypharmacy, falls, functional and cognitive impairment, which present challenges to implementing standard disease-based treatment strategies. Faced with these complexities, patient-centered care in geriatric cardiology strives to direct all management toward the achievement of an individual's prioritized health and life goals by employing shared decision-making to align treatment with goals, utilizing stated goals to navigate situations of treatment uncertainty, and pro-actively mitigating aging-related risks. This fundamental change in cardiovascular medicine from disease-centered management to patient-centered goal-directed care is necessary to facilitate wellness, independence, and favorable quality of life outcomes in the older adult population.

Huang, ES, R Gorawara-Bhat and MH Chin. **Self-reported goals of older patients with type 2 diabetes mellitus**. *Journal of the American Geriatrics Society*. 2005 53(2): 306-311.

OBJECTIVES: New diabetes mellitus guidelines from the American Geriatrics Society promote the individualization of treatment goals and plans for patients aged 65 and older. Communicating with older patients about such complex medical decisions presents new challenges for providers. The self-reported healthcare goals, factors influencing these goals, and self-care practices of older patients with diabetes mellitus were explored. DESIGN: Exploratory study involving semistructured interviews. SETTING: Four clinics of a midwestern, urban academic medical center. PARTICIPANTS: Patients aged 65 and older with type II diabetes mellitus (N=28). MEASUREMENTS: Semistructured, one-on-one interviews were conducted. Interviews were audiotaped, transcribed, and evaluated for recurring themes using a grounded theory approach. RESULTS: The majority of patients expressed their healthcare goals in a social and functional language, in contrast to the biomedical language of risk factor control and complication prevention, even when specifically asked about goals for diabetes mellitus care. Patient's predominant healthcare goals centered on maintaining their independence and their activities of daily living (71%). Medical experiences of friends and family (50%), social comparison with peers (7%), and medical professionals (43%) shaped patients' goals. Self-reported medication adherence and glucose monitoring was high, but more than one-quarter of patients failed to adhere to any dietary recommendations, and one-third failed to adhere to their exercise regimens. CONCLUSION: As diabetes mellitus care recommendations for older patients grow more complex, providers could enhance their communication about such medical decisions by exploring patients' specific circumstances and reframing diabetes mellitus treatment goals in patients' own language. These may be crucial steps to developing successful individualized care plans.

Hullfish, KL, VE Bovbjerg and WD Steers. **Patient-centered goals for pelvic floor dysfunction surgery: Long-term follow-up**. *American Journal of Obstetrics and Gynecology*. 2004 191(1): 201-205.

OBJECTIVE: This study was undertaken to describe long-term postoperative perceived achievement of subjective preoperative goals for pelvic floor dysfunction (PFD) surgery. STUDY DESIGN: From March 2000 through December 2001, 123 PFD surgical patients completed a preoperative open-ended questionnaire on which they described up to 5 personal outcome goals for PFD surgery. Patients were asked to review their original goals list and assess the degree to which they had met their goals on a 5-point scale (-2=strongly disagree the goal had been met to +2=strongly agree that the goal had been met) 3 months after surgery and again between 1 and 3 years after surgery. At the second follow-up, patients also completed the Incontinence Impact Questionnaire (IIQ-7)

and Urogenital Distress Inventory (UDI-6) instruments to assess life impact and symptom distress, respectively. RESULTS: Of 50 women to date with long-term follow-up, 98% were white, 96% had delivered at least 1 child, 38% had previous surgery for PFD, mean weight was 74.2 +/- 14.1 kilos, and mean age was 65.4 +/- 11.5 years. Mean follow-up duration was 1.8 years, and ranged from 0.98 to 3.01 years. Of 194 goals listed by participants, 40.2% had to do with resuming previous activities or lifestyle, 38.1% with symptom relief, 9.3% with improving self-image and social relationships, 7.7% with improving general health, and 4.6% with improving physical appearance. At the individual goal level, 72% of goals were attained at short-term, and 68% attained at long-term follow-up. Long-term goal achievement did not vary significantly by category of goal. Goal achievement was lower only for symptom relief at long-term follow-up (68.9%) than at short-term follow-up (87.4%, P <.001). At the person level, 45.8% of women reported achieving all listed goals in the short term, and 42.0% in the long term. Long-term goal achievement was associated with PFD-specific quality of life (UDI-6 and IIQ-7 scores) and inversely associated with surgical complications, but was not associated with other clinical or demographic variables, including weight, parity, PFD diagnosis, psychiatric comorbidity, surgical route, or previous surgical history. CONCLUSION: Self-reported achievement of preoperatively recorded goals for PFD surgery persisted 1 to 3 years after surgery. The association of goal achievement to IIQ-7 and UDI-6 scores suggests that goal achievement is related to, but not identical to, overall measures of PFD life impact and symptom distress. Future work should examine the association of goal achievement to clinical measures of PFD severity, and compare surgically and medically managed patients. Preoperative assessment of goals may be a useful addition to clinical and subjective data in the long-term management of women with pelvic floor disorders.

Hullmann, SE, SL Robb and KL Rand. **Life goals in patients with cancer: A systematic review of the literature**. *Psychooncology*. 2016 25(4): 387-399.

OBJECTIVE: Purposes of this systematic review of life goal research in cancer patients were to (1) identify life goal characteristics and processes being examined, (2) describe instruments used to assess life goal constructs, (3) identify theoretical models being used to guide research, and (4) summarize what is known about the impact of the cancer experience on life goal characteristics, processes, and psychological outcomes. METHODS: We conducted this systematic review using MEDLINE, PubMed, CINAHL, and PsycINFO databases. Inclusion criteria were as follows: (1) published between 1993 and 2014, (2) English language, (3) cancer patient population, and (4) original research articles that assessed life goal characteristics and/or goal processes. One hundred ninety-seven articles were screened and 27 included in the final review. RESULTS: Seven life goal characteristics and seven life goal processes were identified, and less than half of studies investigated associations between goal characteristics and processes. Conceptual definitions were not provided for about half of the identified life goal constructs. Studies used both validated and author-developed instruments to assess goal constructs. Twenty-four different theoretical models were identified, with self-regulation theory most frequently cited. Overall, the literature suggests that cancer impacts patients' life goal characteristics and processes, and life goal disturbance is related to poorer psychological outcomes. CONCLUSIONS: The impact of the cancer experience on life goals is an important and emerging area of research that would benefit from conceptual and theoretical clarity and measurement consistency.

Imbeault B., Nadeau-Fredette AC. **Optimization of Dialysis Modality Transitions for Improved Patient Care**. Can J Kidney Health Dis. 2019 Oct 16;6:2054358119882664. doi: 10.1177/2054358119882664. eCollection 2019.

Purpose of review: Initial and subsequent modality decisions are important, impacting both clinical outcomes and quality of life. Transition from chronic kidney disease to dialysis and between dialysis modalities are periods were patients may be especially vulnerable. Reviewing our current knowledge surrounding these critical periods and identifying areas for future research may allow us to develop dialysis strategies beneficial to patients. Sources of information: We searched the electronic database PubMed and queried Google Scholar for English peer-reviewed articles using appropriate keywords (non-exhaustive list): dialysis transitions, peritoneal dialysis, home hemodialysis, integrated care pathway, and health-related quality of life. Primary sources were accessed whenever possible. Methods: In this narrative review, we aim to expose the controversies surrounding home-dialysis first strategies and examine the evidence underpinning home-dialysis first strategies as well as home-to-home and home-to-in-center transitions. Key findings: Diverse factors must be taken into consideration when choosing initial and subsequent dialysis modalities. Given the limitations of available data (and lack of convincing benefit or detriment of one modality over the other), patient-centered considerations may prime over suspected mortality benefits of one modality or another. Limitations: Available data stem almost exclusively from retrospective and observational studies, often using large national and international databases, susceptible to bias. Furthermore, this is a narrative review which takes into account the views and opinions of the authors, especially as it pertains to optimal dialysis pathways. Implications: Emphasis must be placed on individual patient goals and preferences during modality selection while planning ahead to achieve timely and appropriate transitions limiting discomfort and anxiety for patients. Further research is required to ascertain specific interventions which may be beneficial to patients.

Karel, MJ, EA Mulligan, A Walder, LA Martin, J Moye and AD Naik. **Valued life abilities among veteran cancer survivors**. *Health Expectations*. 2016 19(3): 679-690.

BACKGROUND: When patients have multiple chronic illnesses, it is not feasible to provide disease-based care when treatments for one condition adversely affect another. Instead, health-care delivery requires a broader person-centred treatment plan based on collaborative, patient-oriented values and goals. OBJECTIVE: We examined the individual variability, thematic content, and sociodemographic correlates of valued life abilities and activities among multimorbid veterans diagnosed with life-altering cancer. SETTING AND PARTICIPANTS: Participants were 144 veterans in the 'Vet-Cares' study who completed a health-care values and goals scale 12 months after diagnosis of head and neck, gastro-oesophageal, or colorectal cancer. They had mean age of 65 years and one quarter identified as Hispanic and/or African American. DESIGN: At twelve months post-diagnosis, participants rated 16 life abilities/activities in their importance to quality of life on a 10-point Likert scale, during an in-person interview. Scale themes were validated via exploratory factor analysis and examining associations with sociodemographic variables. RESULTS: Participants rated most life abilities/activities as extremely important. Variability in responses was sufficient to identify three underlying values themes in exploratory factor analysis: self-sufficiency, enjoyment/comfort, and connection to family, friends and spirituality. Veterans with a spouse/partner rated self-sufficiency as less important. African American veterans rated connection as more important than did White veterans. CONCLUSIONS: It is feasible yet challenging to ask

older, multimorbid patients to rate relative importance of values associated with life abilities/activities. Themes related to self-sufficiency, enjoyment/comfort in daily life and connection are salient and logically consistent with sociodemographic traits. Future studies should explore their role in goal-directed health care.

Ladin, K, N Lin, E Hahn, G Zhang, S Koch-Weser and DE Weiner. **Engagement in decision-making and patient satisfaction: A qualitative study of older patients' perceptions of dialysis initiation and modality decisions.** *Nephrology Dialysis Transplantation*. 2016, Epub date: 2016/09/01, doi: 10.1093/ndt/gfw307.

BACKGROUND: Although shared decision-making (SDM) can better align patient preferences with treatment, barriers remain incompletely understood and the impact on patient satisfaction is unknown. METHODS: This is a qualitative study with semistructured interviews. A purposive sample of prevalent dialysis patients >/=65 years of age at two facilities in Greater Boston were selected for diversity in time from initiation, race, modality and vintage. A codebook was developed and interrater reliability was 89%. Codes were discussed and organized into themes. RESULTS: A total of 31 interviews with 23 in-center hemodialysis patients, 1 home hemodialysis patient and 7 peritoneal dialysis patients were completed. The mean age was 76 +/- 9 years. Two dominant themes (with related subthemes) emerged: decision-making experiences and satisfaction, and barriers to SDM. Subthemes included negative versus positive decision-making experiences, struggling for autonomy, being a 'good patient' and lack of choice. In spite of believing that dialysis initiation should be the patient's choice, no patients perceived that they had made a choice. Patients explained that this is due to the perception of imminent death or that the decision to start dialysis belonged to physicians. Clinicians and family frequently overrode patient preferences, with patient autonomy honored mostly to select dialysis modality. Poor decision-making experiences were associated with low treatment satisfaction. CONCLUSIONS: Despite recommendations for SDM, many older patients were unaware that dialysis initiation was voluntary, held mistaken beliefs about their prognosis and were not engaged in decision-making, resulting in poor satisfaction. Patients desired greater information, specifically focusing on the acuity of their choice, prognosis and goals of care.

Lok CE, Huber TS, Lee T, et al; **KDOQI Vascular Access Guideline Work Group. KDOQI clinical practice guideline for vascular access: 2019 update**. *Am J Kidney Dis*. 2020;75(4)(suppl 2):S1-S164.

The National Kidney Foundation’s Kidney Disease Outcomes Quality Initiative (KDOQI) has provided evidence based guidelines for hemodialysis vascular access since 1996. Since the last update in 2006, there has been a great accumulation of new evidence and sophistication in the guidelines process. The 2019 update to the KDOQI Clinical Practice Guideline for Vascular Access is a comprehensive document intended to assist multidisciplinary practitioners care for chronic kidney disease patients and their vascular access. New topics include the end-stage kidney disease “Life-Plan” and related concepts, guidance on vascular access choice, new targets for arteriovenous access (fistulas and grafts) and central venous catheters, management of specific complications, and renewed approaches to some older topics. Appraisal of the quality of the evidence was independently conducted by using a Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach, and interpretation and application followed the GRADE Evidence to Decision frameworks. As applicable, each guideline statement is accompanied by rationale/background information, a detailed

justification, monitoring and evaluation guidance, implementation considerations, special discussions, and recommendations for future research.

Lok, CE and I Davidson. **Optimal choice of dialysis access for chronic kidney disease patients: Developing a life plan for dialysis access.** *Seminars in Nephrology*. 2012 32(6): 530-537.

Patient-focused dialysis modality and access selection requires a coordinated teamwork approach that emphasizes chronic kidney disease care to be a continuum of care. Individualized and detailed patient history and examination are the mainstays of dialysis modality and access selection. Preoperative vessel mapping by duplex Doppler ultrasonography can be a useful supplementary investigation to the history and physical examination to determine the optimal dialysis access type and site. Dialysis access modality and choice considers many patient factors that can be aided by a clinical risk score, asking key clinical questions, surgical expert opinion, and a multidisciplinary approach to individualized patient care. In many situations, a lifelong access utilization strategy prioritizes peritoneal dialysis as the first dialysis modality followed by appropriately planned hemodialysis. The goal of an integrated patient-focused approach is to achieve complication-free access to help patients achieve their life goals on and off dialysis.

Lu, E., and Chai, E. (2021). **Kidney Supportive Care in Peritoneal Dialysis: Developing a Person-Centered Kidney Disease Care Plan.** *Kidney medicine*, *4*(2), 100392. https://doi.org/10.1016/j.xkme.2021.10.005

Individuals receiving peritoneal dialysis (PD)-similar to those receiving hemodialysis - may experience high mortality coupled with a high symptom burden and reduced health-related quality of life. In this context, a discussion of the risks, benefits, and tradeoffs of PD and/or other kidney treatment modalities should be explored based on individual goals and preferences. Through these principles, kidney supportive care provides a person-centered approach to kidney disease care throughout the spectrum of kidney failure and earlier stages of chronic kidney disease. Kidney supportive care is offered in conjunction with life-prolonging therapies, including dialysis and kidney transplants, and is increasingly recognized as an integral part of advancing the care of PD patients. Using "My Kidney Care Roadmap" for shared decision making, kidney supportive care guides patients undergoing PD and their clinicians to (1) elicit patient goals, values, and priorities; (2) convey medical prognosis and suitable treatment options; and (3) ask "Which of these kidney treatment options will best help me achieve my goals and priorities?" to inform both current and future decisions, including choice of dialysis modalities, time-limited trials, and/or nondialysis management. Recognizing that patient priorities and choices may evolve, this framework ultimately allows patients to continually reassess their PD care to better achieve goal-directed dialysis.

Mold J. (2017). **Goal-Directed Health Care: Redefining Health and Health Care in the Era of Value-Based Care.** Cureus, 9(2), e1043. https://doi.org/10.7759/cureus.1043

Health care reform efforts have increasingly emphasized payment models that reward value (quality/cost). It seems appropriate, therefore, to examine what we value in health care, and that will require that we examine our definition of health. In spite of admonitions from the World Health Organization and others, our current health care system operates under the assumption that health represents the absence of health

problems. While that perspective has led to incredible advances in medical science, it now may be adversely affecting value. Problem-oriented care is clearly one of the drivers of rising costs and it could be adversely affecting the quality of care, depending upon how quality is defined. If we redefined health in terms of patient-centered goals, health care could be focused more directly on meaningful outcomes, reducing the number of irrelevant tests and treatments. Greater emphasis would be placed on prevention, meaningful activities, advance directives and personal growth and development. The role of patients within clinician-patient relationships would be elevated, strengthening therapeutic relationships. Reframing health in terms of health-related goals and directing the health care system to help people achieve them, could both improve quality and reduce costs. In the process, it could also make health care less mechanical and more humane.

Nair KPS. (2003) **Life goals: the concept and its relevance to rehabilitation.** Clinical Rehabilitation. 17(2):192-202. doi:10.1191/0269215503cr599oa

Objective: Life goals are desired states that people seek to obtain maintain or avoid. These goals may influence motivation to participate in the rehabilitation process. The aim of this paper is to review the literature on life goals and the influence of life goals on the rehabilitation process. Methods: The MEDLINE, EMBASE, Psychlit and CINAHL databases were searched with the keywords goals, life goals, aim of life, meaning of life, motivation, assessment (identification) of life goals, goal planning, disability, coping and rehabilitation. Results: The initial search produced 917 abstracts. After going through these abstracts, 39 articles were selected for inclusion in the review. Age, gender, personality, experiences and society and environment influence life goals. Pursuit and attainment of life goals affect sense of well-being. Life goals are accessible to conscious awareness and can be identified. Several questionnaires are available for assessment of life goals. Different questionnaires assess different aspects of life goals. All except one of these questionnaires need to be tested for validity and reliability in a rehabilitation setting. Disabilities interfere with goal striving and result in emotional distress. Motivation to participate in a rehabilitation programme depends on concurrence between a patient's life goals and treatment goals. Incorporation of a subject's life goals into a management programme resulted in better outcomes in various physical and psychiatric disorders. There are no data on the efficacy of life goal-orientated rehabilitation programmes. Conclusions: Life goals influence patients' motivation to participate in and compliance with treatment programmes. We still do not know whether rehabilitation programmes focusing on life goals make any difference in outcome. There is need for further studies in this area.

O'Hare, AM, N Armistead, WL Schrag, L Diamond and AH Moss. **Patient-centered care: An opportunity to accomplish the "three aims" of the national quality strategy in the Medicare ESRD program**. *Clinical Journal of the American Society of Nephrology*. 2014 9(12): 2189-2194.

In light of mounting federal government debt and levels of Medicare spending that are widely viewed as unsustainable, commentators have called for a transformation of the United States health care system to deliver better care at lower costs. This article presents the priorities of the Coalition for Supportive Care of Kidney Patients for how clinicians might achieve this transformation for patients with advanced CKD and their families. The authors suspect that much of the high-intensity, high-cost care currently delivered to patients with advanced kidney disease may be unwanted and that the

"Three Aims" put forth by the National Quality Strategy of better care for the individual, better health for populations, and reduced health care costs may be within reach for patients with CKD and ESRD. This work describes the coalition's vision for a more patient-centered approach to the care of patients with kidney disease and argues for more concerted efforts to align their treatments with their goals, values, and preferences. Key priorities to achieve this vision include using improved prognostic models and decision science to help patients, their families, and their providers better understand what to expect in the future; engaging patients and their families in shared decision-making before the initiation of dialysis and during the course of dialysis treatment; and tailoring treatment strategies throughout the continuum of their care to address what matters most to individual patients.

Periyakoil, V.S., Neri E., Kraemer, H. (2018). **Common Items on a Bucket List.** *Journal of Palliative Medicine*. 21(5), 652-658. http://doi.org.proxy.lib.umich.edu/10.1089/jpm.2017.0512

Background: To provide preference-sensitive care, we propose that clinicians might routinely inquire about their patients' bucket-lists and discuss the impact (if any) of their medical treatments on their life goals. Methods: This cross-sectional, mixed methods online study explores the concept of the bucket list and seeks to identify common bucket list themes. Data were collected in 2015–2016 through an online survey, which was completed by a total of 3056 participants across the United States. Forty participants who had a bucket list were identified randomly and used as the development cohort: their responses were analyzed qualitatively using grounded theory methods to identify the six key bucket list themes. The responses of the remaining 3016 participants were used for the validation study. The codes identified from the development cohort were validated by analyses of responses from 50 randomly drawn subjects from the validation cohort. All the 3016 validation cohort transcripts were coded for presence or absence of each of the six bucket list themes. Results: Around 91.2% participants had a bucket list. Age and spirituality influence the patient's bucket-list. Participants who reported that faith/religion/spirituality was important to them were most likely (95%) to have a bucket list compared with those who reported it to be unimportant (68.2%), $\chi 2 = 37.67$. Six primary themes identified were the desire to travel (78.5%), desire to accomplish a personal goal (78.3%), desire to achieve specific life milestones (51%), desire to spend quality time with friends and family (16.7%), desire to achieve financial stability (24.3%), and desire to do a daring activity (15%). Conclusions: The bucket list is a simple framework that can be used to engage patients about their healthcare decision making. Knowing a patient's bucket list can aid clinicians in relating each treatment option to its potential impact (if any) on the patient's life and life goals to promote informed decision making.

Purkaple, B. A., Nagykaldi, Z. J., Allahyar, A., Todd, R., & Mold, J. W. (2020). **Physicians' Response to Patients' Quality-of-Life Goals.** *Journal of the American Board of Family Medicine : JABFM*, *33*(1), 71–79. https://doi.org/10.3122/jabfm.2020.01.190169

Purpose: Patients are able to participate in quality-of-life (QOL) discussions, but clinicians struggle to incorporate this information into encounters and shared decision making. We designed a study to determine if a clinician-initiated prompt could make patient visits more goal directed. Methods: Patients were given a previsit questionnaire that included QOL questions. Physicians in the control were given no further prompting. The intervention physicians were prompted to ask a QOL question: what things are you unable to do because of your health problems today? A 2-pronged design was used: 1

prepost group where 3 physicians participated in 5 control and 5 intervention encounters (n = 30) and a randomized group in which 11 physicians and their patients were randomly assigned to control or intervention groups (n = 30). Video recordings of the encounters were reviewed to determine if QOL goals were mentioned and if they were utilized in decision making. Results: Fifty-seven (95%) of the 60 patients provided written answers to at least 1 of the QOL questions on the intake form. QOL goals were mentioned during intervention encounters more often than in control groups. QOL information was used in shared decision making in only 4 of the 30 (13%) intervention encounters. Conclusions: Physicians were able to engage in QOL discussions with their patients, but did not translate that information to medical decision making. More research is needed to understand why clinicians opt not to use QOL information and how to make communication more goal directed.

Robinski, M., Mau, W., Wienke, A., Girndt, M. (2017). **The choice of renal replacement therapy (CORETH) project: Dialysis patients' psychosocial characteristics and treatment satisfaction.** *Nephrol Dial Transplant, 32*(2), 315-324.

Background: Until today, research has underestimated the role of psychosocial conditions as contributing factors to dialysis modality choice. The novelty within the Choice of Renal Replacement Therapy (CORETH) project (German Clinical Trials Register #DRKS00006350) is its focus on the multivariate associations between these aspects and their consecutive significance regarding treatment satisfaction (TS) in peritoneal dialysis (PD) versus haemodialysis (HD) patients. In this article, we present the baseline results of a multicentre study, which is supported by a grant from the German Ministry for Education and Research. Methods: Six to 24 months after initiation of dialysis, 780 patients from 55 dialysis centres all over Germany were surveyed. The questionnaire addressed psychosocial, physical, socio-demographic and shared decision-making (SDM) aspects. Furthermore, cognitive functioning was tested. After indexing the measures, two propensity score-matched groups (n = 482) were compared in a first step, after having chosen PD or HD. In a second step, a moderated multiple regression (n = 445) was conducted to initially investigate the multivariate impact of patient characteristics on TS. Results: In comparison with HD patients, PD patients were more satisfied with their treatment (P < 0.001), had a more autonomy-seeking personality (P = 0.04), had better cognitive functioning (P = 0.001), indicated more satisfying SDM (P < 0.001) and had a larger living space (P < 0.001). All patients were more satisfied when they had a good psychological state and received SDM. Especially in HD patients, TS was higher when the patient had a less autonomous personality, lower cognitive functioning, more social support, a poorer physical state and poorer socio-demographic conditions (R2 = 0.26). Conclusions: Psychosocial characteristics play a major role in TS in dialysis patients. Within a multivariate approach, these factors are even more important than physical or environment-related factors. In practice, focusing on SDM and screening patient characteristics at an early stage can foster patients' TS. Changes will be examined in a 1-year follow-up.

Robinski M, Mau W, Wienke A, Girndt M. **Shared decision-making in chronic kidney disease: A retrospection of recently initiated dialysis patients in Germany** Patient Educ Couns. 2016 Apr;99(4):562-570. doi: 10.1016/j.pec.2015.10.014. Epub 2015 Oct 29.

OBJECTIVE: To compare differences in shared decision-making (SDM) and treatment satisfaction (TS) between haemodialysis (HD) and peritoneal dialysis (PD) patients. METHODS: 6-24 months after initiation of dialysis, we surveyed 780 patients from

throughout Germany (CORETH-project) regarding SDM, the reason for modality choice and TS. Data were compared between two age-, comorbidity-, education-, and employment status-matched groups (n=482). RESULTS: PD patients rated all aspects of SDM more positively than did HD patients (total score: MPD=84.6, SD=24.1 vs. MHD=61.9, SD=37.3; $p \leq 0.0001$). The highest difference occurred for the item "announcement of a necessary decision" (delta=1.3 points on a 6-point Likert-scale). PD patients indicated their desire for independence as a motivator for choosing PD (65%), whereas HD patients were subject to medical decisions (23%) or wanted to rely on medical support (20%). We found positive correlations between SDM and TS ($0.16 \leq r \leq 0.48$; $p \leq 0.0001$). CONCLUSION: Our findings increase awareness of a participatory nephrological counseling-culture and imply that SDM can pave the way for quality of life and treatment success for dialysis patients. PRACTICE IMPLICATIONS: Practitioners can facilitate SDM by screening patient preferences at an early stage, being aware of biases in consultation, using easy terminology and encouraging passive patients to participate in the choice.

Rodriguez-Gutierrez, R., Gonzalez-Gonzalez, J. G., Zuñiga-Hernandez, J. A., & McCoy, R. G. (2019). **Benefits and harms of intensive glycemic control in patients with type 2 diabetes.** *BMJ (Clinical research ed.)*, *367*, l5887. https://doi.org/10.1136/bmj.l5887

Diabetes is a major and costly health concern worldwide, with high morbidity, disability, mortality, and impaired quality of life. The vast majority of people living with diabetes have type 2 diabetes. Historically, the main strategy to reduce complications of type 2 diabetes has been intensive glycemic control. However, the body of evidence shows no meaningful benefit of intensive (compared with moderate) glycemic control for microvascular and macrovascular outcomes important to patients, with the exception of reduced rates of non-fatal myocardial infarction. Intensive glycemic control does, however, increase the risk of severe hypoglycemia and incurs additional burden by way of polypharmacy, side effects, and cost. Additionally, data from cardiovascular outcomes trials showed that cardiovascular, kidney, and mortality outcomes may be improved with use of specific classes of glucose lowering drugs largely independently of their glycemic effects. Therefore, delivering evidence based, patient centered care to people with type 2 diabetes requires a paradigm shift and departure from the predominantly glucocentric view of diabetes management. Instead of prioritizing intensive glycemic control, the focus needs to be on ensuring access to adequate diabetes care, aligning glycemic targets to patients' goals and situations, minimizing short term and long term complications, reducing the burden of treatment, and improving quality of life.

Schellinger, SE, EW Anderson, MS Frazer and CL Cain. **Patient self-defined goals**. *American Journal of Hospice and Palliative Medicine*. 2017, Epub date: 2017/01/01, doi: 1049909117699600.

This research, a descriptive qualitative analysis of self-defined serious illness goals, expands the knowledge of what goals are important beyond the physical?making existing disease-specific guidelines more holistic. Integration of goals of care discussions and documentation is standard for quality palliative care but not consistently executed into general and specialty practice. Over 14 months, lay health-care workers (care guides) provided monthly supportive visits for 160 patients with advanced heart failure, cancer, and dementia expected to die in 2 to 3 years. Care guides explored what was most important to patients and documented their self-defined goals on a medical record flow sheet. Using definitions of an expanded set of whole-person domains

adapted from the National Consensus Project (NCP) Clinical Practice Guidelines for Quality Palliative Care, 999 goals and their associated plans were deductively coded and examined. Four themes were identified?medical, nonmedical, multiple, and global. Forty percent of goals were coded into the medical domain; 40% were coded to nonmedical domains?social (9%), ethical (7%), family (6%), financial/legal (5%), psychological (5%), housing (3%), legacy/bereavement (3%), spiritual (1%), and end-of-life care (1%). Sixteen percent of the goals were complex and reflected a mix of medical and nonmedical domains, ?multiple? goals. The remaining goals (4%) were too global to attribute to an NCP domain. Self-defined serious illness goals express experiences beyond physical health and extend into all aspects of whole person. It is feasible to elicit and record serious illness goals. This approach to goals can support meaningful person-centered care, decision-making, and planning that accords with individual preferences of late life.

Schulman-Green, DJ, AD Naik, EH Bradley, R McCorkle and ST Bogardus. **Goal setting as a shared decision making strategy among clinicians and their older patients**. *Patient Education and Counseling*. 2006 63(1-2): 145-151.

OBJECTIVE: Older adults are less likely than other age groups to participate in clinical decision-making. To enhance participation, we sought to understand how older adults consider and discuss their life and health goals during the clinical encounter. METHODS: We conducted six focus groups: four with community-dwelling older persons (n=42), one with geriatricians and internists (n=6), and one with rehabilitation nurses (n=5). Participants were asked to discuss: patients' life and health goals; communication about goals, and perception of agreement about health goals. Group interactions were tape-recorded, transcribed, and analyzed using content analysis. RESULTS: All participants were willing to discuss goals, but varied in the degree to which they did so. Reasons for non-discussion included that goal setting was not a priority given limited time, visits focused on symptoms, mutual perception of disinterest, and the presumption that all patients' goals were the same. CONCLUSION: Interventions to enhance goal setting need to address key barriers to promoting goals discussions. Participants recognized the benefits of goal setting, however, training and instruments are needed to integrate goal setting into medicine. PRACTICE IMPLICATIONS: Setting goals initially and reviewing them periodically may be a comprehensive, time-efficient way of integrating patients' goals into their care plans.

Serrano, V., Spencer-Bonilla, G., Boehmer, K. R., & Montori, V. M. (2017**). Minimally Disruptive Medicine for Patients with Diabetes**. *Current diabetes reports*, *17*(11), 104. https://doi.org/10.1007/s11892-017-0935-7

Purpose of review: Patients with diabetes must deal with the burden of symptoms and complications (burden of illness). Simultaneously, diabetes care demands practical and emotional work from patients and their families, work to access and use healthcare and to enact self-care (burden of treatment). Patient work must compete with the demands of family, job, and community life. Overwhelmed patients may not have the capacity to access care or enact self-care and will thus experience suboptimal diabetes outcomes. Recent findings: Minimally disruptive medicine (MDM) is a patient-centered approach to healthcare that prioritizes patients' goals for life and health while minimizing the healthcare disruption on patients' lives. In patients with diabetes, particularly in those with complex lives and multimorbidity, MDM coordinates healthcare and community responses to improve outcomes, reduce treatment burden, and enable patients to

pursue their life's hopes and dreams.

Thornton, TA and RM Hakim. **Meaningful rehabilitation of the end-stage renal disease patient**. *Seminars in Nephrology*. 1997 17(3): 246-252.

In this highly technological age, health care providers are called to attend to the patient as a whole person, with dreams and goals and a desire for purpose and meaning in life. In this article, we propose a broadened definition of rehabilitation and a rehabilitation program designed to effect an improvement in the quality of life of each renal patient by aiming to restore meaningful existence in each of their lives. An individualized plan for rehabilitation can be constructed and implemented with far-reaching success when the focus is on the life goals of the patient, whether physical, social, psychological, or intellectual. These programs not only enhance the quality of life of the patient with end-stage renal disease, but are cost-effective, both at the societal level and at the level of the dialysis clinic.

University of Michigan Kidney Epidemiology and Cost Center. (2017). **End-stage renal disease patient-reported outcomes technical expert panel summary report**. *Prepared for the Centers for Medicare and Medicaid Services.* Accessed August 30, 2021 https://www.dialysisdata.org/sites/default/files/content/ESRD_Measures/ESRD_Patient_Reported_Outcomes_TEP_Summary_Report.pdf

Waters, D and VS Sierpina. **Goal-directed health care and the chronic pain patient: A new vision of the healing encounter**. *Pain Physician*. 2006 9(4): 353-360.

We introduce a new way to engage the patient with chronic pain, Goal-Directed Health Care (G-DHC). Identifying the patient's major life goals during the medical interview is the key element of this approach along with connecting these life goals to specific health-related goals. The implementation of G-DHC is a shift in process from the usual focus on disease-related goals such as relief of pain, titrating narcotic refills, and working on condition management to broader, long-term, personal goals. It emphasizes the importance of identifying the global life goals of patients and the reasons they wish to be well for and what they would do with improved health once they had it. Utilizing these life goals as a point of reference, discussion, and motivation makes clearer what specified health goals mean, whether or not the patient is ready to work on them, and most significantly, what the underlying motivation is to participate in their own care. We anticipate such a model of patient-centered care will shift the dynamic of the medical encounter with the patient with chronic pain to one that is ultimately more productive and satisfying for both patient and physician. Illustrations of cases, questions to ask patients, and a detail of the process may allow the reader to adopt this method into their practice.

Woo K, Lok CE. **New insights into dialysis vascular access: what is the optimal vascular access type and timing of access creation in CKD and dialysis patients?** Clin J Am Soc Nephrol. 2016;11(8):1487-1494.

Optimal vascular access planning begins when the patient is in the predialysis stages of CKD. The choice of optimal vascular access for an individual patient and determining timing of access creation are dependent on a multitude of factors that can vary widely with each patient, including demographics, comorbidities, anatomy, and personal preferences. It is important to consider every patient’s ESRD life plan (hence, their overall dialysis access life plan for every vascular access creation or placement).

Optimal access type and timing of access creation are also influenced by factors external to the patient, such as surgeon experience and processes of care. In this review, we will discuss the key determinants in optimal access type and timing of access creation for upper extremity arteriovenous fistulas and grafts.

Wuerth D.B., Finkelstein S.H., Schwetz O., Carey H., Kliger A.S., Finkelstein F.O. (2002). **Patients' Descriptions of Specific Factors Leading to Modality Selection of Chronic Peritoneal Dialysis or Hemodialysis.** *Peritoneal Dialysis International*. 22(2):184-190. doi:10.1177/089686080202200204

Objectives: There has been increasing interest in understanding how patients with chronic renal failure choose between chronic peritoneal dialysis (CPD) and hemodialysis (HD) for renal replacement therapy. The purpose of the present study was to examine the influences and specific factors that patients identify as significant in choosing a specific dialysis modality for treatment of their end-stage renal disease (ESRD). Patients and Design: 40 patients (20 CPD, 20 HD) who had started dialysis within the preceding 6 months were randomly selected to participate in the study. A structured interview was conducted with the patients, discussing and exploring what factors patients thought were important in helping them decide their treatment modality. The format of the interview was open-ended. Based on patients' comments, a taxonomy of the specific factors that influenced the patients' decisions was developed. Setting: The study was conducted in a freestanding CPD unit and two freestanding HD units. Results: All 20 CPD patients reported choosing their treatment modality; only 8 of the 20 HD patients reported having a choice of treatment modality. 18 of the 22 patients who participated in predialysis educational programs opted for CPD. 83% of the patients reported that their physician was important in influencing their treatment choice; however, the CPD patients relied more on written material and the opinions of their spouse/significant other or other family members in making their decisions. Issues of autonomy and control were important for 95% of patients choosing CPD. Both CPD and HD patients cited a variety of treatment-specific factors. The three most frequently cited reasons for choosing CPD were (1) flexibility of schedule (19 patients), (2) convenience of performing CPD in their own home (19 patients), and (3) the option of doing dialysis at night while sleeping (8 patients). The 8 HD patients who selected their treatment modality cited the desirability of having a planned schedule (7 patients) and letting nurses or other take care of them (5 patients). Conclusions: The present study explored factors perceived by patients as being important in determining their choice of renal replacement therapy. A taxonomy of patient influences and concerns has been developed to provide caregivers with a framework to structure their educational strategies and assist patients with progressive renal failure in making an informed choice of therapeutic modality for their ESRD treatment.

Zee, J., Zhao, J., Subramanian, L., Bryant, N., McCall, M., Restovic, Y., Torres, D., Robinson, B., Pisoni, R., Tentori, F. (2018). Perceptions about the dialysis modality decision process among peritoneal dialysis and in-center hemodialysis patients. *BMC Nephrol, 19*(1), 298.

Patients reaching end-stage renal disease must make a difficult decision regarding renal replacement therapy (RRT) options. Because the choice between dialysis modalities should include patient preferences, it is critical that patients are engaged in the dialysis modality decision. As part of the Empowering Patients on Choices for RRT (EPOCH-

RRT) study, we assessed dialysis patients' perceptions of their dialysis modality decision-making process and the impact of their chosen modality on their lives. A 39-question survey was developed in collaboration with a multi-stakeholder advisory panel to assess perceptions of patients on either peritoneal dialysis (PD) or in-center hemodialysis (HD). The survey was disseminated to participants in the large US cohorts of the Dialysis Outcomes and Practice Patterns Study (DOPPS) and the Peritoneal DOPPS (PDOPPS). Survey responses were compared between PD and in-center HD patients using descriptive statistics, adjusted logistic generalized estimating equation models, and linear mixed regression models. Six hundred fourteen PD and 1346 in-center HD participants responded. Compared with in-center HD participants, PD participants more frequently reported that they were engaged in the decision-making process, were provided enough information, understood differences between dialysis modalities, and felt satisfied with their modality choice. PD participants also reported more frequently than in-center HD participants that partners or spouses (79% vs. 70%), physician assistants (80% vs. 66%), and nursing staff (78% vs. 60%) had at least some involvement in the dialysis modality decision. Over 35% of PD and in-center HD participants did not know another dialysis patient at the time of their modality decision and over 60% did not know the disadvantages of their modality type. Participants using either dialysis modality perceived a moderate to high impact of dialysis on their lives. PD participants were more engaged in the modality decision process compared to in-center HD participants. For both modalities, there is room for improvement in patient education and other support for patients choosing a dialysis modality.

Importance to Measure and Report: Gap in Care/Disparities (1b.01 - 1b.05)

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.

For people that are on chronic dialysis, discussion of patient life goals with their dialysis facility care team can lead to better understanding of these goals at the facility-, provider- and patient-levels, facilitating decision making that explicitly takes these goals into consideration and enhances patient shared decision making about modality selection, vascular access, and other treatment options for their dialysis and ESRD care. Discussing personal life goals that are important to the patient also results in patient-centered care which is an outcome in its own right (Blake and Brown 2020). This measure is intended to provide information to users about one particular aspect of patient-centered care, identification and discussion of patient life goals with those delivering care.

This new measure, the PaLS, will provide a way for providers, payers, and others to assess how well a facility is doing in identifying and discussing life goals with ESRD chronic dialysis patients as part of the treatment planning process. For example, the PaLS can provide care teams with information about important life goals that they can use to ensure that dialysis treatment modality is synergistic with a patient's specific life goals. A discussion of life goals between a patient and his/her/their provider(s) can inform modality option selection, such as a home dialysis therapy (home hemodialysis, peritoneal dialysis) or kidney transplant, and help ensure that modality selection is aligned with the patient's life goals. Results from the PaLS can also provide payers and other stakeholders national level information about patients' experience with discussions of life goals (or absence of) with their care team, and whether patients feel their treatments are aligned with their life goals.

Reference:

Blake, PG., and Brown, EA. **Person-centered peritoneal dialysis prescription and the role of shared decision-making**. 2020. Peritoneal Dialysis Int, published ahead of print, 1-8.

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.

Performance score data are provided for both the calibration sample and the subsequent validation testing sample.

Calibration sample distribution of *t*-scores:

Dates of data collection: Calibration sample -- 06-03-2020 - 12-29-2020
Number of measured entities (participants): 517

Table 1: Distribution of *t*-scores (overall and by decile)

*	N	Mean	Standard deviation	Minimum	Maximum	Lower Quartile	Upper Quartile	IQR
Overall	517	50.0	9.5	25.9	69.8	43.2	55.8	12.7
*	*	*	*	*	*	*	*	*
Decile	*	*	*	*	*	*	*	*
1	52	34.7	2.9	25.9	37.5	33.4	36.9	3.5
2	51	39.8	1.1	37.5	41.7	39.0	40.8	1.7
3	52	43.2	0.8	41.7	44.4	42.5	43.9	1.4
4	52	45.8	0.8	44.5	47.0	45.1	46.6	1.5
5	53	48.4	0.8	47.1	49.6	47.7	49.2	1.5
6	51	51.1	0.7	49.6	52.0	50.4	51.7	1.2
7	51	53.5	0.9	52.0	55.1	52.8	54.4	1.6
8	52	55.9	0.7	55.1	57.3	55.3	56.4	1.1
9	53	60.1	1.7	57.6	62.8	58.5	61.5	3.0
10	50	68.0	2.4	63.1	69.8	66.4	69.8	3.4

IQR = Upper quartile minus lower quartile
*Cells intentionally left blank

Validation testing sample distribution of *t*-scores:

Dates of data collection: Validation testing sample at baseline assessment -- 04-07-2021 - 11-06-2021
Number of measured entities (participants): 420

Table 2: Distribution of *t*-scores (overall and by decile)

*	N	Mean	Standard deviation	Minimum	Maximum	Lower Quartile	Upper Quartile	IQR
Overall	420	50.1	9.7	25.9	69.8	43.5	56.4	12.9
*	*	*	*	*	*	*	*	*
Decile	*	*	*	*	*	*	*	*
1	42	34.3	3.2	25.9	37.8	32.6	36.7	4.1
2	42	39.7	1.4	37.8	41.9	38.5	41.2	2.7
3	42	43.4	0.8	42.0	44.5	42.7	44.1	1.4
4	42	45.8	0.7	44.6	47.0	45.1	46.5	1.3
5	42	48.3	0.7	47.2	49.5	47.6	48.8	1.1
6	42	50.7	0.9	49.5	52.0	49.9	51.4	1.5
7	43	53.6	0.9	52.0	54.9	52.8	54.5	1.6
8	41	56.3	0.8	55.0	57.7	55.3	56.8	1.5
9	42	60.9	1.8	57.9	63.8	59.3	62.8	3.5
10	42	68.1	2.0	63.9	69.8	66.4	69.8	3.4

IQR = Upper quartile minus lower quartile
*Cells intentionally left blank

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.

N/A

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.

Disparities data are provided for both the calibration sample and the subsequent validation testing sample.

Calibration sample: distribution of *t*-scores:

Dates of data collection: Calibration sample -- 06-03-2020 - 12-29-2020

Number of measured entities (participants): N=517
Number of participants with self-reported race, ethnicity, sex, and education: N=517
Number of participants where matching allowed for the identification of dialysis modality: N=463

Table 3: Distribution of *t*-scores by demographic factors

*	N	Mean	Standard deviation	Minimum	Maximum	Lower Quartile	Upper Quartile	IQR
Race	*	*	*	*	*	*	*	*
White	364	49.6	9.4	25.9	69.8	43.1	55.3	12.2
Black/African American	94	50.9	9.3	33.7	69.8	44.3	56.0	11.7
Native American or Alaska Native	8	50.2	12.1	32.5	69.8	41.5	58.4	16.8
Asian or Pacific Islander	17	51.4	8.1	35.1	69.8	47.4	55.9	8.4
More than one	10	49.2	9.9	25.9	61.5	46.0	55.3	9.3
Do not wish to report	24	51.9	11.2	32.6	69.8	42.5	61.2	18.6
Ethnicity	*	*	*	*	*	*	*	*
Non-Hispanic	426	49.9	9.3	25.9	69.8	43.3	55.8	12.5
Hispanic	48	51.9	11.5	32.1	69.8	42.6	61.1	18.5
Do not wish to report	43	48.9	8.9	25.9	66.0	42.6	55.3	12.7
Sex	*	*	*	*	*	*	*	*
Female	245	50.4	9.5	25.9	69.8	43.6	56.0	12.4
Male	272	49.6	9.5	25.9	69.8	43.0	55.5	12.5
Education	*	*	*	*	*	*	*	*
Some high school but did not graduate	12	57.4	7.3	47.6	69.8	49.3	62.2	12.9
High school graduate or GED	90	49.0	8.9	30.4	69.8	43.8	54.9	11.2
Some college or 2-year degree	192	49.7	9.3	25.9	69.8	42.9	55.3	12.5
4-year college degree	127	50.5	9.5	25.9	69.8	44.4	55.7	11.3
Master's degree or higher	92	50.2	10.2	32.1	69.8	41.8	57.2	15.4
Do not wish to report	4	42.8	7.9	32.6	51.7	37.4	48.2	10.8

IQR = Upper quartile minus lower quartile
*Cells intentionally left blank

There were small differences in mean scores within demographic groups, including race, ethnicity (non-Hispanic participant mean scores were 2 points lower than Hispanic participant scores), sex (males had mean scores that were 0.8 points lower than females) and level of education; there was no statistically significant difference in *t*-scores between groups.

Validation sample distribution of *t*-scores:

Dates of data collection: Validation testing sample at baseline assessment -- 04-07-2021 - 11-06-2021

Number of measured entities (participants): N=420
Number of participants with self-reported race, ethnicity, sex, and education: N=420
Modality: N=335

Table 4: Distribution of *t*-scores by demographics factors

*	N	Mean	Standard deviation	Minimum	Maximum	Lower Quartile	Upper Quartile	IQR
Race	*	*	*	*	*	*	*	*
White	293	50.1	9.9	25.9	69.8	43.0	56.7	13.7
Black/African American	91	50.2	9.8	30.4	69.8	44.2	56.2	12.0
Native American or Alaska Native	9	49.9	7.3	41.5	65.1	44.3	53.7	9.5
Asian or Pacific Islander	8	53.6	11.4	37.2	69.8	45.0	61.9	16.9
More than one	8	47.2	5.2	38.1	55.3	44.6	50.6	6.1
Do not wish to report	11	48.3	7.3	38.0	59.3	39.4	54.5	15.0
Ethnicity	*	*	*	*	*	*	*	*
Non-Hispanic	356	50.1	9.6	25.9	69.8	43.7	56.0	12.3
Hispanic	32	50.7	9.4	37.0	69.8	42.1	57.3	15.2
Do not wish to report	32	49.7	10.8	25.9	69.8	40.9	57.0	16.1
Sex	*	*	*	*	*	*	*	*
Female	238	50.1	9.5	25.9	69.8	43.3	56.7	13.4
Male	182	50.1	9.9	25.9	69.8	44.1	55.3	11.2
Education	*	*	*	*	*	*	*	*
Some high school but did not graduate	7	55.2	9.5	38.3	69.8	51.3	59.5	8.2
High school graduate or GED	86	49.6	10.0	30.4	69.8	43.2	55.3	12.1
Some college or 2-year degree	162	49.8	9.6	25.9	69.8	43.5	56.4	12.9
4-year college degree	89	50.8	9.6	32.2	69.8	44.1	56.7	12.6
Master's degree or higher	72	49.9	9.9	25.9	69.8	42.5	55.3	12.8
Do not wish to report	4	53.3	4.8	48.3	57.5	49.1	57.4	8.3

IQR = Upper quartile minus lower quartile
*Cells intentionally left blank

There were small differences in mean scores within demographic groups, including race, ethnicity (non-Hispanic participant mean scores were 0.6 points lower than Hispanic participant scores), and level of education; there was no statistically significant difference in *t*-scores between groups.

Examination of Disparities of PaLS scores

The percentage of participants that had poor life goals scores (i.e., those participants that were not satisfied with life goals discussions) was calculated by dividing the number of participants with a *t*-score ≤40 in each stratification by the total number of participants in that stratification and then multiplying by 100. To provide a robust evaluative percentage, the number of participants with poor life goals scores included only participants who had answered all six life goals items. We expected >16% of participants to have poor/unsatisfactory experience (i.e., scores that were ≥1 SD away from the mean in the direction of worse experience) to have PaLS scores ≥1 SD below the mean (i.e., *t*-score ≤40; Heaton et al., 2004). Note that, assuming a normal distribution, PaLS scores that are ≤40 would warrant follow-up by the facility. Specifically, scores ≤40 (i.e., ≥1 *SD* below the mean) suggest patient perceptions of care discussions are worse than 84% of their peers.

Moderate to large Cohen's *d* effects sizes (i.e., values between 0.50 and 0.79 were considered "moderate;" values ≥0.80 were considered "large") also provide psychometric support for PaLS *t*-score (Cohen, 1988).

Score disparity analysis for dual eligibility

We did not expect differences in the responses to the six Likert-type scorable PaLS items between participants with and without dual eligibility, i.e., we did not expect to see mean PaLS score disparities for this variable.

Calibration sample

In the calibration sample, there was no statistically significant difference in *t*-scores between participants in the no dual eligibility group compared to the dual eligibility group. 16.3% of participants that were not dual eligible had a poor PaLS score compared to 13.7% of participants that were dual eligible with a poor PaLS score.

Validation testing sample

In the validation testing sample, there were no statistically significant differences in *t*-scores between participants that were not dual eligible compared to participants that were dual eligible. 17.6% of participants that were not dual eligible had a poor PaLS score compared to 12.8% of participants that with dual eligibility with a poor PaLS score.

Table 5: Score disparity analysis for dual eligibility

*	PaLS Scores	*	*	PaLS Scores	*	*	*	*	*
*	No Dual	*	*	Dual	*	*	*	*	*
*	N	Mean (SD) PaLS *t*-score	% of participants with poor PaLS *t*-score	N	Mean (SD) PaLS *t*-score	% of participants with poor PaLS *t*-score	t	p-value	Cohen's d
Calibration	355	49.9 (9.3)	16.3	102	50.1 (9.3)	13.7	-0.13	0.89	-0.01
Validation	245	49.4 (9.8)	17.6	86	50.0 (8.8)	12.8	-0.50	0.61	-0.06

* Cells intentionally left blank

Score disparity analysis for education

We expected to see a disparity for PaLS score based on education. Specifically, we expected participants with a 4-year college degree or more to have higher (better) *t*-scores than participants with less than a 4-year college degree.

Calibration sample

In the calibration sample, there was no statistically significant difference in *t*-scores for participants with a 4-year college degree or more vs. less than a 4-year college degree. 13.5% of participants with a 4-year college degree or more had a poor PaLS score compared to 16.6% of participants with less than a 4-year college degree with a poor PaLS score. While the results are not as hypothesized, we believe that it is reasonable to conclude that the discussion of life goals between facilities and patients should not differ by level of education.

Validation testing sample

In the validation testing sample, there was no statistically significant difference in *t*-scores for participants with a 4-year college degree or more vs. less than a 4-year college degree. 15.6% of participants with a 4-year college degree or more had a poor PaLS score compared to 16.0% of participants with less than a 4-year college degree with a poor PaLS score. While the results are not as hypothesized, we believe that it is reasonable to conclude that the discussion of life goals between facilities and patients should not differ by level of education.

Table 6: Score disparity analysis for education

*	PaLS Scores	*	*	PaLS Scores	*	*	*	*	*
*	4 year college degree or more	*	*	Less than 4 year college degree	*	*	*	*	*
*	N	Mean (SD) PaLS *t*-score	% of participants with poor PaLS *t*-score	N	Mean (SD) PaLS *t*-score	% of participants with poor PaLS *t*-score	t	p-value	Cohen's d
Calibration	215	50.6 (9.7)	13.5	295	49.8 (9.3)	16.6	0.98	0.16	0.09
Validation	160	50.5 (9.7)	15.6	256	49.9 (9.7)	16.0	0.65	0.26	0.07

* Cells intentionally left blank

The results of testing for score-based disparities on the PaLS did not indicate any significant disparities in life goals for individuals with or without dual eligibility or by level of education.

References:

Cohen, J. (1988) *Statistical power analysis for the behavioral sciences*. New York, New York: Lawrence Erlbaum Associates

Heaton, R. K., Miller, S. W., Taylor, J. T., & Grant, I. (2004). *Revised comprehensive norms for an expanded Halstead-Reitan Battery: Demographically adjusted neuropsychological norms for African American and Caucasian adults*. Lutz, FL: Psychological Assessment Resources, Inc.

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.

N/A

Scientific Acceptability: Maintenance (2ma.01 - 2ma.04)

2ma.01) Indicate whether additional empirical reliability testing at the accountable entity level has been conducted. If yes, please provide results in the following section, Scientific Acceptability: Reliability - Testing. Include information on all testing conducted (prior testing as well as any new testing).

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

Current Submission:

Updated testing information here.

Previous Submission:

Testing from the previous submission here.

☐ Yes
☐ No

2ma.02) Indicate whether additional empirical validity testing at the accountable entity level has been conducted. If yes, please provide results in the following section, Scientific Acceptability: Validity - Testing. Include information on all testing conducted (prior testing as well as any new testing).

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

Current Submission:

Updated testing information here.

Previous Submission:

Testing from the previous submission here.

☐ Yes
☐ No

2ma.03) For outcome, patient-reported outcome, resource use, cost, and some process measures, risk adjustment/stratification may be conducted. Did you perform a risk adjustment or stratification analysis?

☐ Yes
☐ No

2ma.04) For maintenance measures in which risk adjustment/stratification has been performed, indicate whether additional risk adjustment testing has been conducted since the most recent maintenance evaluation. This may include updates to the risk adjustment analysis with additional clinical, demographic, and social risk factors.

Please update the Scientific Acceptability: Validity - Other Threats to Validity section.

Note: This section must be updated even if social risk factors are not included in the risk adjustment strategy.

☐ Yes - Additional risk adjustment analysis is included
☐ No additional risk adjustment analysis included

Scientific Acceptability: Reliability - Testing (2a.01 - 2a.12)

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 Battelle staff at PQMsupport@battelle.org 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 Battelle staff at PQMsupport@battelle.org with any questions.
- 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 the 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 Scientific Acceptability sections. For example:

Current Submission:

Updated testing information here.

Previous (Year) Submission:

Testing from the previous submission here.

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

☐ Assessment Data
☒ Claims
☐ Electronic Health Data
☐ Electronic Health Records
☒ Instrument-Based Data
☐ Management Data
☐ Other (please specify here:)
☐ Paper Medical Records
☒ Registry Data

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).

ESRD data were used to confirm self-reported ESRD status and treatment information. These data were derived from the UM-KECC database, a national ESRD patient database that includes information from the Renal Management Information System (REMIS), CROWNWeb facility-reported clinical and administrative data (including CMS-2728 Medical Evidence Form, CMS-2746 Death Notification Form, and CMS-2744 Annual Facility Survey Form and patient tracking data), the Medicare Enrollment Database (EDB), and Medicare dialysis claims data (primarily outpatient). In addition, the UM-KECC database includes transplant data from the Scientific Registry of Transplant Recipients (SRTR), data from the Nursing Home Minimum Dataset, data from the Quality Improvement Evaluation System (QIES) Business Intelligence Center (QBIC; which includes Provider and Survey and Certification data from Automated Survey Processing Environment [ASPEN]), and data from the Dialysis Facility Care Compare (DFCC).

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

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

Calibration sample:

The calibration sample was collected: 06-03-2020 – 12-29-2020. This is the primary development sample for the PaLS measure and was the sample that was used to inform a patient-level *t*-score (based on responses to the six Likert-type scorable PaLS items).

Validation testing sample:

The validation sample was derived from data collected: 04-07-2021 – 04-24-2022. These data were used to confirm the reliability and validity of a patient-level *t*-score (based on responses to the six Likert-type scorable PaLS items).

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.

☐ Accountable Care Organization
☐ Clinician: Group/Practice
☐ Clinician: Individual
☐ Facility
☐ Health Plan
☐ Integrated Delivery System
☒ Other (specify: US Chronic Dialysis Population (patient-level). The measure testing was performed on a sample that reflected the US chronic dialysis population at the patient-level.)

☐ Population: Community, County or City
☐ Population: Regional and State

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.

Calibration sample:

Five-hundred and seventeen (N=517) participants were included in the national field-testing data collection that established the calibration sample. This was a cross-sectional sample and did not include longitudinal follow-up. Participants were recruited using a snowball sampling approach in which recruitment materials were sent via email to several dialysis organizations, nephrology professional organizations, and kidney patient advocacy groups. We asked these stakeholders to disseminate recruitment materials to ESRD patients affiliated with their clinics, patient organizations, or via their own professional or advocacy networks. We only included surveys that met participation eligibility criteria and where participant consent was obtained. Upon request, we provided paper surveys to participants.

Validation testing sample:

Data collection was longitudinal and included baseline, 3-month, and 6-month follow-up study assessments. Four-hundred and twenty (N=420) participants were included in the national data collection that established the validation testing sample at baseline. One hundred and eighty-three (n=183) participants completed the 3-month follow up and one hundred and sixty-seven (n=167) participants completed the 6-month follow up. At each time point, participants completed the PaLS as well as additional surveys that were used to examine different aspects of psychometric validity (e.g., see section 2b.03: Known-groups validity for PROMIS measures). Participants were recruited using a snowball sampling approach in which recruitment materials were sent via email to several dialysis organizations, nephrology professional organizations, and kidney patient advocacy groups. We asked these stakeholders to disseminate recruitment materials to ESRD patients affiliated with their clinics, patient organizations, or via their own professional or advocacy networks. We only included surveys that met participant eligibility criteria and where participant consent was obtained. Upon request, we provided paper surveys to participants.

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.

Calibration sample:

Only participants that met eligibility criteria and provided consent were included in this sample. There were 517 participants in the calibration sample. The average age of participants was 61.6 years old; 47.4% of participants were female; 70.4% were White, 18.1% were Black, 6.8% were Other/Multi-racial/Unknown/Missing and 4.6% did not wish to report race; 9.3% were Hispanic; and 42.4% reported 4 years or more of college.

Validation testing sample:

Only participants that met eligibility criteria and provided consent were included in this sample. There were 420 participants in the validation testing sample at baseline. The average age of participants was 59.6 years old; 56.7% of participants were female; 69.8% were White, 21.7% were Black, 6.0% were Other/Multi-racial/Unknown/Missing and 2.6% did not wish to report race; 7.6% were Hispanic; and 38.3% reported 4 years or more of college.

At 3-month follow up there were 183 participants. The average age of participants was 60.8 years old; 51.4% of participants were female; 73.2% were White, 20.2% were Black, 3.8% were Other/Multi-racial/Unknown/Missing and 2.7% did not wish to report race; 6.6% were Hispanic; and 43.2% reported 4 years or more of college.

At 6-month follow up there were 167 participants. The average age of participants was 61.9 years old; 55.1% of participants were female; 74.9% were White, 18.9% were Black, 3.6% were Other/Multi-racial/Unknown/Missing and 3.0% did not wish to report race; 6.0% were Hispanic;

and 41.9% reported 4 years or more of college.

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.

These two samples were obtained at different time points as noted above.

Calibration sample:

The calibration sample was derived from the field-testing data collected: 06-03-2020 – 12-29-2020. This is the primary development sample for the PaLS measure, which used classical test theory and item response theory analytical approaches. This was the sample that was used to estimate item parameters for calculating a patient-level *t*-score using responses to the six Likert-type scorable PaLS items. This sample was also used to generate preliminary validity data for the PaLS measure.

Validation testing sample:

The validation testing sample was derived from data collected: 04-07-2021 – 04-24-2022. These data were used to validate the patient-level *t*-scores (derived from responses on the six Likert-type scorable PaLS items). Baseline and follow-up data were collected for this sample, along with administration of additional surveys to examine different aspects of psychometric reliability and validity of the patient-level PaLS *t*-score (e.g., see section 2b.03: Known-groups validity for PROMIS measures).

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.

Note: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a.09 check patient or encounter-level data; in 2a.010 enter “see validity testing section of data elements”; and enter “N/A” for 2a.11 and 2a.12.

Calibration sample and Validation testing sample:

Age, Education, Sex, and Race were examined for item bias (using DIF analyses), and dual eligibility status was used to examine score disparities. Social risk factors were analyzed in both the calibration and validation testing samples. Social risk factors were assessed for potential sources of disparities in section 1b.04.

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

Choose one or both levels.

☒ Patient or Encounter-Level (e.g., inter-abstractor reliability; data element reliability

must address ALL critical data elements)

☐ Accountable Entity Level (e.g., signal-to-noise analysis)

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.

Calibration sample

The PaLS was developed according to established PRO development methodology (*PROMIS® Instrument Development and Psychometric Evaluation Scientific Standards,* 2019). This process relies on classical test theory (CTT) and item response theory (IRT) analyses. Specifically, we used exploratory and confirmatory factor analyses (EFA, CFA) to ensure that the six Likert-type scorable PaLS items were unidimensional (a condition for generating an IRT-based *t*-score; Cook et al., 2009; McDonald, 1999; Reise et al., 2007). For EFA, we considered the item set to have unidimensional characteristics if the ratio of eigenvalue 1 to eigenvalue 2 was ≥4 and the proportion of variance accounted for by eigenvalue 1 was ≥0.40 (Hu & Bentler, 1999; Kline, 2005; Lai et al., 2006, 2011).

We excluded items with sparse cells (response categories with n<5 participants), items with low item-adjusted total score correlations (<0.40), and items that were non-monotonic (monotonicity was examined using non-parametric IRT models of item-rest plots and expected score by latent trait plots; Testgraf Software; Ramsay, 2000).

For CFA, we considered an item set to be unidimensional if: the comparative fit index (CFI) was ≥0.90, the Tucker-Lewis index (TLI) was ≥0.90, and the root mean square error of approximation (RMSEA) was <0.10 (Cook et al., 2009; Hu & Bentler, 1999; Kline, 2005; Bentler, 1990; Lai et al., 2011, 2014). For comparative fit purposes, we also obtained the chi-square value for model fit and its associated p value. We considered items for removal if they had low factor loadings (lx<0.50) or were locally dependent (i.e., residual correlation >0.20; correlated error modification index ≥100; Cook et al., 2009; Hair et al., 2009; Kaplan, 1989; Luijben & Boomsma, 1988; McDonald, 1999; Reise et al., 2007; Saris et al., 1987, 2009; Whittaker, 2012).

Next, a graded response model (GRM) was used to estimate item parameters. We excluded items with significant misfit (S-X2 / df effect size >3; Crisan et al., 2017; Drasgow et al., 2017; Stark et al., 2006; Zhao, 2017). We also excluded items with impactful differential item functioning (DIF). Items were evaluated for DIF using the lordif R package, version 0.3-3 (Choi et al., 2011). This statistical package iteratively applies a hybrid logistic ordinal regression (LOR) and IRT approach. Items with McFadden pseudo-R2 change ≥0.02 were flagged for DIF. DIF analyses were conducted for age (median split), education (4 year college degree or more versus less than a 4 year college degree), sex (male versus female), race (white versus other) and modality (in-center versus at home).

We investigated the practical impact of flagged DIF items by quantifying change in individual scores when adjusted for DIF. Two scores were calculated for each patient, one based on item parameters calibrated for the entire sample and another in which items flagged for DIF were calibrated separately (DIF-adjusted). For example, items flagged for DIF based on sex would

have parameters calibrated separately in males and females. Scoring impact was evaluated based on: (a) Pearson correlation, (b) mean difference, (c) root mean squared difference (RMSD), and (d) percentage of score differences > their associated unadjusted score standard error (SE), i.e., >2% of DIF-corrected vs. uncorrected score differences exceeding individual case uncorrected score standard errors (Edelen et al., 2007).

These IRT-based analyses were followed by a final CFA analysis designed to confirm that the final item set remained essentially unidimensional (using the same item-level and overall model fit criteria outlined above).

Following this process, we converted the theta-based score to *t*-score. An IRT score is initially on a theta metric, with a mean of 0 and a SD of 1. We converted the theta score to standardized scores on the *t*-score metric (M=50; SD=10). The conversion from a theta score to a *t*-score can be obtained using the following linear transformation: *t*-score = (theta x 10) +50.

The *t*-score was then used to examine the preliminary reliability and validity of the PaLS *t*-score. Specific methods are described below for each level of testing.

Cronbach's alpha was used to establish the internal consistency of the six Likert-type scorable PaLS items. Cronbach's alpha is defined below, where N is the number of items, $\bar{c}$ is the average inter-item covariance among items and $\bar{v}$ is the average variance (Cronbach, 1951).

$$\alpha = \frac{N\bar{c}}{\bar{v} + (N-1)\bar{c}}$$

Marginal reliability is calculated as the ratio of the true score variance to the total variance, expressed with respect to the estimated latent abilities. Marginal reliability refers to the reliability with regard to the population as a whole. Marginal reliability is defined below with a density of *g* and variance of 1 (Andersson & Xin, 2018).

$$\rho_\theta(\alpha) = \int_{-\infty}^{\infty} \frac{I(\theta;\, \alpha)}{I(\theta;\, \alpha) \,+\, 1} g(\theta) d\theta$$

Response pattern reliability is the reliability based on the median *t*-score standard error and *t*-score standard deviation for *t*-scores ±3 standard deviations from the mean. Thus, response pattern reliability was calculated using the median *t*-score standard error and *t*-score standard deviation where *t*-scores were between 20 and 80. Participants were not excluded if they did not answer all six Likert-type scorable PaLS items. Response pattern reliability is defined below (Pilkonis et al., 2014).

$$Reliability = 1 - \frac{SE^2}{SD^2}$$

Validation testing sample

Cronbach's alpha, marginal reliability and response pattern reliability were also examined for the patient-level *t*-score in the validation testing sample using the methods outlined in the calibration sample description above.

In addition, we calculated test-retest reliability using intraclass correlation coefficients (ICC). Minimum acceptable criteria for test-retest reliability was set at ≥0.70 for intraclass correlations (Cohen, 1969). ICC was calculated using the SAS macro intracc.sas.

We estimated two forms of the ICC, one including both systematic and random error in the estimation denominator and one including only random error. We used a two-way mixed effects ICC model, where people effects were randomized and measure effects were fixed. We report our test-retest reliability estimates based on our ICC results from systematic plus random error-based estimations.

Minimal detectable change (DC_{95}) and standard error of measurement (SEM) were calculated for the PaLS scores. Minimal detectable change identifies the amount of change that can be detected with 95% confidence that it is not due to measurement error from baseline to the time point of interest, in this case 3 months and 6 months. The SEM percentage is calculated by dividing the SEM by the mean of all observations across time points and multiplying by 100. <10% indicates good (i.e. acceptable) measurement error.

Minimal detectable change (DC_{95}) is calculated using the equation:

$$DC_{95} = SEM * 1.96 * \sqrt{2}$$

Standard Error of Measurement (SEM) is calculated using the equation:

$$SEM = 1.96 * \sqrt{1 - ICC}$$

References:

Andersson, B., & Xin, T. (2018). Large sample confidence intervals for item response theory reliability coefficients. *Educational and psychological measurement*, *78*(1), 32–45. https://doi.org/10.1177/0013164417713570

Bentler, P.M. (1990). *Comparative fit indexes in structural models. Psychological Bulletin, 107*(2), 238-246.

Choi, S.W., Gibbons, L.E., & Crane, P.K. (2011). Lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and monte carlo simulations. *Journal of statistical software*, *39*(8), 1–30. https://doi.org/10.18637/jss.v039.i08

Cohen J. (1969). *Statistical Power Analysis for the Behavioral Sciences.* New York: Academic Press.

Cook, K.F., Kallen, M.A., & Amtmann, D. (2009). Having a fit: Impact of number of items and distribution of data on traditional criteria for assessing IRT's unidimensionality assumption. *Quality of Life Research, 18*(4), 447-460.

Crişan, D.R., Tendeiro, J.N., & Meijer, R.R. (2017). Investigating the practical consquences of model misfit in unidimensional IRT models. *Applied Psychological Measurement*, *41*(6), 439–455. https://doi.org/10.1177/0146621617695522

Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*,

16, 297-334.

Drasgow, F., Levine, M. V., Tsien, S., Williams, B. & et al. (1995). Fitting polytomous item response theory models to multiple-choice tests. *Applied Psychological Measurement, 19,* 143-165. doi:10.1177/014662169501900203

Edelen, M.O., & Reeve, B.B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of Life Research*, *16*(Suppl 1), 5–18. https://doi.org/10.1007/s11136-007-9198-0

Hair, J.F., Black, W.C., Babin, B.J., & Anderson, R.E. (Eds.). (2009). *Multivariate data analysis* (7th edition ed.). Upper Saddle River, NJ: Prentice Hall.

Hatcher, L. (1994). *A step-by-step approach to using SAS for factor analysis and structural equation modeling*. Cary, NC: SAS Institute, Inc.

Hu, L.T., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling-a Multidisciplinary Journal, 6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Kaplan D. (1989). Model modification in covariance structure analysis: Application of the expected parameter change statistic. *Multivariate behavioral research*, *24*(3), 285–305. https://doi.org/10.1207/s15327906mbr2403_2

Kline, R.B. (2005). *Principles and practice of structural equation modeling, Second Edition*. New York: Guilford Press.

Lai, J.S., Cella, D., Choi, S., Junghaenel, D.U., Christodoulou, C., Gershon, R., & Stone, A. (2011). How item banks and their application can influence measurement practice in rehabilitation medicine: A PROMIS fatigue item bank example. *Archives of physical medicine and rehabilitation*, *92*(10 Suppl), S20–S27. https://doi.org/10.1016/j.apmr.2010.08.033

Lai, J.S., Crane, P.K., & Cella, D. (2006). Factor analysis techniques for assessing sufficient unidimensionality of cancer related fatigue. *Qual Life Res, 15*(7), 1179-1190.

Lai, J.S., Zelko, F., Butt, Z., Cella, D., Kieran, M.W., Krull, K.R., Magasi, S., & Goldman, S. (2011). Parent-perceived child cognitive function: results from a sample drawn from the US general population. *Child's Nervous System, 27*(2), 285-293.

Lai, J.S., Zelko, F., Krull, K., Cella, D., Nowinski, C., Manley, P., & Goldman, S. (2014). Parent-reported cognition of children with cancer and its potential clinical usefulness. *Quality of Life Research, 23*(4), 1049-1058.

Luijben, T.C., & Boomsma, A. (1988). Statistical guidance for model modification in covariance structure analysis. *Compstat*, 335-340.

McDonald, R.P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Pilkonis, P.A., Yu, L., Dodds, N.E., Johnston, K.L., Maihoefer, C.C., & Lawrence, S.M.

(2014). Validation of the depression item bank from the Patient-Reported Outcomes Measurement Information System (PROMIS) in a three-month observational study. *Journal of psychiatric research*, *56*, 112–119. https://doi.org/10.1016/j.jpsychires.2014.05.010

PROMIS® Instrument Development and Psychometric Evaluation Scientific Standards, http://www.healthmeasures.net/images/PROMIS/PROMISStandards_Vers2.0_Final.pdf. (Vol. 2019).

Ramsay, J.O. (2000). *TestGraf: A program for the graphical analysis of multiple choice test and questionnaire data.* Montreal, Quebec: McGill University. Retrieved from www.psych.mcgill.ca/misc/fda/downloads/testgraf/TestGraf98.doc

Reise, S.P., Morizot, J., & Hays, R.D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*, *16*(Suppl 1), 19–31. https://doi.org/10.1007/s11136-007-9183-7

Saris, W.E., Satorra, A., & Sorbom, D. (1987). The detection and correction of specification errors in structural equation models. In C.C. Clogg (Ed.), *Sociological methodology* (pp. 105-129). San Francisco, CA: Jossey-Bass.

Saris, W.E., Satorra, A., & van der Veld, W.M. (2009). Testing structural equation models for detection of misspecifications. *Structural Equation Modeling, 16*, 561-582. https://doi.org/10.1080/10705510903203433

Stark, S., Chernyshenko, O.S., Drasgow, F., & Williams, B.A. (2006). Examining assumptions about item responding in personality assessment: Should ideal point methods be considered for scale development and scoring? *The Journal of Applied Psychology*, *91*(1), 25–39. https://doi.org/10.1037/0021-9010.91.1.25

Whittaker, T.A. (2012). Using the modification index and standardized expected parameter change for model modification. *Journal of Experimental Education, 80*(1), 2644. https://doi.org/10.1080/00220973.2010.531299

Zhao, Y. (2017). Impact of IRT item misfit on score estimates and severity classifications: An examination of PROMIS depression and pain interference item banks. *Quality of Life Research*, *26*(3), 555–564. https://doi.org/10.1007/s11136-016-1467-3

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, Measure Evaluation Criteria).

Calibration sample

Establishing Unidimensionality.

Table 7: EFA Correlation matrix of the six Likert-type scorable items included in the PaLS

CORRELATION MATRIX	*	*	*	*	*	*
N = 510	*	*	*	*	*	*
*	2A	2B	2C	3A	3B	3C
2a. At least one member of my care team knows about my life goals	*	*	*	*	*	*
2b. I believe it is important at least one member of my care team talks with me about my life goals	0.52	*	*	*	*	*
2c. My treatment plan is consistent with my life goals	0.56	0.36	*	*	*	*
3a. At least one member of my care team talks with me about my life goals	0.63	0.37	0.56	*	*	*
3b. I feel comfortable discussing changes in my life goals with at least one member of my care team	0.53	0.40	0.47	0.72	*	*
3c. At least one member of my care team helps me meet my life goals	0.61	0.34	0.60	0.84	0.70	*

Criterion: $r \geq 0.90$ indicates potentially redundant item content

* Cells intentionally left blank

Table 8: Eigenvalues from EFA analysis

EIGENVALUES FOR SAMPLE CORRELATION MATRIX	*	*	*	*
N = 510	*	*	*	*
*	1	2	Ratio of eigenvalue 1 to 2	% Variance
Eigenvalue	3.8	0.80	4.7	63.0

Criterion: Item set to have unidimensional characteristics if the ratio of eigenvalue 1 to eigenvalue 2 was ≥4 and the proportion of variance accounted for by eigenvalue 1 was ≥0.40.

* Cells intentionally left blank

Table 9: Geomin rotated loadings from EFA analysis

GEOMIN ROTATED LOADINGS	*	*	*	*
N = 510	*	*	*	*
*	Factor 1	*	Factor 1	Factor 2
2a. At least one member of my care team knows about my life goals	0.73‡	*	0.85‡	0.02
2b. I believe it is important at least one member of my care team talks with me about my life goals	0.51‡	*	0.71‡	-0.13
2c. My treatment plan is consistent with my life goals	0.66‡	*	0.38‡	0.33‡
3a. At least one member of my care team talks with me about my life goals	0.91‡	*	0.01	0.91‡
3b. I feel comfortable discussing changes in my life goals with at least one member of my care team	0.77‡	*	0.09	0.71‡
3c. At least one member of my care team helps me meet my life goals	0.90‡	*	-0.03	0.94‡

Criterion: Factor loadings >0.4 indicate items appear to have at least minimal construct validity.

‡ indicates significance at 5%

* Cells intentionally left blank

Table 10: Geomin factor correlation from EFA analysis

GEOMIN FACTOR CORRELATION	*	*	*	*	*
*	F1	*	*	F1	F2
F1	1	*	F1	1	*
*	*	*	F2	0.78‡	1

‡ indicates significance at 5%

* Cells intentionally left blank

Table 11: Fit Statistics from CFA analysis

1-factor model with modinidices at 100	*	*	*	2-factor model with modinidices at 100	*	*
N = 510	*	*	*	N = 510	*	*
Test	Value	90% CI	Probability RMSEA ≤0.05	Value	90% CI	Probability RMSEA ≤0.05
RMSEA	0.14	(0.12, 0.17)	0	0.09	(0.07, 0.12)	0.004
CFI	0.98	*	*	0.99	*	*
TLI	0.97	*	*	0.99	*	*
SRMR	0.06	*	*	0.04	*	*

Criterion: RMSEA <0.10, CFI and TLI ≥0.95, SRMR <0.08

* Cells intentionally left blank

Table 12: Standardized model results from CFA analysis, 1-factor model

Standardized model results - 1-factor model	*	*	*	*
N = 510	*	*	*	*
FACTOR1 by	Estimate	S.E	Est./S.E	Two-tailed p-value
2a. At least one member of my care team knows about my life goals	0.73	0.02	32.9	0
2b. I believe it is important at least one member of my care team talks with me about my life goals	0.51	0.03	16.6	0
2c. My treatment plan is consistent with my life goals	0.66	0.02	27.1	0
3a. At least one member of my care team talks with me about my life goals	0.91	0.01	83.5	0
3b. I feel comfortable discussing changes in my life goals with at least one member of my care team	0.77	0.02	40.5	0
3c. At least one member of my care team helps me meet my life goals	0.90	0.01	75.1	0

Criterion: Poor construct validity indicated by factor loading <0.50.

* Cells intentionally left blank

Table 13: Standardized model results from CFA analysis, 2-factor model

Standardized model results - 2-factor model	*	*	*	*
N = 510	*	*	*	*
FACTOR1 by	Estimate	S.E	Est./S.E	Two-tailed p-value
2a. At least one member of my care team knows about my life goals	0.81	0.02	33.3	0
2b. I believe it is important at least one member of my care team talks with me about my life goals	0.54	0.03	17.5	0
2c. My treatment plan is consistent with my life goals	0.73	0.03	27.5	0
FACTOR2 by				
3a. At least one member of my care team talks with me about my life goals	0.92	0.01	84.4	0
3b. I feel comfortable discussing changes in my life goals with at least one member of my care team	0.78	0.02	40.7	0
3c. At least one member of my care team helps me meet my life goals	0.91	0.01	74.7	0

Criterion: Poor construct validity indicated by factor loading <0.50.

* Cells intentionally left blank

Table 14: Residual correlation matrix from the CFA analysis, 1 factor model

RESIDUAL CORRELATION MATRIX	*	*	*	*	*	*
N = 510	*	*	*	*	*	*
*	2A	2B	2C	3A	3B	3C
2a. At least one member of my care team knows about my life goals	*	*	*	*	*	*
2b. I believe it is important at least one member of my care team talks with me about my life goals	0.15	*	*	*	*	*
2c. My treatment plan is consistent with my life goals	0.08	0.03	*	*	*	*
3a. At least one member of my care team talks with me about my life goals	-0.03	-0.09	-0.04	*	*	*
3b. I feel comfortable discussing changes in my life goals with at least one member of my care team	-0.03	0.01	-0.04	0.02	*	*
3c. At least one member of my care team helps me meet my life goals	-0.05	-0.11	0.00	0.02	0.01	*

Criterion: Local item dependency indicated by residual correlation >0.20.

* Cells intentionally left blank

Table 15: Item fit statistics from CFA Analysis, 1-factor model

Item	X^2	df	Probability	Item fit
2a. At least one member of my care team knows about my life goals	84.4	53	0.0039	1.6
2b. I believe it is important at least one member of my care team talks with me about my life goals	115.5	57	0.0001	2.0
2c. My treatment plan is consistent with my life goals	72.8	54	0.04	1.3
3a. At least one member of my care team talks with me about my life goals	57.1	39	0.03	1.5
3b. I feel comfortable discussing changes in my life goals with at least one member of my care team	87.6	49	0.0006	1.8
3c. At least one member of my care team helps me meet my life goals	66.8	41	0.0067	1.6

Criterion: Good item fit indicated by item fit ≤3

* Cells intentionally left blank

Table 16: Item Bias

Total (N=517)	Count (%)
Sex	*
Female	245 (47.4)
Male	272 (52.6)
Race	*
White	364 (70.4)
Other[i]	153 (29.6)
Education	*
4-year college degree or more	298 (57.6)
Less than 4-year college degree*	219 (42.4)
Age	*
Greater or equal to median age split	258 (49.9)
Less than median age split[i]	259 (50.1)
[i]Other includes: Black/African American, Native American or Alaska Native, Asian, Pacific Islander, Do not wish to report, and More than One	*
*Less than 4-year college degree includes those that chose not to report	*
Age < median age split includes missing age	*

* Cells intentionally left blank

Table 17: Differential Item Function (DIF), Nagelkerke pseudo-R^2

*	Age	*	*	Education	*	*	Gender	*	*	Race	*	*
*	Model 1 vs. 2	Model 1 vs. 3	Model 2 vs. 3	Model 1 vs. 2	Model 1 vs. 3	Model 2 vs. 3	Model 1 vs. 2	Model 1 vs. 3	Model 2 vs. 3	Model 1 vs. 2	Model 1 vs. 3	Model 2 vs. 3
2a	0.0001	0.001	0.0009	0.0038	0.0038	0	0.0015	0.0029	0.0015	0.001	0.001	0.0001
2b	0.0042	0.0042	0	0.0148	0.0171	0.0023	0.0009	0.0016	0.0007	0.0012	0.0077	0.0066
2c	0.0034	0.0034	0	0.0009	0.0009	0	0	0.0004	0.0004	0.0008	0.002	0.0012
3a	0.002	0.002	0	0.0001	0.0005	0.0004	0.0013	0.0013	0.0001	0.0015	0.0019	0.0004
3b	0.0001	0.0007	0.0006	0.0003	0.0009	0.0006	0	0.0004	0.0004	0.003	0.0047	0.0017
3c	0.0023	0.0028	0.0005	0.0006	0.001	0.0004	0.0008	0.0008	0.0001	0.0001	0.0001	0

* Cells intentionally left blank

Reliability Analyses.

Table 18: Reliability values of the PaLS

Reliability test	Calibration Sample	Validation testing Sample
Response pattern reliability	0.91	0.91
Median t-score Standard Error	2.86	2.85
T-score standard deviation	9.47	9.69
Cronbach alpha reliability	0.84	0.85
Marginal reliability	0.90	0.91

Table 19: Item-level Cronbach's alpha if an item is deleted

N = 510	*	*	*	*
Deleted Variables	Raw Variables	*	Standardized Variables	*
*	Correlation with Total	Alpha	Correlation with Total	Alpha
2a. At least one member of my care team knows about my life goals	0.65	0.82	0.66	0.81
2b. I believe it is important at least one member of my care team talks with me about my life goals	0.40	0.86	0.40	0.86
2c. My treatment plan is consistent with my life goals	0.57	0.84	0.57	0.83
3a. At least one member of my care team talks with me about my life goals	0.77	0.80	0.75	0.80
3b. I feel comfortable discussing changes in my life goals with at least one member of my care team	0.66	0.82	0.65	0.81
3c. At least one member of my care team helps me meet my life goals	0.75	0.80	0.74	0.80

* Cells intentionally left blank

Table 20: Pearson correlation coefficients among the six Likert-type scorable PaLS items

*	2A	2B	2C	3A	3B	3C
2a. At least one member of my care team knows about my life goals	1	0.42 <.0001	0.49 <.0001	0.56 <.0001	0.47 <.0001	0.54 <.0001
2b. I believe it is important at least one member of my care team talks with me about my life goals	*	1	0.29 <.0001	0.30 <.0001	0.33 <.0001	0.28 <.0001
2c. My treatment plan is consistent with my life goals	*	*	1	0.48 <.0001	0.41 <.0001	0.52 <.0001
3a. At least one member of my care team talks with me about my life goals	*	*	*	1	0.64 <.0001	0.78 <.0001
3b. I feel comfortable discussing changes in my life goals with at least one member of my care team	*	*	*	*	1	0.62 <.0001
3c. At least one member of my care team helps me meet my life goals	*	*	*	*	*	1

* Cells intentionally left blank

Validation testing sample:

Reliability Analyses.

Table 21: Item level Cronbach's alpha if an item is deleted

N = 416	*	*	*	*
Deleted Variables	Raw Variables	*	Standardized Variables	*
*	Correlation with Total	Alpha	Correlation with Total	Alpha
2a. At least one member of my care team knows about my life goals	0.66	0.84	0.66	0.83
2b. I believe it is important at least one member of my care team talks with me about my life goals	0.40	0.88	0.40	0.87
2c. My treatment plan is consistent with my life goals	0.58	0.85	0.58	0.84
3a. At least one member of my care team talks with me about my life goals	0.79	0.81	0.78	0.80
3b. I feel comfortable discussing changes in my life goals with at least one member of my care team	0.71	0.83	0.69	0.82
3c. At least one member of my care team helps me meet my life goals	0.78	0.81	0.77	0.81

* Cells intentionally left blank

Table 22: Pearson correlation coefficients among six Likert-type scorable PaLS items

*	2A	2B	2C	3A	3B	3C
2a. At least one member of my care team knows about my life goals	1	0.39 <.0001	0.49 <.0001	0.59 <.0001	0.50 <.0001	0.57 <.0001
2b. I believe it is important at least one member of my care team talks with me about my life goals	*	1	0.28 <.0001	0.33 <.0001	0.31 <.0001	0.31 <.0001
2c. My treatment plan is consistent with my life goals	*	*	1	0.51 <.0001	0.44 <.0001	0.53 <.0001
3a. At least one member of my care team talks with me about my life goals	*	*	*	1	0.70 <.0001	0.79 <.0001
3b. I feel comfortable discussing changes in my life goals with at least one member of my care team	*	*	*	*	1	0.70 <.0001
3c. At least one member of my care team helps me meet my life goals	*	*	*	*	*	1

* Cells intentionally left blank

Table 23: Test-retest Reliability, Minimal Detectable Change, and Standard Error of Measurement of PaLS *t*-score, time points combined

*	ICC	SEM	SEM %	$DC\%_{95}$ (LDC, UDC)
PaLS *t*-score	0.80	4.2	8.4	11.7 (-10.7, 12.6)

Criterion: Minimum acceptable criteria for the intraclass correlation used to analyze test-retest reliability was set at ≥0.70 (Cohen, 1969). Standard Error of Measurement (SEM) percent less 10% indicates good (i.e., acceptable) measurement error.

* Cells intentionally left blank

Reference:

Cohen J. *Statistical power analysis for the behavioral sciences.* New York: Academic Press; 1969.

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?)

The first set of analyses was focused on the identification of a unidimensional set of items and included EFA and CFA analysis. The findings from these analyses indicated that the six Likert-type scorable PaLS items were essentially unidimensional. An examination of item bias and DIF analysis did not identify any problem items for factors investigated, and a final CFA indicated good fit statistics supporting essential unidimensionality. Following these analyses we examined reliability at both the score and item-level, as well as measurement error. Specific results are discussed below.

Calibration sample:

Establishing a Unidimensional Set of Items.

CTT:

Item-adjusted total score correlations should be ≥0.40. The six Likert-type scorable PaLS items all had item-adjusted total score correlations above 0.40.

EFA:

In EFA, correlations among items that ≥0.90 may indicate redundant item content, i.e., that items are measuring the same thing (Hu & Bentler, 1999; Kline, 2005; Lai et al., 2006, 2011). Therefore, we would like item correlations to be below 0.90. For EFA, we considered the item set to have unidimensional characteristics if the ratio of eigenvalue 1 to eigenvalue 2 was ≥4 and the proportion of variance accounted for by eigenvalue 1 was ≥0.40. For geomin rotated factor loadings, a factor loading ≥0.40 supports item construct validity.

In the calibration sample, all six Likert-type scorable PaLS items had correlations below 0.90. The ratio of eigenvalue 1 to 2 for the calibration sample life goals survey was 4.7, indicating a unidimensional model appeared appropriate. For geomin rotated factor loadings, in the 1-factor model, all factor loadings were above 0.4, indicating that all items appeared to have construct validity.

CFA:

Good fit in CFA is indicated by RMSEA <0.1, CFI ≥0.95, TLI ≥0.95, and SRMR <0.08 (Cook et al., 2009; Kline, 2005; Bentler, 1990; Hu & Bentler, 1999; Hatcher, 1994; Lai et al., 2011, 2014). Signs of a poor model and fit include factor loadings <0.50, residual correlations >0.20, and correlated error modification index values ≥100 (Cook et al., 2009; Kaplan, 1989; Luijben et al., 1988; McDonald, 1999; Reise et al., 2007; Saris et al., 1987, 2009; Whittaker, 2012). In the calibration sample, the RMSEA was 0.14, which was higher than the criterion of 0.1. A 2-factor model was explored; the 2-factor model was not deemed appropriate based on modeling results (presented above). CFI and TLI were 0.98 and 0.97, respectively which met the criteria for indicating good fit. Factor loadings (presented above) were 0.50 or higher. Residual correlations (presented above) were <0.20, which indicated no problems with local item dependency. No items were flagged for MI-based correlated errors (therefore results not shown), which further supported there being no problems with local item dependency.

IRT:

Using the IRT-based approach to calculate item fit, poor fit was identified if the misfit quotient value was >3 (Crisan et al., 2017; Drasgow et al., 1995; Stark et al., 2006; Zhao, 2017). Using the IRT-based item fit assessment, all fit values were below 3, indicating good item fit.

Item bias:

For item bias, we expected an equivalent estimation of item parameters across tested groups. We did not expect to see items flagged for differential item functioning (DIF; i.e., based on the methods described in section 2a.10). Overall there was no evidence for item bias for the factors of sex, education, age, or race, i.e., no items were flagged for DIF for any of the DIF factors investigated.

Calibration sample and Validation testing sample

Reliability of the PaLS t-scores

In the calibration and validation testing samples, response pattern reliability, Cronbach's alpha reliability, and marginal reliability all supported the internal consistency reliability of the set of six Likert-type scorable items included in the PaLS instrument (i.e., all values were ≥0.84, indicating "very good" internal consistency; Cohen, 1969). Additionally, all measures of reliability were consistent across both testing samples.

Validation testing sample

T-score Level Reliability and Measurement Error

Minimum acceptable criteria for the intraclass correlation used to analyze test-retest reliability was set at ≥0.70 (Cohen, 1969). Standard Error of Measurement (SEM) percent <10% indicates good (i.e., acceptable) measurement error (Flansbjer et al., 2005).

In the validation testing sample, the test-retest reliability was very good, with an ICC = 0.80. The SEM percent was 8.4%, indicating low-level measurement error. Minimal detectable change was 11.7%, with a 95% confidence of -10.7 to 12.6.

References:

Bentler, P.M. (1990). *Comparative Fit Indexes in Structural Models. Psychological Bulletin, 107*(2), 238-246.

Cohen J. (1969). *Statistical power analysis for the behavioral sciences.* New York: Academic Press.

Cook, K. F., Kallen, M. A., & Amtmann, D. (2009). Having a fit: Impact of number of items and distribution of data on traditional criteria for assessing IRT's unidimensionality assumption. *Quality of Life Research, 18*(4), 447-460.

Crişan, D.R., Tendeiro, J.N., & Meijer, R.R. (2017). Investigating the practical consequences of model misfit in unidimensional IRT models. *Applied Psychological Measurement*, *41*(6), 439–455. https://doi.org/10.1177/0146621617695522

Drasgow, F., Levine, M.V., Tsien, S., Williams, B. & et al. (1995). Fitting polytomous item response theory models to multiple-choice tests. *Applied Psychological Measurement, 19,* 143-165. Doi:10.1177/014662169501900203

Flansbjer, U.B., Holmbäck, A.M., Downham, D., Patten, C., & Lexell, J. (2005). Reliability of gait performance tests in men and women with hemiparesis after stroke. *Journal of rehabilitation medicine*, *37*(2), 75–82. https://doi.org/10.1080/16501970410017215

Hatcher, L. (1994). *A step-by-step approach to using SAS for factor analysis and structural equation modeling*. Cary, NC: SAS Institute, Inc.

Hu, L.T., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling-a Multidisciplinary Journal, 6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Kaplan, D. (1989). Model modification in covariance structure analysis: Application of the expected parameter change statistic. *Multivariate behavioral research*, *24*(3), 285–305. https://doi.org/10.1207/s15327906mbr2403_2

Kline, R.B. (2005). *Principles and practice of structural equation modeling, Second Edition*. New York: Guilford Press.

Lai, J.S., Cella, D., Choi, S., Junghaenel, D.U., Christodoulou, C., Gershon, R., & Stone, A. (2011). How item banks and their application can influence measurement practice in rehabilitation medicine: a PROMIS fatigue item bank example. *Archives of Physical Medicine and Rehabilitation*, *92*(10 Suppl), S20–S27. https://doi.org/10.1016/j.apmr.2010.08.033

Lai, J.S., Crane, P.K., & Cella, D. (2006). Factor analysis techniques for assessing sufficient unidimensionality of cancer related fatigue. *Qual Life Res, 15*(7), 1179-1190.

Lai, J.S., Zelko, F., Butt, Z., Cella, D., Kieran, M. W., Krull, K. R., Magasi, S., & Goldman, S. (2011). Parent-perceived child cognitive function: results from a sample drawn from the US general population. *Child's Nervous System, 27*(2), 285-293.

Lai, J. S., Zelko, F., Krull, K., Cella, D., Nowinski, C., Manley, P., & Goldman, S. (2014). Parent-reported cognition of children with cancer and its potential clinical usefulness. *Quality of Life Research, 23*(4), 1049-1058.

Luijben, T. C., & Boomsma, A. (1988). Statistical guidance for model modification in covariance structure analysis. *Compstat*, 335-340.

McDonald, R.P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Reise, S.P., Morizot, J., & Hays, R.D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*, *16*(Suppl 1), 19–31. https://doi.org/10.1007/s11136-007-9183-7

Saris, W.E., Satorra, A., & Sorbom, D. (1987). The detection and correction of specification errors in structural equation models. In C. C. Clogg (Ed.), *Sociological methodology* (pp.

105-129). San Francisco, CA: Jossey-Bass.

Saris, W.E., Satorra, A., & van der Veld, W.M. (2009). Testing structural equation models for detection of misspecifications. *Structural Equation Modeling, 16*, 561-582. https://doi.org/10.1080/10705510903203433

Stark, S., Chernyshenko, O.S., Drasgow, F., & Williams, B.A. (2006). Examining assumptions about item responding in personality assessment: should ideal point methods be considered for scale development and scoring?. *The Journal of applied psychology*, *91*(1), 25–39. https://doi.org/10.1037/0021-9010.91.1.25

Whittaker, T.A. (2012). Using the modification index and standardized expected parameter change for model modification. *Journal of Experimental Education, 80*(1), 26–44. https://doi.org/10.1080/00220973.2010.531299

Zhao, Y. (2017). Impact of IRT item misfit on score estimates and severity classifications: An examination of PROMIS depression and pain interference item banks. *Quality of Life Research, 26*(3), 555–564. https://doi.org/10.1007/s11136-016-1467-3

Scientific Acceptability: Validity - Testing (2b.01 - 2b.04)

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

☒ Patient or Encounter-Level (data element validity must address ALL critical data elements)

☐ Accountable Entity Level (e.g., hospitals, clinicians)

☐ Empirical validity testing of the measure score

☐ 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)

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.

We performed several series of tests to assess different aspects of validity for the PaLS *t*-scores in the calibration and validation testing samples for participants who completed all six Likert-type scorable PaLS items. The methods and samples used for each set of validity tests are described below.

As noted in sections 2b.08-10), <2% of the samples had missing data. Given these low rates for missing scorable items, as well as the fact that those with incomplete versus complete data for scorable items did not differ on descriptive characteristics, the generalizability of our results would not be affected by the decision to include or exclude individuals with incomplete data. We elected to focus our analyses on those with complete data.

Calibration sample and Validation testing sample

Known-groups validity. Known-groups validity was used to analyze whether there were differences in PaLS *t*-score for groups expected to have different mean responses based on their demographic or clinical characteristics. In the calibration and validation testing samples, we looked for mean score differences based on dialysis modality. For modality, a one-sided upper-tail *t*-test was computed for the *t*-score, stratified by variable characteristic of interest. The *t*-test was limited to participants that answered all six Likert-type scorable PaLS items. Pooled *t* statistics were used for stratifications with equal variance based on the *F*-test.

In the validation sample we looked for PaLS mean *t*-score differences using "high" versus "low" health-related quality of life (HRQOL) for several PROMIS domain scores (i.e., Global Physical Health, Global Mental Health, Meaning and Purpose, Ability to Participate, Depression, and Self Efficacy). Mean *t*-tests were used to compare patients with low vs. high PROMIS measure-specific HRQOL status. The Folded *F*-test was used to determine whether variances were

equal between the two groups; the pooled degrees of freedom method was used when variances were equal, and Satterthwaite degrees of freedom method was used when variances were unequal. We hypothesized that participants with poor mean HRQOL scores (i.e., >0.5 SD from the mean in the "worse health" direction) would report greater dissatisfaction with their life goals discussion (i.e., have poorer mean PaLS scores) than participants with good HRQOL scores (i.e., >0.5 SD from the mean in the "better health" direction).

Base Rates for Patients with Dissatisfaction with Life Goals discussions. Dissatisfaction with life goals discussions (i.e., participants whose PaLS *t*-score was ≤40) were evaluated to determine if participants that have poor HRQOL were at increased risk for dissatisfaction with life goals discussions compared to those with good HRQOL. We expected that, based on the mean, 16% of the US general population would have poor scores on the PaLS. We anticipated this rate would be higher for those with poor versus good HRQOL; rates >16% for those with poor HRQOL would indicate greater dissatisfaction than expected.

Effect Sizes. Cohen's d effect sizes for the patient-level PaLS *t*-score was computed for the high versus low HRQOL groups for each of the PROMIS domain measures using the standard equation:

$$Cohen's\ d = \frac{mean_{group\ 1} - mean_{group\ 2}}{standard\ deviation_{pooled}}$$

Values of *d* between 0.20 and 0.49 were considered "small", values between 0.50 and 0.79 were considered "moderate," and values ≥0.80 were considered "large" (Cohen, 1988).

Floor and ceiling effects. Floor and ceiling effects assessments were used to determine if a high proportion of participants was responding "strongly disagree" or "strongly agree" to all six Likert-type scorable PaLS items. High floor and ceiling effects can be a threat to validity. The floor and ceiling effects were calculated by creating a measure score for each participant based on the numeric value of how the participant responded to the item. The number of participants with the lowest possible score of 6 was divided by the total number of participants to complete all six Likert-type scorable PaLS items for the floor effect. The number of participants with the highest possible score of 30 was divided by the total number of participants to complete all six Likert-type scorable PaLS items for the ceiling effect. N, mean, and standard deviation descriptive statistics were obtained for participants answering all six Likert-type scorable PaLS items. *A priori* criterion for acceptable floor and ceiling effects was specified as ≤20% (Andresen, 2000; Cramer & Howitt, 2001).

Validation testing sample

Convergent and discriminant validity. We investigated convergent and discriminant validity for the patient-level PaLS *t-score* using Pearson correlations. Convergent validity indicates whether the PaLS *t-score* is moderately or highly correlated with similar measured concepts. Discriminant validity is used to test that two measures have low or no correlation and are measuring unrelated concepts.

The PaLS *t-score* was compared to the PROMIS domain measure scores administered in the validation testing sample. Convergent validity would be supported by observing "moderate" to "high" correlations between the PaLS *t-score* and PROMIS measure scores. Discriminant

validity would be supported by “low” correlations between the PaLS *t-score* and PROMIS measure scores. “Low”, “moderate”, and “high” were defined as: “low” = $r \leq 0.35$, “moderate” = $r \geq 0.36 – 0.67$, “high” = *r* between 0.68 and 0.89 (Campbell & Fiske, 1959). We expected evidence of convergent validity for the PaLS *t*-scores’ association with PROMIS Meaning and Purpose.

Responsiveness. Responsiveness was tested using a modified version of the Life Events survey as well as the 3- and 6-month assessments of the PaLS. The Life Event survey (Holmes & Rahe, 1969) is a self-report survey where participants select from a list of personal, relational, health, financial and social related events experienced within the last 3 months. In order to tailor this to the ESRD chronic dialysis population for our testing, the existing Life Events survey was modified to remove events not applicable and add several specific ESRD and health-related events including switched to a different dialysis modality, switched vascular access used, were offered a kidney, switched dialysis facilities, switched to a different kidney doctor, had a change in care team, were hospitalized for any reason, had a health event requiring immediate care, death (for any reason) of someone you known that is on dialysis, COVID-19 infection or death in family member or close friend. These modifications were guided by clinical nephrologist input. Response options for all life events included “Have not experienced in the past 3 months”, extremely negative impact on life (-3) to extremely positive impact on life (+3). We anticipated a relationship between magnitude of life events experienced (agnostic to positive or negative) and changes in a participant’s PaLS *t-score*.

The absolute difference in PaLS *t-score* was calculated between baseline and 3 months as well as baseline and 6 months. Baseline to 3 months results are presented below; baseline to 6 months were explored but not found to be statistically significant (data not reported). The ***number of life events*** a participant experienced at each time point was summed to explore differences in *t-score* based on the number of life events a participant experienced. We hypothesized that the more life events a participant had at a given time point, the more likely there would also be a significant change in *t-score* for these associated time points.

We compared participants with between zero life events and four life events to participants with five or more life events. We also calculated the median split of life events, which was found to be three life events, comparing participants with ≤3 life events vs. participants with >3 life events.

Finally, the ***impact of life events*** was explored using the rating each participant gave to selected life events. The absolute sum of the impact of life events was calculated and the median split of impact was found to be seven. Participants that had life events with an impact ≤7 were compared to participants with an impact >7. Participants that did not experience any life events at each time point were excluded from the impact analysis.

References:

Andresen E.M. (2000). Criteria for assessing the tools of disability outcomes research. *Archives of Physical Medicine and rehabilitation*, *81*(12 Suppl 2), S15–S20. https://doi.org/10.1053/apmr.2000.20619

Campbell, D.T., & Fiske, D.W. (1959). Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix. *Psychological Bulletin, 56*, 81-105.

Cramer, D., & Howitt, D.L. (2004). *The Sage dictionary of statistics.* Thousand Oaks, CA:

Sage.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences.* New York, New York: Lawrence

Erlbaum Associates

Holmes, T.H., & Rahe, R.H. (1967). The Social Readjustment Rating Scale. *Journal of Psychosomatic Research*, *11*(2), 213–218. https://doi.org/10.1016/0022-3999(67)90010-4

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

Examples may include correlations or t-test results.

Known-groups Validity

Table 24: Known-groups validity for dialysis modality

*	PaLS Scores	*	*	PaLS Scores	*	*	*	*	*
*	HHD; PD	*	*	ICHD	*	*	*	*	*
Sample	N	Mean (SD) PaLS *t*-score	% of participants with poor PaLS *t*-score	N	Mean (SD) PaLS *t*-score	% of participants with poor PaLS *t*-score	t	p-value	Cohen's d
Calibration	125	53.3 (8.9)	8.8	333	48.8 (9.1)	18.3	4.7	<.0001	0.50
Validation	106	50.7 (9.9)	15.1	225	49.0 (9.3)	16.9	1.5	0.07	0.18

ICHD = in-center hemodialysis; HHD = home hemodialysis; PD = peritoneal dialysis.

* Cells intentionally left blank

Table 25: Known-groups validity for PROMIS Measures

*	PaLS Scores	*	*	PaLS Scores	*	*	*	*	*
*	Low HRQOL[a]	*	*	High HRQOL[b]	*	*	*	*	*
PROMIS Measure	N	Mean (SD) PaLS *t*-score	% of participants with poor PaLS *t*-score	N	Mean (SD) PaLS *t*-score	% of participants with poor PaLS *t*-score	t	p-value	Cohen's d
Global Physical Health	282	49.3 (9.3)	16.3	25	56.5 (11.5)	16.0	3.6	0.0002	0.76
Global Mental Health	217	48.2 (9.1)	19.4	58	54.6 (11.7)	17.2	3.9	0.0001	0.67
Meaning and Purpose	152	45.7 (7.4)	24.3	120	55.2 (10.4)	10.8	8.7	<0.0001	1.1
Ability to Participate	256	48.7 (9.2)	18.4	31	56.3 (11.6)	16.1	4.2	<0.0001	0.80
Depression‡	197	47.7 (9.1)	20.3	97	52.9 (10.2)	11.3	-4.5	<0.0001	-0.56
Self efficacy	172	47.4 (8.6)	19.8	67	54.9 (11.8)	16.4	4.7	<0.0001	0.78

Criterion: [a] for each PROMIS domain, Low HRQOL reflects participants that are 0.5 SD away from the mean in the "worse health" direction; [b] for each PROMIS domain, High HRQOL reflects participants that are 0.5 SD away from the mean in the "better health" direction; ‡Depression negatively worded concept. All other PROMIS measures are positively worded concepts

* Cells intentionally left blank

Floor and Ceiling Effects

Calibration sample and Validation testing sample

Table 26: Floor and ceiling effects for measure score

Sample	N	Floor effect (%) (a priori <20%)	Ceiling effect (%) (a priori <20%)	Mean measure score	STD of measure score
Calibration	510	0.39	6.1	20.4	5.6
Validation	416	0.48	5.8	20.2	5.7

* Cells intentionally left blank

Convergent and discriminant validity

Validation testing sample

Table 27: Pearson correlation examining convergent validity of PaLS *t*-score

*	PaLS (a priori ≥0.36)
PROMIS Meaning and Purpose	0.46

Criterion: r<0.36 discriminant validity, r≥0.36 convergent validity

* Cells intentionally left blank

Table 28: Pearson correlation examining discriminant validity of PaLS *t*-score

PROMIS measure	PaLS (a priori <0.36)
PROMIS Ability to participate	0.27
PROMIS Depression	-0.29
PROMIS Global Health Mental Health	0.30
PROMIS Global Health Physical Health	0.22
PROMIS Self Efficacy	0.34

Responsiveness

Validation testing sample

Table 29: 3-month responsiveness of the absolute difference in PaLS *t*-score relative to different groupings of self-reported number of life events

*	N	Mean (std)	N	Mean (std)	t	p-value
*	BETWEEN ZERO AND FOUR LIFE EVENTS	*	FIVE OR MORE LIFE EVENTS	*	*	*
PaLS	111	4.8 (4.6)	69	6.6 (5.8)	-2.3	0.03
*	BETWEEN ZERO AND THREE LIFE EVENTS	*	≥ THREE LIFE EVENTS	*	*	*
PaLS	96	4.4 (4.2)	84	6.7 (5.8)	-2.9	0.004
*	≤ IMPACT OF SEVEN	*	>IMPACT OF SEVEN	*	*	*
PaLS	85	4.4 (4.8)	73	6.9 (5.7)	-3.1	0.003

* Cells intentionally left blank

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?)

Results from validity testing met our respective criteria for known-groups validity, floor and ceiling effects, convergent and discriminant validity, and responsiveness. Discussion of each set of validity results for each respective testing sample (calibration or validation testing) follows below.

Known-groups validation

Modality. Known-groups validity was investigated in both the calibration sample and the validation testing sample. For the calibration sample, as expected, patients on a home dialysis modality report greater life goals satisfaction than those on in-center dialysis, supporting known-groups validity for this clinical factor. While we saw a similar trend in the validation testing sample, this finding did not meet conventional levels of significance (i.e., p=0.07). The absence of a statistically significant difference may possibly be attributed to a slightly smaller sample size in the validation testing sample. However, the effect size was 0.50 in the calibration sample, which is considered moderate and 0.18 in the validation testing sample, which is considered negligible.

HRQOL. Known-groups validity testing for these analyses was conducted using the validation testing sample only. In all cases results were in accordance with proposed hypotheses. As expected, we found that participants with poor HRQOL reported significantly greater dissatisfaction with life goals discussions than participants with good HRQOL. Specifically, those with poor global physical health, poor global mental health, poor meaning and purpose, poor ability to participate in social roles and activities, higher self-reported depression and poorer self-efficacy reported significantly more dissatisfaction with patient life goals discussions than those with good global physical health, good global mental health, good meaning and purpose, good ability to participate in social roles and activities, lower self-reported depression and good self-efficacy, respectively.

Baseline Rates for Patients with Dissatisfaction with Life Goals discussions. As expected, baseline rates of dissatisfaction with life goals discussions were consistently higher for those on in-center hemodialysis versus those on a home dialysis modality. Baseline rates were also higher for those with poor versus good HRQOL (for all measured PROMIS domains). In addition, rates of dissatisfaction consistently met or exceeded what was expected (16%), both for those on in-center hemodialysis and for those individuals with poor HRQOL.

Effect Sizes. For Cohen's d effect size, we aimed to have moderate or high effect sizes for the different groupings that we examined. These groupings included modality and the different HRQOL domains (as measured by the PROMIS measures). As expected, effect sizes were generally moderate to large; the largest effect size was seen between the PaLS groups for meaning and purpose. The sole exception was the negligible effect size that was seen for the validation testing sample for dialysis modality.

Floor and ceiling effects

We found no evidence of floor or ceiling effects in either results from our calibration sample or the validation testing sample. In the calibration and validation testing samples, floor and ceiling effects were both below 20% (floor effects 0.39% and ceiling effects 6.1% in the calibration sample; floor effects 0.48% and ceiling effects 5.8% in the validation testing sample). This indicates that only a small percentage of participants selected the lowest (worst) or highest (best) responses about their satisfaction with their care team's discussions about life goals, and that most participants had varied responses across the six Likert-type scorable PaLS items.

Convergent and discriminant validity

We expected convergent validity would be supported for the association between patient-level PaLS *t*-score and PROMIS Meaning and Purpose scores in our validation testing sample. Note that convergent validity was considered supported if the between-score correlation was ≥0.36. Discriminant validity was supported if the correlation was <0.36. PROMIS Meaning and Purpose scores and patient-level PaLS *t*-score had a correlation of 0.46, indicating evidence of convergent validity, as hypothesized. All other correlations between PROMIS measure scores and PaLS scores were <0.36, therefore supporting our expectation of providing evidence of discriminant validity.

Responsiveness

3- and 6-month responsiveness was explored in our validation testing sample. We report on 3-month responsiveness below; baseline to 6-months responsiveness was explored but not found to be statistically significant and thus is not reported. As we hypothesized, the number of life

events reported by participants yielded greater absolute PaLS *t*-score changes compared to those below each specified threshold of life events. This indicates that participants reported greater dissatisfaction with their care team's discussion about life goals within 3 months of experiencing a life event. We assume that life events, particularly a culmination of those over a relatively short period of time, may result in individuals re-evaluating their life goals, including whether their treatment plan needs to be modified. These time periods then present opportunities to revisit these discussions with the dialysis care team.

Participants with five or more life events had a greater absolute change in *t*-score compared to participants with between zero and four life events at 3 months (Table 29, p=0.03). Additionally, participants that had greater than the median split of life events (i.e., three life events) had a greater absolute change in the PaLS *t*-score compared to participants with ≤ the median split of life events at 3 months (Table 29, p=0.004).

When looking at the impact of life events, participants that experienced more than the median split of number of impactful life events (i.e., seven) had a greater change in PaLS *t*-score compared to participants with fewer impactful life events (Table 29, p=0.003).

Scientific Acceptability: Validity - Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data) (2b.05 - 2b.14)

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.

Calibration and Validation testing sample

Clinically meaningful. We investigated the change in patient-level *t*-scores between baseline, 3-month, and 6-month time points for participants included in the validation testing sample. Time points included 1) validation testing baseline to validation testing 3-month time point, and 2) validation testing baseline to validation testing 6-month time point. Participants were not required to complete 3-month or 6-month follow ups.

Additionally, we investigated the change in patient-level *t*-score for participants that were included in both the calibration sample and the validation testing sample. Time points included 1) calibration testing to validation testing baseline, 2) validation testing baseline to validation testing 3-month time point, and 3) validation testing baseline to validation testing 6-month time point.

Meaningful differences among patient-level *t*-score was used because data were not available to support facility-level *t*-score calculation at this stage. We expected to see variability in a patient-level *t*-score over time, as life goals may change for patients at different times. Differences in a patient-level *t*-score ≥5 (i.e., ≥0.5 SDs) was considered clinically meaningful (Heaton et al., 2004).

Validation testing sample

Statistical significance. The Guyatt's Responsiveness Statistic (RS) and Standardized Response Mean (SRM) effect sizes were calculated to examine the responsiveness of the PaLS measure scores. RS was calculated by dividing the mean change of PaLS *t*-scores for each PROMIS change group by the standard deviation of change in PaLS *t*-score in the "no change" group (Guyatt et al., 1987). SRMs were calculated by dividing the mean change of PaLS *t*-score for each group by the standard deviation of change of PaLS *t*-score for that group. RS and SRM were calculated relative to PROMIS Global Physical Health, Global Mental Health, and Meaning and Purpose. "No change" in PROMIS *t*-score was defined as a change <|5| *t*-score points (i.e., <0.5 SDs) between baseline and follow up. "Change" in PROMIS *t*-score is defined as a change ≥|5| *t*-score points (≥0.5 SDs). Effect sizes between 0.00 and |0.19| were considered "negligible", |0.20| to |0.49| were small, |0.50| to |0.79| were moderate, and ≥|0.80| were large.

For participants with a "change" in PROMIS *t*-score, we predicted small RS/SRMs. For participants with "no change" in PROMIS *t*-score we predicted "negligible" RS/SRMs.

General linear models (GLMs) were used to examine change over time (from baseline to 3-month and baseline to 6-month time points) for PaLS *t*-score relative to PROMIS Global Physical Health, PROMIS Global Mental Health, and PROMIS Meaning and Purpose scores (Cohen, 1992; Kopjar, 1996). Each model included predictors for "change" in PROMIS *t*-score. Least-square means and standard errors were calculated for each change group to determine whether change over time was significantly different from zero.

Responsiveness was supported by significant change in PaLS *t*-score relative to change in PROMIS measure *t*-score.

References:

Cohen, J.A. (1992). Power primer. *Psychol Bull, 112*(1), 155–159.

Guyatt, G., Walter, S., & Norman, G. (1987). Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis, 40*(2), 171–178.

Heaton, R.K., Miller, S.W., Taylor, J.T., & Grant, I. (2004). *Revised comprehensive norms for an expanded Halstead-Reitan Battery: Demographically adjusted neuropsychological norms for African American and Caucasian adults*. Lutz, FL: Psychological Assessment Resources, Inc.

Kopjar, B. (1996). The SF-36 health survey: a valid measure of changes in health status after injury. *Inj Prev,* 2, 135–139.

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.

Calibration and Validation testing sample

Table 30: Difference in PaLS *t*-score between survey time points, calibration testing and validation testing baseline, 3-month, and 6-month time points

*	No change	*	Change >0 and <2	*	Change ≥2 and <5	*	Change ≥5	*
Validation testing	N	%	N	%	N	%	N	%
Baseline to 3-month (N=183)	9	4.9	33	18.0	62	33.9	79	43.2
Baseline to 6-month (N=167)	8	4.8	35	21.0	54	32.3	70	41.9
Calibration testing and Validation testing sample	*	*	*	*	*	*	*	*
Calibration testing to Baseline (N=95)	5	5.3	31	32.6	25	26.3	34	35.8
Baseline to 3-month (N=57)	3	5.3	9	15.8	26	45.6	19	33.3
Baseline to 6-month (N=45)	4	8.9	12	26.7	14	31.1	15	33.3

* Cells intentionally left blank

Validation testing sample

Table 31: Guyatt's Responsiveness statistic for changes in PaLS *t*-score

Validation testing	Baseline to 3-month	*	*	*	*	*	Baseline to 6-month	*	*	*	*	*
Global Physical Health	No change	*	*	Change	*	*	No change	*	*	Change	*	*
*	N	RS	SRM	N	RS	SRM	N	RS	SRM	N	RS	SRM
PaLS	111	0.25	0.25	66	-0.002	-0.002	106	0.18	0.18	56	0.006	0.004
*	*	*	*	*	*	*	*	*	*	*	*	*
Global Mental Health	No change	*	*	Change	*	*	No change	*	*	Change	*	*
	N	RS	SRM	N	RS	SRM	N	RS	SRM	N	RS	SRM
PaLS	109	0.14	0.14	68	0.17	0.16	93	0.16	0.16	69	0.03	0.03
*	*	*	*	*	*	*	*	*	*	*	*	*
Meaning and Purpose	No change	*	*	Change	*	*	No change	*	*	Change	*	*
*	N	RS	SRM	N	RS	SRM	N	RS	SRM	N	RS	SRM
PaLS	114	0.26	0.26	62	0.07	0.04	86	-0.01	-0.01	74	0.28	0.25

* Cells intentionally left blank

Table 32: Responsiveness relative to change in PROMIS measure score

Validation testing	Baseline to 3-month	*	*	*	*	*	Baseline to 6-month	*	*	*	*	*
Global Physical Health	No change	*	*	Change	*	*	No change	*	*	Change	*	*
*	Least squared mean	SE	p-value	Least squared mean	SE	p-value	Least squared mean	SE	p-value	Least squared mean	SE	p-value
PaLS	1.78	-0.71	0.01	-0.01	0.92	0.99	1.17	0.74	0.12	0.04	1.02	0.97
*	*	*	*	*	*	*	*	*	*	*	*	*
Global Mental Health	No change	*	*	Change	*	*	No change	*	*	Change	*	*
	Least squared mean	SE	p-value	Least squared mean	SE	p-value	Least squared mean	SE	p-value	Least squared mean	SE	p-value
PaLS	1.00	0.72	0.17	1.27	0.92	0.17	1.20	0.79	0.13	0.22	0.92	0.82
*	*	*	*	*	*	*	*	*	*	*	*	*
Meaning and Purpose	No change	*	*	Change	*	*	No change	*	*	Change	*	*
*	Least squared mean	SE	p-value	Least squared mean	SE	p-value	Least squared mean	SE	p-value	Least squared mean	SE	p-value
PaLS	1.58	0.70	0.03	0.43	0.95	0.65	-0.08	0.82	0.93	1.99	0.88	0.02

* Cells intentionally left blank

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?

Calibration and Validation testing sample

Clinical significance. As hypothesized, participant PaLS *t*-scores changed over time. In the validation testing, 43.2% of participants had a meaningful change in *t*-score (change in *t*-score ≥5) between baseline and 3 months. Additionally, 41.9% of participants had a meaningful change in *t*-score between baseline and 6 months.

For participants in both the calibration sample and validation testing sample, 35.8% of participants had a meaningful change in PaLS *t*-score between the time of response in the calibration sample and validation testing baseline. 33.3% of participants experienced a meaningful change in PaLS *t*-score between validation testing baseline and 3 months as well as between validation testing and 6 months.

While the differences presented above are agnostic to positive or negative changes in patient-level PaLS *t*-score, the results illustrate that a patient's life goals may change over time and should be discussed with the care team regularly. This also shows that the PaLS survey can differentiate patient responses over time.

Validation testing sample

Statistical significance. The SRMs were "negligible" for Global Physical Health from baseline to 6 months, and for Global Mental Health from baseline to 3 months and from baseline to 6 months. The SRM was small for Meaning and Purpose from baseline to 6 months.

Responsiveness was supported in the "no change" baseline to 3-month group of Global Physical Health and Meaning and Purpose. Additionally, responsiveness was supported in the "change" baseline to 6-month group of Meaning and Purpose.

While we expected change in PROMIS measure scores to be predictive of change in PaLS *t*-score, they were not. Participants may not have experienced conversations with their care team initiating changes in life goals between baseline and follow up even if they experienced a change in physical health, mental health, or their feelings of meaning and purpose indicated in the PROMIS measures.

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.

To identify the extent of missing data, we looked at the count of participants that skipped each of the six Likert-type scorable PaLS items. Because the PaLS *t*-score can be calculated with missing items, low levels of missingness do not bias our results. Thus, due to observed low levels of missingness, differences between participants that responded to all items and participants that did not respond to all items was not investigated.

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).

Calibration sample

Table 33: Count and percent of participants with missing scorable Likert-type items

Item (N = 517)	Count (%) participants missing
2a. At least one member of my care team knows about my life goals	4 (0.77)
2b. I believe it is important at least one member of my care team talks with me about my life goals	2 (0.39)
2c. My treatment plan is consistent with my life goals	2 (0.39)
3a. At least one member of my care team talks with me about my life goals	3 (0.58)
3b. I feel comfortable discussing changes in my life goals with at least one member of my care team	1 (0.19)
3c. At least one member of my care team help me meet my life goals	3 (0.58)
Total count of participants that skipped at least one life goals item	7 (1.4)

Validation testing sample

Table 34: Count and percent of participants with missing scorable Likert-type items

Item (N = 420)	Count (%) participants missing
2a. At least one member of my care team knows about my life goals	0 (0)
2b. I believe it is important at least one member of my care team talks with me about my life goals	1 (0.24)
2c. My treatment plan is consistent with my life goals	1 (0.24)
3a. At least one member of my care team talks with me about my life goals	0 (0)
3b. I feel comfortable discussing changes in my life goals with my care team	1 (0.24)
3c. At least one member of my care team helps me meet my life goals	1 (0.24)
Total count of participants that skipped at least one life goals item	4 (0.95)

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.

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.

In both the calibration testing and validation testing samples, there was a low level of missingness (i.e., <2% in the calibration sample; <1% in the validation testing sample). Note that, even if one item is missing from the six Likert-type scorable PaLS items, a *t*-score can still be calculated. Low levels of PaLS item response missingness demonstrate that performance results were not biased due to missing data.

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

☐ Yes, there is more than one set of specifications for this measure
☒ No, there is only one set of specifications for this measure

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.

N/A

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.

N/A

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.

N/A

Scientific Acceptability: Validity - Other Threats to Validity (Exclusions, Risk Adjustment) (2b.15 - 2b.32)

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

☒ N/A or no exclusions
☐ Yes, the measure uses exclusions.

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?

N/A

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.

N/A

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.

N/A

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

☐ Statistical risk model with risk factors (specify number of risk factors)
☐ Stratification by risk category (specify number of categories)
☐ Other (please specify here:)
☒ No risk adjustment or stratification

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.

N/A

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.

N/A

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

☐ Published literature
☐ Internal data analysis
☐ Other (please specify here:)

N/A

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).

N/A

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.

N/A

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.

N/A

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.

N/A

2b.27) Provide risk model discrimination statistics.

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

N/A

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

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.

N/A

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

N/A

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?

N/A

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.

N/A

Feasibility (3.01 - 3.07)

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

☒ Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score)
☒ Coded by someone other than person obtaining original information (e.g., DRG, ICD-10 codes on claims)
☐ Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)
☐ Other (Please describe)

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. ALL data elements are in defined fields in electronic health records (EHRs)

☐ ALL data elements are in defined fields in electronic claims
☐ ALL data elements are in defined fields in electronic clinical data (e.g., clinical registry, nursing home MDS, home health OASIS)
☐ ALL data elements are in defined fields in a combination of electronic sources
☐ Some data elements are in defined fields in electronic sources
☐ No data elements are in defined fields in electronic sources
☒ Patient/family reported information (may be electronic or paper)

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.

The PaLS was administered using both paper and electronic forms. The questions were identical in the two modes. The PaLS is a new survey, not yet available to providers. For general implementation, the survey can be administered in paper or electronic modes.

3.04) Describe any efforts to develop an eCQM.

N/A

3.05) Complete and attach the eCQM-Feasibility-Scorecard.xls file.

N/A

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

N/A

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.

N/A

Use (4a.01 – 4a.10)

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.

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**

☐ Public Reporting
☐ Public Health/Disease Surveillance
☐ Payment Program
☐ Regulatory and Accreditation Programs
☐ Professional Certification or Recognition Program
☐ Quality Improvement with Benchmarking (external benchmarking to multiple organizations)
☐ Quality Improvement (Internal to the specific organization)
☒ Not in use
☐ Use unknown
☐ Other (please specify here:)

4a.02) Check all planned uses.

☒ Public reporting
☐ Public Health/Disease Surveillance
☒ Payment Program
☐ Regulatory and Accreditation Program
☐ Professional Certification or Recognition Program
☐ Quality Improvement with Benchmarking (external benchmarking to multiple organizations)
☐ Quality Improvement (internal to the specific organization)
☐ Measure Currently in Use
☐ Other (please specify here:)

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?

The measure is undergoing initial endorsement review.

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.

CMS will determine if/when to use this measure in a public reporting/payment program. Potential applications for the measure include the ESRD Quality Incentive Program (ESRD QIP) or the Dialysis Facility Care Compare program on Medicare.gov.

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.

Results have not been disseminated to those being measured as part of the development process.

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.

Results have not been disseminated to those being measured as part of the development process.

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

N/A. The measure is not yet implemented, and results have not been disseminated.

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

N/A. The measure is not yet implemented, and results have not been disseminated. Results have not been disseminated to those being measured as part of the development process.

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

N/A. The measure is not yet implemented, and results have not been disseminated.

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.

N/A. The measure is not yet implemented, and results have not been disseminated.

Usability (4b.01 - 4b.03)

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.

The measure is not yet implemented in a public reporting program, so improvement could not be evaluated. CMS currently anticipates future implementation of this measure. Once implemented, performance on the measure can be evaluated to determine if the measure has supported and detected quality improvement in patient life goals discussions among the target population.

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

None, as the measure is not yet implemented.

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

None, as the measure is not yet implemented.

Related and Competing (5.01 - 5.06)

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 endorsed. Please review and update questions 5.01, 5.02, and 5.03 accordingly.

5.01) Search and select all endorsed related measures (conceptually, either same measure focus or target population) by going to the PQM website.

(Can search and select measures.)

No related measures.

5.02) Search and select all endorsed competing measures (conceptually, the measures have both the same measure focus or target population) by going to the PQM website.

(Can search and select measures.)

No competing measures.

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

N/A

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

☐ Yes
☐ No

N/A

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

N/A

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.

N/A

Additional (1 - 9)

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.

☐ Available in attached file
☐ No appendix
☒ Available at measure-specific web page URL identified in sp.09

Attached PaLS Flowchart

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

Describe the members' role in measure development.

These individuals participated in the Patient Report Outcomes Technical Expert Panel that provided the conceptual framework for the PaLS measure developed by UM-KECC:

Jennifer Flythe, MD, MPH

Assistant Professor and Research Fellow, University of North Carolina at Chapel Hill

Michelle M. Richardson, PharmD.

Director of Dialysis Outcomes Programs, Director of Communications, and Assistant Professor of Medicine, Tufts Medical Center

Director, Outcomes Monitoring Program, Dialysis Clinic Incorporated

Kerri Cavanaugh, MD, MS

Medical Director, Vanderbilt Dialysis Clinic-Campus

Associate Professor of Medicine, Division of Nephrology & Hypertension, Vanderbilt University Medical Center

Paul T. Conway, BA

President, American Association of Kidney Patients

Board Member, Mid-Atlantic Renal Coalition (MARC) and Polycystic Kidney Disease Foundation (PKDF)

Lorien S. Dalrymple, MD, MPH

Vice President of Epidemiology and Research, Fresenius Medical Care North America

Volunteer Clinical Faculty, Associate Professor, Division of Nephrology University of California, Davis

Derek Forfang

Board of Directors Member and Kidney Patient Advisory Council Chair, National Forum of ESRD Networks

Kidney Advocacy Committee Member and Public Policy Committee Member, National Kidney Foundation

Patrick O. Gee

Patient Advisory Committee Chair and Subject Matter Expert, Quality Insights Mid-Atlantic Renal Coalition Network 5

Kidney Advocacy Committee Member, National Kidney Foundation

Jennifer Geiger, MSW, LSW, NSW-C

Regional Lead Social Worker, Fresenius Medical Care North America (FMCNA)

Medical Review Board Member, Quality Insights Renal Network 3

Amanda Grandinetti, MPH

Senior Specialist, Performance Measures and Analysis, American Academy of Dermatology

Kidney Action Committee Member, National Kidney Foundation

Lori Hartwell

Founder and President, Renal Support Network

Board of Directors Member, Kidney Care Partners

NQF Renal Standing Committee Member, ESRD National Quality Forum

Daniel Iniguez

Secretary of the Board of Directors, Dialysis Patient Citizens

Jacqueline Javier-Burns, RN

Director of Patient Care Services, Queens Long Island Renal Institute

Michael “Jack” Lennon, MBA

Program Manager, Division of Nephrology and Hypertension, Cincinnati Children’s Hospital Medical Center

Medical Advisory Committee Member, National Kidney Foundation of Greater Cincinnati

Klemens Meyer, MD

Director of Dialysis Services, Tufts Medical Center

Professor of Medicine, Tufts University School of Medicine

Medical Director, Dialysis Clinic, Incorporated

Sherry Rivera, DNP, APRN, ANP-C

Nurse Practitioner, New Orleans Nephrology Associates

Brigitte Schiller, MD

Chief Medical Officer and Vice President of Scientific Affairs, Satellite Healthcare, Inc.

Consulting Associate Professor, Department of Medicine, Division of Nephrology, Stanford University

Nancy L. Scott

President, Dialysis Patient Citizens Education Center

Chairperson of the Board of Directors, Henrietta Johnson Medical Center

Francesca Tentori, MD, MS

Medical Director, Outcomes Research, DaVita Clinical Research

Adjunct Instructor in Medicine, Department of Internal Medicine, Division of Nephrology, Vanderbilt University Medical Center

Medical Advisory Board Member, American Association of Kidney Patients

John E. Ware, PhD

President, Founder, and Chief Science Officer, John Ware Research Group

Professor and Division Chief, Quantitative Health Sciences, University of Massachusetts Medical School

Research Professor, Department of Medicine Tufts University School of Medicine

3) Indicate the year the measure was first released.

2022

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

04/2023

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

Annual

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

04/2024

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

N/A

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

N/A

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

N/A