Primary Care and Chronic Illness Spring 2023 Cycle: Pre-evaluation Comments

Contents

The analyses upon which this publication is based were performed under Contract Number 75FCMC23C0010, entitled, "National Consensus Development and Strategic Planning for Health Care Quality Measurement," sponsored by the Department of Health and Human Services, Centers for Medicare & Medicaid Services.

CBE 3210e HIV Viral Suppression (Health Resources and Services Administration - HIV/AIDS Bureau)

Pre-evaluation Public Comments

Public comments received for committee consideration of this measure can be found here: HIV viral suppression (3210e) | Partnership for Quality Measurement (p4qm.org)

Pre-evaluation Standing Committee Comments

1a. Evidence

- Evidence supports maintaining the measure with the newer, more stringent cutoff goals
- Evidence is solid. Pass. No concerns
- This appears to be a well-constructed measure which does not have sufficient data on use and usability. I would like to hear more specific information regarding plans for use in the future.
- No concerns
- No concerns
- No concerns

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

- Performance has clearly improved since 2017. While overall adherence is excellent, criteria for topping out of the measure are not met. Disparities in age and race are noted in current data set. Thus there is still a modest performance gap. My only question: what is the rate of patient unresponsiveness to ARV therapy and if more than trivial, how is it distributed across the affected population?
- Gap in care especially for minority populations. There is a Moderate opportunity for room for improvement.
- No Concerns
- No concerns, there is room for improvement
- Moderate gap
- No concerns

2a. Reliability

- No concerns
- No concerns. Moderate level of reliability
- No concerns
- No Concerns
- Moderate reliability, tested in part through Ryan White Clinics
- No concerns

2b. Validity

- No concerns
- Results valid. Moderate level of validity

- No concerns
- No concerns
- no concerns
- No concerns

2b2-2b6. Potential Threats to Validity

- Exclusions
  - No concerns
  - No exclusions. No concerns
  - No concerns
  - No exclusions in measure
  - no concerns
  - No exclusions, No concerns
- Risk Adjustment
  - The TEP was split re: need for risk adjustment. It seems reasonable to agree with developers for no adjustments, but subgroup analyses should continue and inform future adjustments to the measure.
  - No concerns. No risk adjustment
  - No concerns
  - No risk adjustment provided. This may provide challenges in measured performance in some clinical settings.
  - No concerns.
  - No risk adjustment, No concerns
  - Not risk adjusted or stratified.
- Meaningful Difference
  - Given the data about relationship of suppressed or absent viral titer with development of AIDS, the small differences in decile performance on measure do identify meaningful] differences about quality.
  - Wide range of performance demonstrating meaningful differences
  - No concerns
  - No concerns
  - no concerns
  - No concerns
- Comparability of Data Sources
  - No concerns
  - No concerns
  - No concerns
  - No concerns. Some EMRs did not have a data element for HIV, but this is being addressed, and is most likely apparent in a very small number of EMRs currently.
  - No concerns if EHR has structured fields.
  - No concerns
  - Only uses one set of data no concerns
- Missing Data
  - No concerns
  - No concern
  - No concerns

    - See Q11. Chart search can provide the data.
    - No concerns
    - See previous comment regarding HIV diagnosis in structured fields (denominator definition).

3. Feasibility

- No concerns
- Feasible. All accuracy issues were with data elements not required for measure calculation. No concerns
- No concerns
- No concerns
- Moderate feasibility
- Two of seven sites included in testing did not consistently capture HIV diagnoses and/or diagnosis dates in structured fields, one of the sites has plans to change workflow.
- Concerns around HIV diagnosis but they are addressing the formats

4a. Use

- Use and usability cannot be assessed in this newly refined measure.
- No concerns. Measure will be useful for accountability at the clinician level appropriately
- I would like to hear from staff on the context. Why has the measure not been implemented?
- Planning to use as a measure, but not yet implemented.
- Developer notes not in use. This presents a concern as measure initially reviewed several years ago
- Not currently in use, HRSA plans to replace MIPS CQM version with eCQM
- Yes, a plan is in place No concerns. No feedback therefore no pass--needs further discussion regarding their planned use accountability program.

4a. Usability

- There are no data on use or usability for this refined metric
- Yes. No concerns
- Insufficient evidence presented.
- This measure is at the clinician level. Feedback at the clinician level can improve knowledge and performance in measures such as this one. Additionally, measuring at the entity level can capture how a system is engaging all providers in improving care, especially when many systems are reliant on team based care.
- Not in use despite prior approval (is a maintenance measure)
- No concerns
- No data submitted to support progress on improvement--Insufficient.

5: Related and Competing Measures

- [Standing Committee feedback]

CBE #3742 ESRD Dialysis Patient Life Goals Survey (University of Michigan Kidney Epidemiology and Cost Center/ Centers for Medicare & Medicaid Services)

Pre-evaluation Public Comments

Public comments received for committee consideration of this measure can be found here: ESRD Dialysis Patient Life Goals Survey (PaLS) (3742) | Partnership for Quality Measurement (p4qm.org)

Pre-evaluation Standing Committee Comments

1a. Evidence

- There is no specific evidence that this measure relates directly to desired outcomes. However, there is general agreement in the literature that care planning correlates with patient-reported satisfaction. Since this measure is not intended for public reporting, endorsement of such a measure will provide a database for future study.
- Moderate. No concerns
- This is a process measure and per the evidence algorithm the evidence should demonstrate that the target population values the measured process or structure and finds it meaningful. The comments from the public do not support this statement. In fact, several of the comments by dialysis patients and dialysis patient groups are not in favor of the measure as constructed. The logic model states that the patient "goals of life " survey will lead to a discussion of different treatment plans (for example dialysis or transplant modality, vascular access type), which will lead to shared decision making, which will lead to alignment of treatment plan with life goals, which will lead to patient centered care. However, no data are presented to demonstrate that there is relationship between the measure, and measurement tool as constructed, and these patient outcomes. Also, there are no data presented to support that a single annual query offers an adequate or realistic evaluation of a patient's "life-goals," which are often impacted by multiple factors other than dialysis, or that nephrologists can reduce treatment times or number to accommodate a patient life goals (e.g. Independence, time with family, travel, etc.) and still simultaneously achieve adequate dialysis therapy, as is required by CMS and other payors
- Agree with Moderate however there may be new papers published around shared decision making that may expand evidence from 2020-23.

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

- No concerns
- Gap in care identified. Moderate No concerns
- Unclear if this tool demonstrates a gap on care. 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. 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.
- No concerns

2a. Reliability

- No concerns
- Agree with clarification on specs. Moderate. No other concerns
- survey fatigue may unpredictably influence the reliability of responses.
- More clarification needed for the survey selection processes.
- The exclusions of non-English speaking patients was not specifically addressed but this may significantly influence the results for some facilities"
- Reliable metric. No concerns
- No concerns

2b. Validity

- In the absence of validity testing - - cannot determine validity
- Absence of face validity/instrument tested...not performance on metric. Insufficient validity
- unclear if this tool, as constructed, is a valid measure of provider or facility quality. can the provider realistically accommodate the life goals listed in the measurement tool ?
- Empirical and Face validity testing were not conducted, testing was done on the instrument. Agree with Insufficient rating.

2b2-2b6. Potential Threats to Validity

- Exclusions
    - the exclusion of non English speaking patients is of concern
    - None. No concerns. Moderate
    - No concerns"
- Risk Adjustment
    - none
    - No risk adjustment. No concerns. Moderate
    - No concerns
    - Need to hear more from the group discussion.
- Meaningful Difference
    - While differences in the measure are demonstrated, "meaningful difference" cannot be adequately assessed.
    - Meaningful diff at 3 months. Insufficient
    - if a patient answers at least one question, that will be viewed as a satisfactory response, not clear if that is valid measure of a meaningful difference in quality between facilities
    - PROMIS measure scores show little quality change.
- Comparability of Data Sources
    - No concerns
    - No concerns
    - No concerns

- Missing Data
    - Few misses on Likert scale were noted, and number did not threaten validity
    - No concerns
    - No concerns
    - The number of patients missing the six Likert-type items may be larger in a larger real-world population.

3. Feasibility

- This measure requires that patients engage in an assessment tool. While the feasibility of this additional task to the current care process on a large scale is not clear, I believe it will indeed be feasible.
- Measure is Feasible. Moderate. No concerns.
- moderate responses can be paper or electronic, and responses may be influenced by survey fatigue,
- No concerns

4a. Use

- It is credible that results of this measure (not publicly disclosed) will have a positive effect on increasing goal planning between caregivers and patients.
- Pass. No concerns
- not in use now. Developers plan to implement
- No concerns.
- No evidence on feedback, effect of feedback or potential to incorporate feedback in future changes to the measure.
- Feedback in measure testing phase. Moderate. No concerns
- none available
- No feedback have been shared to date of report.

4a. Usability

- No concerns
- Moderate usability. No concerns
- uncertain if tool as planned reflects quality of care , Do support eCQMs . and patient cantered care and PROMS , just not sure if this tools as planned defines facility quality
- No rationale shared other than CMS will determine.
- No harms/No concerns
- patient comments raised concerns regarding how results will be used by providers, industry and payors, raising concerns regarding negative, unexpected consequence's
- Currently only available in English so those in marginalized communities may not respond to the current instrument--has it passed the reading level and language review?

CBE #3752e HIV Annual Retention in Care (Health Resources and Services Administration - HIV/AIDS Bureau)

Pre-evaluation Public Comments

Public comments received for committee consideration of this measure can be found here: HIV Annual Retention in Care (3752e) | Partnership for Quality Measurement (p4qm.org)

Pre-evaluation Standing Committee Comments

1a. Evidence

- No concerns
- Guidelines recommend monitoring for patients in care. This measure does that. Pass Moderate. No concerns
- No concerns
- reasonable evidence presented ,, but several similar measures that should be harmonized.
- No Concerns
- Evidence supports a retention measure, but the measure as structured may not fully align with evidence. For example, an individual may be seen in August (therefore qualifying for the measure) and is required to meet the 2 encounters or 1 encounter/lab visit, separated by 90 days, in an ~ 4-month remaining observation window.
- No Concerns. Sources suggest that care retention is associated with mortality.

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

- Performance test group was a select group, that developers speculated might have performed at a better level than the general physician population. This is important because current performance of this group was excellent: mean adherence was nearly 90%, with decile performance between 80%-100%. There were no statistically significant disparities. While this cannot be considered a "topped off" metric, the opportunity for improvement is relatively low.
- Yes. Moderate level gap
- No concerns
- moderate
- No Concerns
- Among clinicians with 11+ patients, median performance 91.4% with IQR of 9.5
- Overall rates around 90%. Differences were seen in population examination, but none were statistically significant (are the clinically meaningful?)

2a. Reliability

- there are several similar measures it would be good to harmonize before approving competing measure.
- No Concerns
- No concerns

- Reliable measure. No concerns. High
- No concerns
- no concerns
- No Concerns
- No concerns

2b. Validity

- while 10% of clinicians performed significantly lower than average, and 23% higher than average, the spread of scores was relatively small. No concerns about validity testing.
- Valid measure. Moderate
- No concerns
- moderate
- No Concerns
- For face validity, 4 of 7 (57%) clinicians agreed that the measure could distinguish between good and poor quality care whereas 88% of TEP agreed measure was important and related to quality of care.
- no concerns

2b2-2b6. Potential Threats to Validity

- Exclusions
    - No exclusions
    - No exclusions. No concerns
    - No concerns
    - none
    - No Concerns
    - No exclusions
    - no concerns
- Risk Adjustment
    - No risk adjustment
    - Not required. No concerns
    - No concerns
    - none
    - Not risk adjusted,
    - No risk adjustment
    - no concerns
- Meaningful Difference
    - While the spread of performance is small, patients of physicians performing significantly lower than expected (poor patient retention) are at higher risk of poor outcome - so there are meaningful differences.
    - Variation in performance. Moderate meaningful differences. No concerns
    - No concerns
    - some of data are form Ryan while so quality comparisons. may be difficult., but meaningful difference reported.
    - Performance is high, but there is room for improvement
    - Of 48 clinicians, 5 (10.4%) performed worse than the sample average; measure

developers state that broader application of measure may identify lower performance scores (testing data reflect patients receiving care in Ryan White HIV/AIDS Program)
    - analysis indicated that some performed better and some performed worse than national average. 4 of 7 clinicians agreed that measure can distinguish between good and poor quality.
- Comparability of Data Sources\
    - No multiple sets of specs - - so no concerns
    - One set of specs. No concerns
    - Not yet addressed
    - no concerns
    - No Concerns
    - N/A
    - no concerns
- Missing Data
    - No concerns
    - No threat to validity. No concerns
    - No concerns
    - no concerns
    - No Concerns
    - Per developer, data is based on presence of encounters, VL tests, and HIV diagnosis dates
    - no concerns

3. Feasibility

- No concerns
- Feasible. No concerns. Moderate
- No concerns
- moderate
- Some EMR did not have a structured data field for diagnosis, but the diagnosis was able to be obtained thru other methods.
- 2 of 7 testing sites did not consistently capture HIV diagnoses and/or diagnosis dates in structured fields
- two sites had some issues with dates / diagnosis in structured way. How generalizable are these sites to other providers that could be measured?

4a. Use

- No concerns
- Pass
- No concerns
- not yet implemented in a program,,, plans to add to accountability program
- Purportedly, a plan is in place to have this publicly reported.
- Not in use, plan for use as a clinician-level measure in MIPS
- says will be submitted to MUC by april 2023. DId that happen? When know outcome?
- New measure, so no data on feedback. To the extent that poor performance reflects No

Show return visits, or failure to do ordered testing, this is a critical question: will feedback increase adherence to physician visits and testing? Will need to collect data.

- Yes. No concerns. Pass
- Not yet addressed
- feedback invited during development., nit in use no broader feedback
- Yes - No Concerns
- has not been done to date.

4a. Usability

- Not clear to me how results of this metric might be used effectively to increase adherence to best practice.
- Moderate Pass. No concerns
- No concerns
- No Concerns
- No concerns
- no concerns
- Would be helpful to discuss as a committee

CBE #3753 Delay in Progression of Chronic Kidney Disease (CKD) Measure (Yale New Haven Health Services Corporation – Center for Outcomes Research and Evaluation (CORE)/ Centers for Medicare & Medicaid Services)

Pre-evaluation Public Comments

Public comments received for committee consideration of this measure can be found here: Delay in Progression of Chronic Kidney Disease (CKD) Measure (3753) | Partnership for Quality Measurement (p4qm.org)

Pre-evaluation Standing Committee Comments

1a. Evidence

- The evidence is incomplete, and bears only tangentially on outcome being measured. The definition of this measure is problematic and imprecise. While I agree with the statement that slowing progression of CKD is valuable, this measure does not accurately measure slowing - see below. 1. Patients qualify for inclusion when their eGFR can be anywhere between 15 - 30 ml/min. Time from inclusion to dialysis start is directly related to initial eGFR, so there is predictably large variation dependent NOT on quality of care, but on eGFR at enrollment. Practices that enroll at higher eGFR will do better - - but time of referral to a nephrologist is not a function of the nephrologist, but of the referring physician. 2. The time of initiating dialysis is highly variable, dependent on eGFR, patient symptoms, physical exam and patient autonomous decision. Most importantly, delaying dialysis for patients with low eGFR may be HARMFUL and not beneficial... the proposed measure has no way to distinguish between its harmful effects and its beneficial effects. 3. The evidence that ACE and ARB medicines delay progression of kidney disease show only a very modest effect in RCTs, and it has not been shown in empiric trials that initiating this therapy significantly delays progression if initiated. In contrast, recent studies not cited by developers show a much more powerful effect to delay or stop progression by SGLT-2 inhibitors, GLP1 agonists and nonsteroidal RAAS inhibitors. Empiric studies examining their use in more real-life circumstances have just begun to appear - need to await their finding to confirm utility of these measures to reduce progression in real-life situations. 4. Developers quote studies showing the ineffectiveness of early-start dialysis - - this is not the same as delaying the start of dialysis. 5. Developers quote loss of executive function and cognitive function once dialysis begins as evidence of harm from dialysis. These findings are more likely the result of progressive ESKD than from the dialysis procedure - - not good evidence for this metric. 6. I agree with developers and patients that non-dialysis conservative care may better align with some patient priorities. However, patients who choose non-dialysis palliative care are a very small sub-set of patients with ESKD - - most choose dialysis or transplantation. 7. Developers discuss how improved care coordination may delay progression. While this may be true, the responsibility for such care coordination cannot (in our current care system) be attributed to the consulting nephrologist - - but perhaps better in the hands of the referring primary care physician

- Pass. No concerns
- outcome measure.: no evidence presented that 2 visits to nephrologists over measurement cycle can slow progression. of CKD. Risk adjustment is based on out patient billing codes which may not accurately reflect comorbidities (outpatient billing is not linked to listing of charted comorbidities or diagnoses). no mention of GFR or eGFR so CKD classification of patients may be inaccurate. not clear if only nephrologist classification of CKD stage counts or does any billing entry which stages CKD count , even by non-nephrologist count, ? no evidence presented to suppot that after risk adjustment all disease progression can be slowed or stopped , billing and coding may be key factors and those groups who code most effectively for co-morbidities may appear to be giving best care, when is actualality they may just be "best coders"
- The empirical data cited is from studies published in 2006 or earlier and focuses predominantly on ACEi use. Evidence includes reference to an intervention (e.g., ESA) that is not consistent with standard of care.
- no concerns
- No concerns

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

- Data were provided showing a classic "bell-shaped" distribution of performance by nephrologists. No surprise - but this is NOT evidence that those with longer times from CKD4 to dialysis were from better management, but could well have been from other factors - like very late times of referral to the nephrologist relative to the eGFR, delaying dialysis start to uremic symptoms or sign which harm patients, and factors other than clinical care aimed at reducing progression.
- Moderate documented gap in care. No concerns
- unclear if actual quality performance gap
- Amon practices with 25+ patients, Median RSR 0.993 (IQR 0.922, 1.083)
- stated that the range of RSRs indicate gap in performance between providers. Population analyses do not suggest differences between groups.
- No concerns

2a. Reliability

- Data elements are clear, but insufficient. Patients qualify if claims report filed for stage 4 CKD. Since eGFR is not specified, there is very large variability in entry time relative to risk for ESKD. The measure therefore would need adjustment for entry eGFR. Likewise, time of starting dialysis does not necessarily reflect progression of CKD - - it is determined by the patient and nephrologist based on their judgment of many factors of care. Perhaps a better end of period time would be either start of dialysis OR eGFR < 10 ml/min.
- Measure specs clear and concise. No concerns. Moderate
- data elements of CKD stage with no reference to eGFR might pose reliability issue .
- Clarify start of risk time after accounting for 2 qualifying encounters and timing of stage 4 CKD diagnosis, especially when attribution occurs in the performance year (not preceding performance year). As an example, patient has first nephrology encounter ever in January with diagnosis of Stage 4 CKD and second encounter in February with

diagnosis of Stage 5 CKD, is enrolled is ESRD/ESRD-Dialysis Medicare coverage in March - is this individual excluded from the denominator?
- indicate that some data cleaning was necessary pre analysis. Is this a realistic process?
- No concerns
- Reliable measure. Moderate. No concerns
- tested
- Among nephrologists with 25+ cases, median signal-to-noise reliability was 0.821 versus 0.696 for all providers
- what does testing suggest? only good reliability levels at practices >25?

2b. Validity

- See previous concerns - applies to face validity of the measure
- Valid testing. Moderate. No concerns
- low validity as a quality measure of nephology care
- Data element validity: clarify whether also examined what proportion of those with lab values consistent with eGFR 15 -29 (ideally multiple values with an appropriate interval) had a Stage 4 CKD diagnosis; review with measure developer findings related to enrollment in ESRD versus claims - does ESRD enrollment overestimate outcome compared to ESRD claims? 11/15 TEP members somewhat or strongly agreed measure can be used to distinguish quality of care. Separately, use of stage 4 CKD (with an eGFR range of 15-29) as qualifying event for denominator and ESRD as outcome is not the same as measuring progression using annual changes in eGFR. Last, censoring for death or transplant could be problematic: censoring for death may result in practices with higher mortality performing better on this measure and transplantation does reflect progression, so this censoring event seems to be added presumably because it is viewed as an optimal type of RRT.
- No concerns agree with moderate rating.

2b2-2b6. Potential Threats to Validity

- Exclusions
    - The exclusions are appropriate, but not necessarily related to the evidence - since the evidence to support this measure is very weak (see above).
    - Appropriate exclusions. Moderate. No concerns
    - incomplete .. outpatient cancer codes are also relevant
    - Exclusions are dependent on IP claim for metastatic and advanced cancers, unclear why restricted to IP claim ascertainment. Furthermore, unclear why other exclusions were not considered given rationale provided for exclusion of advanced cancers.
    - no concerns
    - No concerns
- Risk Adjustment
    - Risk adjustment is critically dependent on entry eGFR of each patient. Since no such adjustment has been done, Risk adjustment is inadequate.
    - Risk adjustment is reasonable and necessary. Moderate. No concerns
    - extensive risk-adjustment but based on billing codes and this need to be

reviewed and validated against chart.
    - Interestingly, the risk adjusted model includes stage 5 CKD which has the highest HR (3.538), need to confirm timing of claims for co-morbidity adjustment because if these are claims preceding at-risk time and measurement year then this suggests potential for misclassification (note: percentage is 14.56%)
    - expected events used to risk adjust. social factors not includedt (no significant)
    - No concerns
- Meaningful Difference
    - Analyses show meaningful differences in the interval between identifying patients at stage 4 CKD and start of dialysis. However, there is no evidence that the measure as presented identifies meaningful differences about quality.
    - Meaningful differences noted. Moderate, No concerns
    - unclear if differences noted reflect quality alone , other issues like coding skills
    - Distributions differ by minimum number of patients required
    - significant differences between providers. majority of clinicians believe measure can differentiate quality of care.
    - No concerns
- Comparability of Data Sources
    - Data source is clear - no multiple sets of specifications
    - One set of specs for metric. No concerns. Moderate
    - no issues
    - Only one set of specifications, no concerns about comparability
    - no concerns
    - No concerns
- Missing Data
    - No concerns
    - No concerns. Moderate
    - not an issue
    - No concerns
    - seems to be unknow / not examined
    - No concerns

3. Feasibility

- Feasibility is good - - but the data element of claims-based CKD4 reporting as the determinant of inclusion is poor or incomplete - - need to have eGFR and then adjustment for it. Also, data element of start of dialysis to end the measure is clear - - but is poor or incomplete. A better end measure would be eGFR<10 or initiation of dialysis/transplantation.
- Feasible. Moderate. No concerns
- data and charts readily available
- Uses claims and administrative data
- state that all measures are in electronic sources
- No concerns

4a. Use

- No concerns
- Pass. No concerns
- not a good QIP measure,
- Plan for use in the Kidney Care Choices model (as soon as 2024)
- not in use. plan for implementation in 2024
- No concerns
- I have seen no evidence that feedback on such a measure will have a measurable impact on practice
- Feedback provided. No concerns. Moderate
- minimal feed back now

4a. Usability

- Performance results could encourage use of SGLT-2 inhibitors, GLP1 agonists or non-steroidal RAAS inhibitors, or the less effective ACE/ARB therapies. However, such results could also encourage late start dialysis that will harm patients.
- No concerns. Moderate
- unclear if this measure, as currently designed , will help improve care
- As constructed the measure would potentially identify practices with higher mortality prior to ESRD as higher quality (see comment about censoring for death).
- No concerns
- There is a clear danger of harm. This measure as currently defined could have the unintended consequence of encouraging delay of dialysis start to the time patients experience uremic symptoms, or later, increasing morbidity and mortality.
- No harms/No concerns. Moderate
- no clear harm to patirns
- Unintended consequences possible due to construction of the measure and selected censoring events

CBE #3754 Risk Standardized Mortality Ratio for Late-Stage Chronic Kidney Disease (CKD) and End Stage Renal Disease (ESRD) (Yale New Haven Health Services Corporation – Center for Outcomes Research and Evaluation (CORE)/ Centers for Medicare & Medicaid Services)

Pre-evaluation Public Comments

Public comments received for committee consideration of this measure can be found here: Risk Standardized Mortality Ratio for Late-Stage Chronic Kidney Disease (CKD) and End Stage Renal Disease (ESRD) (3754) | Partnership for Quality Measurement (p4qm.org)

Pre-evaluation Standing Committee Comments

1a. Evidence

- The underlying presumption of this measure is that for physician practices, differences in mortality adjusted for known risk factors reflect "variation in (physician) performance." Similarly, that a "large gap in performance (mortality) can be explained by a meaningful difference in (physician) performance." Is there evidence to support that claim? What percent of the outcome does physician performance determine? Are there unmeasured factors that add significantly to the outcome - and what percent of outcome may be unaccounted? The underlying evidence does not address these critical questions. While the previously endorsed and MCS-utilized SMR for dialysis patients on the facility level of measurement remains among the NQF-endorsed measures, this expansion to CKD4-5 patients, and analysis at the individual practice level add substantial variation, likely increasing the role that unmeasured factors impact mortality. It appears that the evidence supporting a connection between "physician performance" and mortality is inadequate.
- Sufficient evidence. Pass. No concerns
- This outcome measure provides a logic model that states actions taken by a nephrologist for patients with CKD stage IV and V and ESRD, can definitively reduce their all- cause mortality. The model uses a risk- stratified mortality ratio and evaluates the mortality hazard between the "best and Worst-quality nephrologists" after the adjustments for case-mix. If it is assumed that the risk-adjustment is perfect, then the construct of the measure implies that all of the residual mortality risk is due to the actions of the nephrologist, who, per the measure, may have seen the patient as few as 2 times during the measurement period. The developers do not present any evidence to substantiate this attribution of mortality risk. Additionally, with regards to the risk adjustment model itself, for patients with CKD 4 and 5, outpatient billing records are the main data source used to generate the comorbidity-based risk adjustment models. Outpatient billing is not stratified by, or depended upon, a listing of patient comorbidities, but rather billing is based on the amount of time spent with the patient, and by the complexity of medical decision making. It does not appear that the developers evaluated whether the comorbidities present in the patient's medical record were

comprehensively represented within the billing codes submitted. This is critical since the billing codes were, in turn, used as to create the risk- stratified mortality ratio. Thus, is possible that the differences noted between the High and low preforming groups have nothing to do with their quality of care, but rather reflects their billing habits and the codes they chose. This is not a trivial issue. There are over 70,000 billing codes, and if they are to be used as the basis for risk stratification the details and specificity of their use by the providers must be known. A patient can have multiple important problems that are not reflected in the particular billing codes chosen, and thus the bill may not adequately reflect the comorbidities listed in the patient's chart or problem list. Thus, the evidence behind this measure is potentially very faulty.

- Only 2 references provided in the empirical data section. Measure developers cited a study conducted from 2005 - 2006 in the UK that included people with Stages 4 and 5 CKD, and ESRD. The study was not designed to estimate the effect of the intervention on mortality. The second article, a review, primarily focused on CV-risk reduction (not all-cause mortality) and at least one of the key studies cited that did examine all-cause mortality was only in people with type 2 DM.
- No concerns

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

- Developers have clearly demonstrated wide variation in CKD4-5 and ESKD mortality in both small and large (>25 pts) practices. However, it is not clear this is a "performance" gap. 43 risk factors were used in the model to adjust for mortality. Were these adequate, or might other unrecognized or unmeasured factors have played a significant role in mortality? Social risk factors were not included - and developers provide a careful explanation for that... yet long clinical practice suggests that social, racial and financial risk factors do have an impact on mortality. And finally - the ability to capture 43 or more risk factors for each included patient is not clear - - some may have never or rarely been hospitalized, others cared for outside the US or in places where data are absent or questionable... Before we can ascribe the "gap" in outcomes to physician performance, these questions must be answered.
- Significant Gap in care demonstrating disparities and clear opportunity for improvement. Moderate. No concerns
- uncertain if the gap noted represents actual quality of care gap ;. after accounting for differences in clinical case mix, risk of death at a practice does not depend substantially on the proportion of patients served who are Black, dual-eligible, low-SES, or urban residents. Notably, the variation in outcomes within each quintile is much greater than any variation between quintiles.
- Among practices with 25+ patients, median RSMR 0.994, IQR 0.928 - 1.068
- No concerns

2a. Reliability

- My only substantial concerns about the specifications: 1. Problems in the risk/case mix adjustment as outlined in the previous "gap" question. 2. Developer data show the model has clear predictive ability to differentiate outcomes in clinically distinct subgroups. Overall outcome shows substantially lower mortality in CKD than ESKD patients. Yet the

model does not adjust for eGFR in CKD, or weigh the percent of ESKD patients in the physician's sample vs percent of CKD. These adjustments are needed to reliably interpret the results.

- Specs clear and concise. Moderate. No concerns
- moderate may not accurately reflect entire clinical status
- Reliance on claims may result in misclassification of population (e.g., Stage 4 or 5 CKD not on dialysis)
- No concerns
- Mean reliability score was 0.62. I agree with pre-review that this represents moderate reliability
- Reliable metric. Moderate. No concerns
- data readily obtained but may not accurately reflect entire clinical status
- Median signal-to-noise reliability was 0.703 for all practices compared to 0.783 for practices with 25+ cases. Note IQR among all practices (0.430 - 0.867)

2b. Validity

- No concerns
- Valid metric. Moderate. No concerns
- Clarify how/if performance estimates account for relative number of Stages 4 CKD, Stage 5 CKD, and ESRD in a practice given nephrology practices may vary in the proportion of patients in each of these group with associated differences in mortality risk among those with CKD versus ESRD. All-cause mortality may not be an appropriate focus (as compared to cause-specific mortality). Attributing mortality in the first 90 days of ESRD to the pre-dialysis nephrology practice requires justification. Data element validity for stage 4 & 5 CKD diagnosis versus laboratory values conducted. Clarify findings with respect to ESRD enrollment compared to claims.
- No concerns

2b2-2b6. Potential Threats to Validity

- Exclusions
    - Patients with metastatic cancer and Hospice patients are excluded. however, patients who choose palliative care rather than dialysis, patients who refuse further treatment, those who commit suicide, and patients who choose to stop dialysis treatments before their natural death, should likewise be excluded.
    - Appropriate exclusions. No concerns. Moderate
    - no concerns
    - Exclusion of metastatic or advanced cancer relies on IP claims, unclear why OP claims are not considered for exclusion. Other exclusions could be considered to be appropriate given rationale for excluding advanced cancers.
    - No concerns
- Risk Adjustment
    - See above. Risk adjustment is done well. however, risk adjustment models do not presume that[they capture ALL of the factors associated with risk of death, and that all any difference between these measured effects and 100% represents "performance of the nephrologist." This is a major defect in this model - - - risk

adjustment should specify the percent of variation accounted for in the risk factors. The remainder may include unmeasured factors which include physician performance. But the remainder cannot and should not be ascribed solely to physician performance.
    - Appropriate risk adjustment. No concerns. Moderate
    - as noted above
    - Several types of cancers included in the risk-adjusted model, some of the associated risk estimates raise the question of incomplete exclusion of advanced cancers (? reliance on IP claims resulted in under ascertainment of exclusion). Can discuss whether risk-adjustment is sufficient.
    - No concerns
- Meaningful Difference
    - Given that risk adjustment cannot clearly identify the role of "physician performance" in the risk of death, meaningful differences in physician practice-level mortality may be meaningful - but should not be ascribed to physician performance solely.
    - Meaningful differences detected. No concerns. Moderate
    - uncertain if differences noted are reflective of care or if they actually reflect out patient billing training or use habits./patterns . ESRD data for co-morbidities may be problematic, as well . comorbidities only captured once , at time of first enrollment in eSRD/Medicare
    - Demonstrated difference between the bottom and top quintile
    - No concerns
- Comparability of Data Sources
    - My major concern here is the ability to capture all (or most of) the elements in the risk model for pre-dialysis patients as described above.
    - Single spec set. Moderate. No concerns
    - no issues
    - N/A
    - No concerns
- Missing Data
    - Yes, as above for risk model variables
    - No concerns. Moderate
    - no issues
    - Measure developer discussed claims data (absence treated as not having condition)
    - No concerns

3. Feasibility

- Claims data may not adequately capture needed data elements for the risk model - - unclear if any other data sources are feasible.
- Feasible. No concern. Moderate
- data readily reachable
- Relies on claims and administrative data
- No concerns

4a. Use

- Given the problem that this measured gap in causes of mortality may not necessarily represent physician performance, I don't see a plan to measure that causal gap and seek other possible contributing factors.
- Pass. No concern.
- not valid for quality of care /payment measure
- Planned for implementation in voluntary Kidney Care Choices model (as soon as 2024)
- No concerns
- This is a new measure, and thus no feedback as yet. However, it is related to the SMR measure for dialysis patients. It would be valuable to know whether feedback of those metrics have been used or have changed practice in dialysis facilities.
- Feedback provided. No concern. Moderate
- the FACILITY SMR is a very different measure , this new measure is not in use-minimal feedback from affected user groups.
- they have a plan no concerns

4a. Usability

- Performance results may induce clinicians with higher than expected mortality to examine their practice. However, uncertainty about whether their findings are related to unmeasured risk factors or their practice modes make it difficult for these practitioners to understand the meaning of these findings.
- No concerns. Moderate
- low
- See prior concerns about validity
- No concerns
- There surely is potential harm... In a world where there is currently a shortage of nephrologists in some areas of the US, practitioners potentially affected by financial threats may choose to care for only those patients from higher socio-economic populations, or those with better overall health status.
- No concern. Moderate
- no clear discussion, provided but likely of low risk
- Measure as constructed may not reflect the quality of care provided by the nephrologist

CBE #3755e STI Testing for People with HIV (Health Resources and Services Administration - HIV/AIDS Bureau)

Pre-evaluation Public Comments

Public comments received for committee consideration of this measure can be found here: STI Testing for People with HIV (3755e) | Partnership for Quality Measurement (p4qm.org)

Pre-evaluation Standing Committee Comments

1a. Evidence

- This measure calls for annual 3-STD testing in patients with HIV. Three sets of guidelines recommend testing - one recommends "routine testing," another recommends annual testing without grading the evidence. In reviewing the USPSTF guidance, the developers cite a study by Patel ("2012") reportedly showing the utility of annual STD testing. The Patel citation I could find was from 2021, examining new and repeated syphilis infection. The study assessed brief risk reduction counseling with biannual STD testing, and found that it did reduce syphilis incidence. The latter study did not examine the efficacy of annual STD testing - the focus of this measure. Thus, I found only expert opinion, but no clear higher level evidence showing the efficacy of annual testing for STDs. Thus the level of evidence is LOW.
- Guideline driven. No concerns. Moderate
- Graded Evidence for testing all hiv pts annually for each of the listed std is not included and may not be available. Testing sexually active msm annually is strongest
- No concerns
- USPSTF recommendation focuses on syphilis screening (Grade A), other guidelines no grade assigned.
- Evidence seems to suggest testing annually or even more frequently for certain individuals. This measure does not seem to have the ability to differentiate between those groups / individuals and only indicates if the tests have been done one or more times.
- No concerns

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

- While the studied population was small, the very wide variation in decile performance shows a clear gap. There were disparities by race/age, but I agree with staff reviewers that small numbers make us careful in interpreting these findings.
- Performance gap noted. Opportunity for improvement. Moderate. No concerns
- Moderate
- Disparities information provided. No concerns
- No concerns
- overall average around 55%. Differences were seen in different groups.
- Supports disparities in care with small sample size-IDU may not seek care as often as others.

2a. Reliability

- No concerns
- Specs defined. Pass. High, No concerns
- No concerns
- Definitions clear. No case-mix adjustment. Otherwise, No concerns
- Confirm numerator includes STI testing in the measurement period that precedes qualifying encounter
- no concerns
- No concerns
- Reliable measure. No concerns. High

2b. Validity

- I do have some concerns. If patients are tested months before the observation period, it is likely that no repeat testing is needed in the subsequent year - there is NO clear evidence that annual testing is needed.
- Valid measure, no concerns. Moderate
- Moderate
- No concerns
- 3/7 (42%) of clinicians agreed the measure can distinguish quality of care whereas 100% of TEP agreed measure was important and related to quality of care
- no concerns
- No concerns

2b2-2b6. Potential Threats to Validity

- Exclusions
    - No exclusions - so no concerns
    - No exclusions
    - No concerns
    - No exclusions
    - Clinicians interviewed raised concerns about the lack of exclusion of patients who are not sexually active
- Risk Adjustment
    - No risk adjustment - so no concerns
    - Not required
    - No concerns
    - No risk adjustment done
    - No risk adjustment, no concerns
    - No risk adjustment
- Meaningful Difference
    - The are clear differences in deciles of performance. Clearly there is a meaningful difference between the lowest decile (0%) and highest decile (100%). There is no evidence to distinguish among the other deciles - - are infections more common in the 3rd vs 7th decile of performance??
    - High meaningful differences
    - No concerns

    - No concerns
    - No concerns
    - less than half of clinicians believe measure can distinguish care quality
    - No concerns
- Comparability of Data Sources
    - No multiple sets of specs... so no concerns
    - One set of specs. Pass. High
    - Need to harmonize with other existing measures . The criteria shoukd all be aligned
    - No concerns
    - N/A, No concerns
    - no concerns
    - No concerns
- Missing Data
    - No concerns
    - No threats, No concerns, High
    - No concerns
    - No concerns
    - No concerns
    - no concerns
    - No concerns

3. Feasibility

- No concerns
- Feasible, no concerns. Moderate
- Moderate
- No concerns
- 2/7 sites do not consistently capture HIV diagnoses and/or diagnosis dates (1 site has a plan to change workflow)
- HIV diagnosis date data issue in two locations
- No concerns

4a. Use

- No concerns
- Pass. Moderate
- Not sure if it should be limited to specific subsets of pts rather than every pt with hiv.
- Purportedly a plan to add this as a measure for future use as an eMeasure
- Plan for use as a clinician-level measure in MIPS
- Has measure been submitted to MUC list?
- No concerns
- No evidence to answer whether feedback improves performance - but it is very likely it will
- Feedback given. Moderate pass

4a. Usability

- No evidence - but performance is likely to improve when providers know this is a performance measure
- Moderate pass
- Moderate
- No concerns, as in other HIV measures under review this cycle
- See discussion of benefits vs harms
- no concerns
- This measure will help providers focus toward testing and treating STI's in patient with HIV.
- Not determined, but No concerns
- Measure may promote screening for gonorrhea and chlamydia that is not clinically indicated
- no concerns
- Stigmatization, Privacy concerns, Resource allocation, and Compliance issues-- As of my last update in September 2021, I don't have access to information on any specific unintended consequences related to testing for sexually transmitted infections (STIs) in persons with HIV beyond that point. However, I can discuss potential unintended consequences and how the benefits of such measures may outweigh them based on what was known up until that time. Unintended Consequences: Stigmatization: Introducing mandatory or routine STI testing for persons with HIV may unintentionally perpetuate stigma around HIV and increase discrimination against those living with the virus. People with HIV might feel singled out and face further marginalization due to additional testing requirements. Privacy concerns: Implementing more extensive testing protocols might raise concerns about patient privacy. People might worry that their sensitive health information could be exposed or misused, leading to potential reluctance in seeking care and disclosing their HIV status. Resource allocation: Expanding testing protocols could potentially strain healthcare resources, including finances, staff, and equipment. This may lead to delays in receiving test results or other health services for individuals. Compliance issues: Requiring additional testing may lead to decreased compliance with healthcare guidelines. Some individuals might be hesitant to undergo frequent testing, leading to a potential decrease in overall STI detection rates and timely treatment. Benefits Outweighing the Unintended Consequences: Early detection and treatment: Routine STI testing in persons with HIV can help identify infections early, leading to timely treatment. This can prevent complications and reduce the risk of transmitting STIs to sexual partners, including other individuals with HIV. Improved public health outcomes: Identifying and treating STIs promptly not only benefits the individual but also contributes to the overall reduction of STI transmission rates in the population. It can be an essential step in controlling the spread of STIs. Comprehensive healthcare: By integrating STI testing into the regular care of persons with HIV, healthcare providers can address multiple health concerns simultaneously, leading to more comprehensive and holistic care. Partner notification and prevention: Early STI detection allows for better partner notification and targeted prevention efforts, helping to break the chain of transmission and protect sexual partners from infection.