NATIONAL QUALITY FORUM

Measure Evaluation 4.1
December 2009

This form contains the measure information submitted by stewards. Blank fields indicate no information was provided. Attachments also may have been submitted and are provided to reviewers. The subcriteria and most of the footnotes from the evaluation criteria are provided in Word comments within the form and will appear if your cursor is over the highlighted area. Hyperlinks to the evaluation criteria and ratings are provided in each section.

TAP/Workgroup (if utilized)**:** Complete all yellow highlighted areas of the form. Evaluate the extent to which each subcriterion is met. Based on your evaluation, summarize the strengths and weaknesses in each section.

Note: *If there is no TAP or workgroup, the SC also evaluates the subcriteria (yellow highlighted areas).*

Steering Committee: Complete all pink highlighted areas of the form. Review the workgroup/TAP assessment of the subcriteria, noting any areas of disagreement; then evaluate the extent to which each major criterion is met; and finally, indicate your recommendation for the endorsement. Provide the rationale for your ratings.

Evaluation ratings of the extent to which the criteria are met
C = Completely (unquestionably demonstrated to meet the criterion)
P = Partially (demonstrated to partially meet the criterion)
M = Minimally (addressed BUT demonstrated to only minimally meet the criterion)
N = Not at all (NOT addressed; OR incorrectly addressed; OR demonstrated to NOT meet the criterion)
NA = Not applicable (only an option for a few subcriteria as indicated)

(for NQF staff use) **NQF Review #:** 1426 **NQF Project:** End Stage Renal Disease
MEASURE DESCRIPTIVE INFORMATION
De.1 **Measure Title:** Assessment of Iron Stores
De.2 **Brief description of measure***:* Percentage of all adult (>= 18 years old) dialysis patients for whom serum ferritin and transferrin saturation percentage (TSAT) are measured simultaneously at least once during the three-month study period.
1.1-2 **Type of Measure:** Process De.3 **If included in a composite or paired with another measure, please identify composite or paired measure** N/A
De.4 **National Priority Partners Priority Area:** Population health De.5 **IOM Quality Domain**: Effectiveness De.6 **Consumer Care Need:** Living with illness

CONDITIONS FOR CONSIDERATION BY NQF	
Four conditions must be met before proposed measures may be considered and evaluated for suitability as voluntary consensus standards:	**NQF Staff**
A. The measure is in the public domain or an intellectual property (measure steward agreement) is signed. *Public domain only applies to governmental organizations. All non-government organizations must sign a measure steward agreement even if measures are made publicly and freely available.* A.1 **Do you attest that the measure steward holds intellectual property rights to the measure and the right to use aspects of the measure owned by another entity (e.g., risk model, code set)?** Yes A.2 **Indicate if Proprietary Measure** (*as defined in measure steward agreement*)**:** A.3 **Measure Steward Agreement:** Government entity and in the public domain - no agreement necessary A.4 **Measure Steward Agreement attached:**	**A** Y☒ N☐
B. The measure owner/steward verifies there is an identified responsible entity and process to maintain and	**B**

update the measure on a schedule that is commensurate with the rate of clinical innovation, but at least every 3 years. Yes, information provided in contact section	Y☒ N☐
C. The intended use of the measure includes both public reporting and quality improvement. ►**Purpose:** Public reporting, Internal quality improvement	C Y☒ N☐
D. The requested measure submission information is complete. Generally, measures should be fully developed and tested so that all the evaluation criteria have been addressed and information needed to evaluate the measure is provided. Measures that have not been tested are only potentially eligible for a time-limited endorsement and in that case, measure owners must verify that testing will be completed within 12 months of endorsement. D.1**Testing:** Yes, fully developed and tested D.2 **Have NQF-endorsed measures been reviewed to identify if there are similar or related measures?** Yes	D Y☐ N☒
(for NQF staff use) **Have all conditions for consideration been met?** **Staff Notes to Steward** (*if submission returned*)**:**	**Met** Y☐ N☐
Staff Notes to Reviewers (*issues or questions regarding any criteria*)**:** I	
Staff Reviewer Name(s):	

TAP/Workgroup Reviewer Name:	
Steering Committee Reviewer Name:	
1. IMPORTANCE TO MEASURE AND REPORT	
Extent to which the specific measure focus is important to making significant gains in health care quality (safety, timeliness, effectiveness, efficiency, equity, patient-centeredness) and improving health outcomes for a specific high impact aspect of healthcare where there is variation in or overall poor performance. *Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria.* (evaluation criteria) **1a. High Impact**	Eval Ratin g
(for NQF staff use) Specific NPP goal:	
1a.1 **Demonstrated High Impact Aspect of Healthcare:** Affects large numbers, Frequently performed procedure, High resource use 1a.2 1a.3 **Summary of Evidence of High Impact:** The measure focus is important because prudent use of IV iron in dialysis patients improves management of anemia; lowers the dose of erythropoietin stimulating agent (ESA) needed to maintain the hemoglobin in the target range; avoids potential harm of excess iron administration; and encourages optimum utilization of pharmacologic and laboratory resources. 1a.4 **Citations for Evidence of High Impact:** 1) Singh AK, Szczech L, Tang KL, Barnhart H, Sapp S, Wolfson M, Reddan, CHOIR Investigators: Correction of anemia with epoetin alfa in chronic kidney disease. N Engl J Med 355: 2085–2098, 2006. 2) Besarab A, Bolton WK, Browne JK, Egrie JC, Nissenson AR, Okamoto DM, Schwab SJ, Goodkin DA: The effects of normal as compared with low hematocrit values in patients with cardiac disease who are receiving hemodialysis and epoetin. N Engl J Med 339: 584–590, 1998. 3) Phrommintikul A, Haas SJ, Elsik M, Klum H: Mortality and target haemoglobin concentrations in anemia patients with chronic kidney disease treated with erythropoietin: A meta-analysis. Lancet 369: 381–388, 2007.	**1a** C☐ P☐ M☐ N☐

Comment [KP1]: 1a. The measure focus addresses:
•a specific national health goal/priority identified by NQF's National Priorities Partners; OR
•a demonstrated high impact aspect of healthcare (e.g., affects large numbers, leading cause of morbidity/mortality, high resource use (current and/or future), severity of illness, and patient/societal consequences of poor quality).

4) Kapoian T, O'Mara NB, Singh AK, Moran J, Rizkala AR, Geronemus R,Kopelman RC, Dahl NV, Coyne DW: Ferric gluconate reduces epoetin requirements in hemodialysis patients with elevated ferritin. J Am Soc Nephrol 19: 372–379, 2008.

5) Pizzi LT, Bunz TJ, Coyne DW, Goldfarb DS, Singh AK: Ferric gluconate treatment provides cost savings in patients with high ferritin and low transferrin saturation. Kidney Int 74: 1588–1595, 2008.

6) Pfeffer MA, Burdmann EA, Chen CY, et al. A Trial of Darbepoetin Alfa in Type 2 Diabetes and Chronic Kidney Disease. New England Journal of Medicine 361: 2019-2032, 2009.

7) Phrommintikul A, Haas SJ, Elsik M, Klum H. Mortality and target haemoglobin concentrations in anemia patients with chronic kidney disease treated with erythropoietin: A meta-analysis. Lancet 369: 381-388, 2007.

8) Palmer SC, Navaneethan SD, Craig JC, et al. Meta-analysis: erythropoiesis-stimulating agents in patients with chronic kidney disease. Annals of Internal Medicine 153: 23-33, 2010.

9) Suetonia C. Palmer, Ann Intern Med July 6, 2010 153:23-33; published ahead of print May 3, 2010

1b. Opportunity for Improvement

1b.1 Benefits (improvements in quality) envisioned by use of this measure: Erythropoesis-stimulating agent (ESA) and iron therapies correct anemia in dialysis patients. Routine assessment of iron stores allows for prudent use of intravenous (IV) iron which can lower the dose of ESA's needed to maintain hemoglobin (Hgb) in the target range.

1b.2 Summary of data demonstrating performance gap (variation or overall poor performance) across providers:
In the test calculation of the measure using July-September 2009 CROWNWeb data, the facility-level mean was 85% (SD=13.5%). The median, 25th, and 75th percentiles were 88%, 81% and 94%, respectively.

1b.3 Citations for data on performance gap:
CROWNWeb Phase II test data, July-September 2009.

1b.4 Summary of Data on disparities by population group:
N/A

1b.5 Citations for data on Disparities:
N/A

1b
C☐ P☐ M☐ N☐

1c. Outcome or Evidence to Support Measure Focus

1c.1 Relationship to Outcomes (*For non-outcome measures, briefly describe the relationship to desired outcome. For outcomes, describe why it is relevant to the target population*)**:** ESA and iron therapies correct anemia in dialysis patients. Routine assessment of iron stores allows for prudent use of IV iron which can lower the dose of ESA's needed to maintain Hgb in the target range.

1c.2-3. Type of Evidence: Randomized controlled trial

1c.4 Summary of Evidence (*as described in the criteria; for outcomes, summarize any evidence that healthcare services/care processes influence the outcome*)**:**
Recent clinical trials provide evidence that targeting higher Hgb levels when treating anemia in patients with chronic kidney disease (CKD) may increase the risk of adverse outcomes. The Trial to Reduce Cardiovascular Endpoints with Aranesp Therapy (TREAT) study found higher rates of stroke, thromboembolism, and cancer-related deaths in patients with CKD and diabetes who were treated to the higher Hgb target. The Correction of Hemoglobin and Outcomes in Renal Insufficiency (CHOIR) study (CKD patients) [Singh AK, 2006] and the Normal Hematocrit study (dialysis patients at high cardiovascular risk) [Besarab A, 1998] both found higher rates of death and cardiovascular complications among patients

1c
C☐ P☐ M☐ N☐

Comment [KP2]: 1b. Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating considerable variation, or overall poor performance, in the quality of care across providers and/or population groups (disparities in care).

Comment [k3]: 1 Examples of data on opportunity for improvement include, but are not limited to: prior studies, epidemiologic data, measure data from pilot testing or implementation. If data are not available, the measure focus is systematically assessed (e.g., expert panel rating) and judged to be a quality problem.

Comment [k4]: 1c. The measure focus is:
•an outcome (e.g., morbidity, mortality, function, health-related quality of life) that is relevant to, or associated with, a national health goal/priority, the condition, population, and/or care being addressed;
OR
•if an intermediate outcome, process, structure, etc., there is **evidence** that supports the specific measure focus as follows:
oIntermediate outcome – evidence that the measured intermediate outcome (e.g., blood pressure, Hba1c) leads to improved health/avoidance of harm or cost/benefit.
oProcess – evidence that the measured clinical or administrative process leads to improved health/avoidance of harm and
if the measure focus is on one step in a multi-step care process, it measures the step that has the greatest effect on improving the specified desired outcome(s).
oStructure – evidence that the measured structure supports the consistent delivery of effective processes or access that lead to improved health/avoidance of harm or cost/benefit.
oPatient experience – evidence that an association exists between the measure of patient experience of health care and the outcomes, values and preferences of individuals/ the public.
oAccess – evidence that an association exists between access to a health service and the outcomes of, or experience with, care.
oEfficiency – demonstration of an association between the measured resource use and level of performance with respect to one or more of the other five IOM aims of quality.

Comment [k5]: 4 Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multi-step process, the step with the greatest effect on the desired outcome should be selected as the focus of measurement. For example, although assessment of immunization status and recommending immunization are necessary steps, they are not sufficient to achieve the desired impact on health status – patients must be vaccinated to achieve immunity. This does not preclude consideration of measures of preventive screening interventions where there is a strong link with desired outcomes (e.g., … [1]

treated to higher Hgb targets. Two meta-analyses, which included both dialysis and non-dialysis CKD studies, also supported these findings [Phrommintikul A, 2007; KDOQI, 2006].
Although the cause of higher event rates among patients randomized to higher Hgb targets remains incompletely understood, higher ESA doses have been implicated as a possible explanation, and recent opinion in the nephrology community has coalesced around strategies to limit ESA dose when possible. To this end, alternate methods to facilitate ESA-mediated erythropoiesis, and support Hgb levels with lower ESA doses, are increasingly recommended, and the judicious use of IV iron therapy remains central to this strategy [Kapoian T, 2008; Pizzi LT, 2008; Singh AK, 2010].
At the same time, the TEP recognizes evidence limitations with respect to long-term safety of IV iron therapy. As standard practice, IV iron dosing decisions are based on clinical measures of iron stores including ferritin and transferrin saturation (TSAT) levels. The proposed CPMs leave most treatment decisions about IV iron dosing to the judgment of the practitioner, with the exception of values notably out of normal range. For example, no judgment is made about IV iron dosing to patients with ferritin in the 100 to 1200 ng/mL range or with TSAT <50%

1c.5 Rating of strength/quality of evidence (*also provide narrative description of the rating and by whom***):**
Overall, Grade B evidence. Randomized clinical trials were conducted in mainly in CKD patients not on dialysis.

1c.6 Method for rating evidence: United States Preventive Services Task Force (USPSTF)

1c.7 Summary of Controversy/Contradictory Evidence: There is no controversy over the importance of routine iron assessment in dialysis patients.

1c.8 Citations for Evidence (*other than guidelines***):** 1)Singh AK, Szczech L, Tang KL, Barnhart H, Sapp S, Wolfson M, Reddan, CHOIR Investigators: Correction of anemia with epoetin alfa in chronic kidney disease. N Engl J Med 355: 2085–2098, 2006.

2) Besarab A, Bolton WK, Browne JK, Egrie JC, Nissenson AR, Okamoto DM, Schwab SJ, Goodkin DA: The effects of normal as compared with low hematocrit values in patients with cardiac disease who are receiving hemodialysis and epoetin. N Engl J Med 339: 584–590, 1998.

3) Phrommintikul A, Haas SJ, Elsik M, Klum H: Mortality and target haemoglobin concentrations in anemia patients with chronic kidney disease treated with erythropoietin: A meta-analysis. Lancet 369: 381–388, 2007.

4) Kapoian T, O'Mara NB, Singh AK, Moran J, Rizkala AR, Geronemus R,Kopelman RC, Dahl NV, Coyne DW: Ferric gluconate reduces epoetin requirements in hemodialysis patients with elevated ferritin. J Am Soc Nephrol 19: 372–379, 2008.

5) Pizzi LT, Bunz TJ, Coyne DW, Goldfarb DS, Singh AK: Ferric gluconate treatment provides cost savings in patients with high ferritin and low transferrin saturation. Kidney Int 74: 1588–1595, 2008.

6) Pfeffer MA, Burdmann EA, Chen CY, et al. A Trial of Darbepoetin Alfa in Type 2 Diabetes and Chronic Kidney Disease. New England Journal of Medicine 361: 2019-2032, 2009.

7) Phrommintikul A, Haas SJ, Elsik M, Klum H. Mortality and target haemoglobin concentrations in anemia patients with chronic kidney disease treated with erythropoietin: A meta-analysis. Lancet 369: 381-388, 2007.

8) Palmer SC, Navaneethan SD, Craig JC, et al. Meta-analysis: erythropoiesis-stimulating agents in patients with chronic kidney disease. Annals of Internal Medicine 153: 23-33, 2010.

9) Suetonia C. Palmer, Annals of Internal Medicine 153: 23-33, 2010; published ahead of print May 3, 2010.

1c.9 Quote the Specific guideline recommendation (*including guideline number and/or page number***):**

Comment [k6]: 3 The strength of the body of evidence for the specific measure focus should be systematically assessed and rated (e.g., USPSTF grading system http://www.ahrq.gov/clinic/uspstf07/methods/benefit.htm). If the USPSTF grading system was not used, the grading system is explained including how it relates to the USPSTF grades or why it does not. However, evidence is not limited to quantitative studies and the best type of evidence depends upon the question being studied (e.g., randomized controlled trials appropriate for studying drug efficacy are not well suited for complex system changes). When qualitative studies are used, appropriate qualitative research criteria are used to judge the strength of the evidence.

3.2.1 Frequency of iron status tests Iron status tests should be performed: • Every month during initial ESA treatment • At least every 3 months during stable ESA treatment or in patients with HD-CKD not treated with an ESA 3.2.2 Interpretation of iron status tests Results of iron status tests, Hb, and ESA dose should be interpreted together to guide iron therapy 1c.10 **Clinical Practice Guideline Citation:** KDOQI; National Kidney Foundation. II. Clinical practice guidelines and clinical practice recommendations for anemia in chronic kidney disease in adults. Am J Kidney Dis 47[Suppl 3]: S16-S85, 2006. 1c.11 **National Guideline Clearinghouse or other URL:** N/A 1c.12 **Rating of strength of recommendation (***also provide narrative description of the rating and by whom***):** Kidney Disease Outcomes Quality Initiative (KDOQI) Anemia Guidelines are opinion-based. 1c.13 **Method for rating strength of recommendation** (*If different from* USPSTF system*, also describe rating and how it relates to USPSTF)***:** N/A 1c.14 **Rationale for using this guideline over others:** N/A	
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Importance to Measure and Report?*	**1**
Steering Committee: Was the threshold criterion, *Importance to Measure and Report***, met?** **Rationale:**	**1** **Y☐** **N☐**
2. SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES	
Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)	Eval Rating
2a. MEASURE SPECIFICATIONS	
S.1 **Do you have a web page where current detailed measure specifications can be obtained?** S.2 **If yes, provide web page URL:** **2a. Precisely Specified**	
2a.1 **Numerator Statement (***Brief, text description of the numerator - what is being measured about the target population, e.g. target condition, event, or outcome***):** Number of patients in the denominator for whom serum ferritin and TSAT are measured simultaneously at least once during the study period. Simultaneous measurements are those reported with the same collection date. 2a.2 **Numerator Time Window (***The time period in which cases are eligible for inclusion in the numerator***):** Rolling three-month study period. 2a.3 **Numerator Details (***All information required to collect/calculate the numerator, including all codes, logic, and definitions***):** The numerator will be determined by counting the patients in the denominator who meet the following criteria in the three-month study period: Serum Ferritin is populated, AND Iron Saturation Percentage is populated AND Serum Ferritin Collection Date is equal to Iron Saturation Percentage Collection Date.	**2a-specs** **C☐** **P☐** **M☐** **N☐**

Comment [k7]: USPSTF grading system **http://www.ahrq.gov/clinic/uspstf/grades.htm**: **A -** The USPSTF recommends the service. There is high certainty that the net benefit is substantial. **B -** The USPSTF recommends the service. There is high certainty that the net benefit is moderate or there is moderate certainty that the net benefit is moderate to substantial. **C -** The USPSTF recommends against routinely providing the service. There may be considerations that support providing the service in an individual patient. There is at least moderate certainty that the net benefit is small. Offer or provide this service only if other considerations support the offering or providing the service in an individual patient. **D -** The USPSTF recommends against the service. There is moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits. **I -** The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service. Evidence is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined.

Comment [KP8]: 2a. The measure is well defined and precisely specified so that it can be implemented consistently within and across organizations and allow for comparability. The required data elements are of high quality as defined by NQF's Health Information Technology Expert Panel (HITEP) .

2a.4 Denominator Statement (*Brief, text description of the denominator - target population being measured***):**
All adult (>=18 years old) hemodialysis or peritoneal dialysis patients in the facility for the entire three-month study period.

2a.5 Target population gender: Female, Male
2a.6 Target population age range: Adults 18 years or older.

2a.7 Denominator Time Window (*The time period in which cases are eligible for inclusion in the denominator***):**
Rolling three-month study period.

2a.8 Denominator Details (*All information required to collect/calculate the denominator - the target population being measured - including all codes, logic, and definitions***):**
Patients are included in the denominator if they are >= 18 years old, and on dialysis and in the facility for the entire study period.
The patient's age will be determined by subtracting the patient's date of birth from the first day of the study period. Patients in a facility and on dialysis for the entire study period are defined as follows: Admit Date to the specified facility is prior or equal to the first day of the study period, AND the patient has not been discharged (Discharge Date is null or blank), OR Discharge Date from the facility is greater than or equal to the last day of the study period AND Primary Type of Treatment is Hemodialysis, continuous ambulatory peritoneal dialysis (CAPD) or continuous cycling peritoneal dialysis (CCPD) in each month of the study period.

2a.9 Denominator Exclusions (*Brief text description of exclusions from the target population***):** None.

2a.10 Denominator Exclusion Details (*All information required to collect exclusions to the denominator, including all codes, logic, and definitions***):**
N/A

2a.11 Stratification Details/Variables (*All information required to stratify the measure including the stratification variables, all codes, logic, and definitions***):**
N/A

2a.12-13 Risk Adjustment Type: No risk adjustment necessary

2a.14 Risk Adjustment Methodology/Variables (*List risk adjustment variables and describe conceptual models, statistical models, or other aspects of model or method***):**
N/A

2a.15-17 Detailed risk model available Web page URL or attachment:

2a.18-19 Type of Score: Rate/proportion
2a.20 Interpretation of Score: Better quality = Higher score
2a.21 Calculation Algorithm (*Describe the calculation of the measure as a flowchart or series of steps***):**
Patients are included in the denominator if they are >= 18 years old, and on dialysis and in the facility for the entire study period.
The patient's age will be determined by subtracting the patient's date of birth from the first day of the study period. Patients in a facility and on dialysis for the entire study period are defined as follows: Admit Date to the specified facility is prior or equal to the first day of the study period, AND the patient has not been discharged (Discharge Date is null or blank), OR Discharge Date from the facility is greater than or equal to the last day of the study period AND Primary Type of Treatment is Hemodialysis, CAPD or CCPD in each month of the study period.
The numerator will be determined by counting the patients in the denominator who meet the following criteria in the three-month study period: Serum Ferritin is populated, AND Iron Saturation Percentage is populated AND Serum Ferritin Collection Date is equal to Iron Saturation Percentage Collection Date.

2a.22 Describe the method for discriminating performance *(e.g., significance testing)***:**
The performance of the facility will be compared to state, Network and national performance.

2a.23 Sampling (Survey) Methodology *If measure is based on a sample (or survey), provide instructions for*

Comment [k9]: 11 Risk factors that influence outcomes should not be specified as exclusions.
12 Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

obtaining the sample, conducting the survey and guidance on minimum sample size (response rate):
N/A

2a.24 **Data Source (***Check the source(s) for which the measure is specified and tested)*
Electronic clinical data

2a.25 **Data source/data collection instrument (***Identify the specific data source/data collection instrument, e.g. name of database, clinical registry, collection instrument, etc.***):**
CROWNWeb

2a.26-28 **Data source/data collection instrument reference web page URL or attachment:** URL
http://www.projectcrownweb.org/crown/index.php

2a.29-31 **Data dictionary/code table web page URL or attachment:** URL
http://www.projectcrownweb.org/crown/index.php?page=Public_Documents&subPage=Release_Documents

2a.32-35 **Level of Measurement/Analysis** (*Check the level(s) for which the measure is specified and tested)*
Facility/Agency

2a.36-37 **Care Settings (***Check the setting(s) for which the measure is specified and tested)*
Dialysis Facility

2a.38-41 **Clinical Services** (*Healthcare services being measured, check all that apply*)
Dialysis

TESTING/ANALYSIS

2b. Reliability testing

2b.1 **Data/sample** *(description of data/sample and size)***:** The measure has not been tested for reliability.

2b.2 **Analytic Method** *(type of reliability & rationale, method for testing)***:**
Since the data are submitted electronically, we anticipate highly reliable measures. No elements for the measure would be abstracted from records, and no elements would be susceptible to inter-rater variability. Reliability testing of the CROWNWeb data has not yet been performed although monthly reports are currently being distributed to facilities participating in Phase 1 and 2 to compare the metrics to their own data.

2b.3 **Testing Results** *(reliability statistics, assessment of adequacy in the context of norms for the test conducted)***:**
N/A

2b C☐ P☐ M☐ N☐

2c. Validity testing

2c.1 **Data/sample** *(description of data/sample and size)***:** N/A

2c.2 **Analytic Method** *(type of validity & rationale, method for testing)***:**
Face validity is the only validity assessed, therefore testing is not applicable.

2c.3 **Testing Results** *(statistical results, assessment of adequacy in the context of norms for the test conducted)***:**
N/A

2c C☐ P☐ M☐ N☐

2d. Exclusions Justified

2d.1 **Summary of Evidence supporting exclusion(s):**
N/A

2d C☐ P☐ M☐ N☐

Comment [KP10]: 2b. Reliability testing demonstrates the measure results are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period.

Comment [k11]: 8 Examples of reliability testing 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 may address the data items or final measure score.

Comment [KP12]: 2c. Validity testing demonstrates that the measure reflects the quality of care provided, adequately distinguishing good and poor quality. If face validity is the only validity addressed, it is systematically assessed.

Comment [k13]: 9 Examples of validity testing include, but are not limited to: determining if measure scores adequately distinguish between providers known to have good or poor quality assessed by another valid method; correlation of measure scores with another valid indicator of quality for the specific topic; ability of measure scores to predict scores on some other related valid measure; content validity for multi-item scales/tests. Face validity is a subjective assessment by experts of whether the measure reflects the quality of care (e.g., whether the proportion of patients with BP < 140/90 is a marker of quality). If face validity is the only validity addressed, it is systematically assessed (e.g., ratings by relevant stakeholders) and the measure is judged to represent quality care for the specific topic and that the measure focus is the most important aspect of quality for the specific topic.

Comment [KP14]: 2d. Clinically necessary measure exclusions are identified and must be:
•supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion;
AND
•a clinically appropriate exception (e.g., contraindication) to eligibility for the measure focus;
AND
•precisely defined and specified:
–if there is substantial variability in exclusions across providers, the measure is specified so that exclusions are computable and the effect on the measure is transparent (i.e., impact clearly delineated, such as number of cases excluded, exclusion rates by type of exclusion);
if patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that it strongly impacts performance on the measure and the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category ... [2]

Comment [k15]: 10 Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, sensitivity analyses with and without the exclusion, and variability of exclusions across providers.

2d.2 Citations for Evidence: N/A **2d.3 Data/sample** *(description of data/sample and size)*: N/A **2d.4 Analytic Method** *(type analysis & rationale)*: N/A **2d.5 Testing Results** *(e.g., frequency, variability, sensitivity analyses)*: N/A	**NA**☐
2e. Risk Adjustment for Outcomes/ Resource Use Measures **2e.1 Data/sample** *(description of data/sample and size)*: N/A **2e.2 Analytic Method** *(type of risk adjustment, analysis, & rationale)*: N/A **2e.3 Testing Results** *(risk model performance metrics)*: N/A **2e.4 If outcome or resource use measure is not risk adjusted, provide rationale:** N/A	**2e** **C**☐ **P**☐ **M**☐ **N**☐ **NA**☐
2f. Identification of Meaningful Differences in Performance **2f.1 Data/sample from Testing or Current Use** *(description of data/sample and size)*: A test calculation of the measure was performed using CROWNWeb Phase II data from July-September 2009. The calculation included data for 3384 facilities. **2f.2 Methods to identify statistically significant and practically/meaningfully differences in performance** *(type of analysis & rationale)*: The distribution of percent of patients meeting measure criteria by facility was examined. **2f.3 Provide Measure Scores from Testing or Current Use** *(description of scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance)*: In the test calculation of the measure using July-September 2009 CROWNWeb data, the facility-level mean was 85%. The median, 25th, and 75th percentiles were 88%, 81% and 94%, respectively.	**2f** **C**☐ **P**☐ **M**☐ **N**☐
2g. Comparability of Multiple Data Sources/Methods **2g.1 Data/sample** *(description of data/sample and size)*: N/A **2g.2 Analytic Method** *(type of analysis & rationale)*: N/A **2g.3 Testing Results** *(e.g., correlation statistics, comparison of rankings)*: N/A	**2g** **C**☐ **P**☐ **M**☐ **N**☐ **NA**☐
2h. Disparities in Care **2h.1 If measure is stratified, provide stratified results** *(scores by stratified categories/cohorts)*: N/A **2h.2 If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans:** N/A	**2h** **C**☐ **P**☐ **M**☐ **N**☐ **NA**☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Scientific Acceptability of Measure Properties?*	**2**
Steering Committee: Overall, to what extent was the criterion, *Scientific Acceptability of Measure Properties*, **met?**	*2* **C**☐

Comment [KP16]: 2e. For outcome measures and other measures (e.g., resource use) when indicated:
•an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified and is based on patient clinical factors that influence the measured outcome (but not disparities in care) and are present at start of care;[Error! Bookmark not defined.] OR rationale/data support no risk adjustment.

Comment [k17]: 13 Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care such as race, socioeconomic status, gender (e.g., poorer treatment outcomes of African American men with prostate cancer, inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than adjusting out differences.

Comment [KP18]: 2f. Data analysis demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful differences in performance.

Comment [k19]: 14 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% v. 75%) 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 poor performance may not demonstrate much variability across providers.

Comment [KP20]: 2g. If multiple data sources/methods are allowed, there is demonstration they produce comparable results.

Comment [KP21]: 2h. If disparities in care have been identified, measure specifications, scoring, and analysis allow for identification of disparities through stratification of results (e.g., by race, ethnicity, socioeconomic status, gender);OR rationale/data justifies why stratification is not necessary or not feasible.

Rationale:	**P☐ M☐ N☐**
3. USABILITY	
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. (evaluation criteria)	Eval Rating
3a. Meaningful, Understandable, and Useful Information **3a.1 Current Use:** In use **3a.2 Use in a public reporting initiative (disclosure of performance results to the public at large)** (*If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). If not publicly reported, state the plans to achieve public reporting within 3 years*)**:** N/A **3a.3 If used in other programs/initiatives (***If used in quality improvement or other programs/initiatives, name of initiative(s), locations, Web page URL(s). If not used for QI, state the plans to achieve use for QI within 3 years*)**:** A similar measure is calculated in the monthly CROWNWeb Phase III CPM reporting (Anemia Management CPM IIa: Assessment of Iron Stores). Our new proposed process measure has been proposed to replace CPM IIa. **Testing of Interpretability** (*Testing that demonstrates the results are understood by the potential users for public reporting and quality improvement*) **3a.4 Data/sample** *(description of data/sample and size)***:** N/A **3a.5 Methods** *(e.g., focus group, survey, QI project)***:** N/A **3a.6 Results** *(qualitative and/or quantitative results and conclusions)***:** N/A	**3a** **C☐** **P☐** **M☐** **N☐**
3b/3c. Relation to other NQF-endorsed measures **3b.1 NQF # and Title of similar or related measures:** NFQ # 0252 Phase III ESRD CPM Anemia Management CPM IIa: Assessment of Iron Stores	
(for NQF staff use) **Notes on similar/related** endorsed **or submitted measures:**	
3b. Harmonization If this measure is related to measure(s) already endorsed by NQF (e.g., same topic, but different target population/setting/data source or different topic but same target population): **3b.2 Are the measure specifications harmonized? If not, why?** Yes. Our new proposed process measure has been proposed to replace CPM IIa.	**3b** **C☐** **P☐** **M☐** **N☐** **NA☐**
3c. Distinctive or Additive Value **3c.1 Describe the distinctive, improved, or additive value this measure provides to existing NQF-endorsed measures:** Revisions to current CPM: • The specification that serum ferritin and TSAT be measured on the same day. This requirement is based on the rationale that the clinical utility of serum ferritin and TSAT is highest when measured and interpreted together. Based on CROWNWeb test data, 99% of patients with ferritin and TSAT measured within one month had the measurements on the same day. • The specification to measure serum ferritin and TSAT for all patients, not just those receiving ESAs or with Hgb <11 g/dL: o This approach is consistent with the 2006 KDOQI recommendations. It is also consistent with the	**3c** **C☐** **P☐** **M☐** **N☐** **NA☐**

Comment [KP22]: 3a. Demonstration that information produced by the measure is meaningful, understandable, and useful to the intended audience(s) for both public reporting (e.g., focus group, cognitive testing) and informing quality improvement (e.g., quality improvement initiatives). An important outcome that may not have an identified improvement strategy still can be useful for informing quality improvement by identifying the need for and stimulating new approaches to improvement.

Comment [KP23]: 3b. The measure specifications are harmonized with other measures, and are applicable to multiple levels and settings.

Comment [k24]: 16 Measure harmonization refers to the standardization of specifications for similar measures on the same topic (e.g., *influenza immunization* of patients in hospitals or nursing homes), or related measures for the same target population (e.g., eye exam and HbA1c for *patients with diabetes*), or definitions applicable to many measures (e.g., age designation for children) so that they are uniform or compatible, unless differences are dictated by the evidence. The dimensions of harmonization can include numerator, denominator, exclusions, and data source and collection instructions. The extent of harmonization depends on the relationship of the measures, the evidence for the specific measure focus, and differences in data sources.

Comment [KP25]: 3c. Review of existing endorsed measures and measure sets demonstrates that the measure provides a distinctive or additive value to existing NQF-endorsed measures (e.g., provides a more complete picture of quality for a particular condition or aspect of healthcare, is a more valid or efficient way to measure).

current trend in practice to limit ESA therapy when possible, as judicious use of intravenous iron decreases ESA requirements and in some patients can support Hgb levels without the need for ESA therapy for several months or more. o Occasionally, patients have Hgb levels in the normal or near-normal range without requiring any long-term pharmacologic (ESA or iron) therapy. While some practitioners may not measure iron stores routinely in these patients, the prevalence of this condition is very low, estimated at 2% or less of dialysis patients. o Preliminary CROWNWeb data indicate that ferritin and TSAT levels are measured in ~95% of patients, and that this practice does not vary by Hgb level. • Dropping the use of reticulocyte hemoglobin content (CHr) as an alternative to TSAT levels for assessment or iron stores: o Although the use of CHr was added to the 2006 CPM to harmonize with the KDOQI guidelines, the utility of measuring CHr instead of TSAT for the assessment of iron stores is uncertain. o Additionally, the practice of CHr measurement remains uncommon in US dialysis facilities. Based on preliminary CROWNWeb data reported for 226,210 patients in December 2009, 204,905 (91%) had TSAT values; 10,363 (4.6%) had CHr values; and only 545 (0.2%) had CHr but not TSAT values. The addition of CHr to the CPM can be considered should future data support its utility, and its measurement becomes more common. 5.1 **If this measure is similar to measure(s) already endorsed by NQF (i.e., on the same topic and the same target population), Describe why it is a more valid or efficient way to measure quality:** Revisions to current CPM: • The specification that serum ferritin and TSAT be measured on the same day. This requirement is based on the rationale that the clinical utility of serum ferritin and TSAT is highest when measured and interpreted together. Based on CROWNWeb test data, 99% of patients with ferritin and TSAT measured within one month had the measurements on the same day. • The specification to measure serum ferritin and TSAT for all patients, not just those receiving ESAs or with Hgb <11 g/dL: o This approach is consistent with the 2006 KDOQI recommendations. It is also consistent with the current trend in practice to limit ESA therapy when possible, as judicious use of intravenous iron decreases ESA requirements and in some patients can support Hgb levels without the need for ESA therapy for several months or more. o Occasionally, patients have Hgb levels in the normal or near-normal range without requiring any long-term pharmacologic (ESA or iron) therapy. While some practitioners may not measure iron stores routinely in these patients, the prevalence of this condition is very low, estimated at 2% or less of dialysis patients. o Preliminary CROWNWeb data indicate that ferritin and TSAT levels are measured in ~95% of patients, and that this practice does not vary by Hgb level. • Dropping the use of reticulocyte hemoglobin content (CHr) as an alternative to TSAT levels for assessment or iron stores: o Although the use of CHr was added to the 2006 CPM to harmonize with the KDOQI guidelines, the utility of measuring CHr instead of TSAT for the assessment of iron stores is uncertain. o Additionally, the practice of CHr measurement remains uncommon in US dialysis facilities. Based on preliminary CROWNWeb data reported for 226,210 patients in December 2009, 204,905 (91%) had TSAT values; 10,363 (4.6%) had CHr values; and only 545 (0.2%) had CHr but not TSAT values. The addition of CHr to the CPM can be considered should future data support its utility, and its measurement becomes more common.	
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Usability*?	**3**
Steering Committee: Overall, to what extent was the criterion, *Usability*, met? **Rationale:**	**3** **C☐** **P☐** **M☐** **N☐**
4. FEASIBILITY	
Extent to which the required data are readily available, retrievable without undue burden, and can be	Eval

implemented for performance measurement. (evaluation criteria)	Rating
4a. Data Generated as a Byproduct of Care Processes **4a.1-2 How are the data elements that are needed to compute measure scores generated?** Data generated as byproduct of care processes during care delivery (Data are generated and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition)	**4a** C☐ P☐ M☐ N☐
4b. Electronic Sources **4b.1 Are all the data elements available electronically?** (*elements that are needed to compute measure scores are in defined, computer-readable fields, e.g., electronic health record, electronic claims*) Yes **4b.2 If not, specify the near-term path to achieve electronic capture by most providers.**	**4b** C☐ P☐ M☐ N☐
4c. Exclusions **4c.1 Do the specified exclusions require additional data sources beyond what is required for the numerator and denominator specifications?** No **4c.2 If yes, provide justification.**	**4c** C☐ P☐ M☐ N☐ NA☐
4d. Susceptibility to Inaccuracies, Errors, or Unintended Consequences **4d.1 Identify susceptibility to inaccuracies, errors, or unintended consequences of the measure and describe how these potential problems could be audited. If audited, provide results.** Data are from the electronic CROWNWeb system, and are minimally susceptible to inaccuracies and errors.	**4d** C☐ P☐ M☐ N☐
4e. Data Collection Strategy/Implementation **4e.1 Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data/missing data, timing/frequency of data collection, patient confidentiality, time/cost of data collection, other feasibility/ implementation issues:** Data are already collected in the CROWNWeb system. **4e.2 Costs to implement the measure** (*costs of data collection, fees associated with proprietary measures*)**:** Data are already collected in the CROWNWeb system. **4e.3 Evidence for costs:** N/A **4e.4 Business case documentation:** Iron status testing is an important step in hemoglobin management. Maintaining hemoglobin (hgb) within a normal range is essential to reducing patient risk of adverse outcomes, often resulting in hospitalization or intensified patient care. This measure is also intended to encourage optimum utilization of pharmacologic and laboratory resources.	**4e** C☐ P☐ M☐ N☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Feasibility*?	**4**
Steering Committee: Overall, to what extent was the criterion, *Feasibility*, **met?** **Rationale:**	**4** C☐ P☐ M☐ N☐

Comment [KP26]: 4a. For clinical measures, required data elements are routinely generated concurrent with and as a byproduct of care processes during care delivery. (e.g., BP recorded in the electronic record, not abstracted from the record later by other personnel; patient self-assessment tools, e.g., depression scale; lab values, meds, etc.)

Comment [KP27]: 4b. The required data elements are available in electronic sources. If the required data are not in existing electronic sources, a credible, near-term path to electronic collection by most providers is specified and clinical data elements are specified for transition to the electronic health record.

Comment [KP28]: 4c. Exclusions should not require additional data sources beyond what is required for scoring the measure (e.g., numerator and denominator) unless justified as supporting measure validity.

Comment [KP29]: 4d. Susceptibility to inaccuracies, errors, or unintended consequences and the ability to audit the data items to detect such problems are identified.

Comment [KP30]: 4e. Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, etc.) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use).

RECOMMENDATION	
(for NQF staff use) **Check if measure is untested and only eligible for time-limited endorsement.**	**Time-limited** ☐
Steering Committee: Do you recommend for endorsement? **Comments:**	**Y**☐ **N**☐ **A**☐

CONTACT INFORMATION

Co.1 Measure Steward (Intellectual Property Owner)
Co.1 Organization
Centers for Medicare and Medicaid Services, 7500 Security Boulevard, Baltimore, Maryland, 21244

Co.2 Point of Contact
Thomas, Dudley, Thomas.Dudley@cms.hhs.gov, 410-786-1442-

Measure Developer If different from Measure Steward
Co.3 Organization
Arbor Research/UM-KECC, 315 W. Huron, Ann Arbor, Michigan, 48103

Co.4 Point of Contact
Adrienne, Janney, adrienne.janney@arborresearch.org, 734-665-4108-

Co.5 Submitter If different from Measure Steward POC
Thomas, Dudley, Thomas.Dudley@cms.hhs.gov, 410-786-1442-, Centers for Medicare and Medicaid Services

Co.6 Additional organizations that sponsored/participated in measure development

ADDITIONAL INFORMATION

Workgroup/Expert Panel involved in measure development
Ad.1 Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
Dr. David VanWyck, panel chair (Vice President, Clinical Services, DaVita)
Dr. Lynda Szczech (Duke University School of Medicine, Durham, NC)
Dr. John Stivelman (University of Washington School of Medicine/Northwest Kidney Centers, Seattle, WA)
Dr. David Gilbertson (USRDS, Minneapolis, MN)
Dr. Michael Lazarus (Senior Executive Vice President, Fresenius Medical Care NA)
Dr. Ajay Singh (Brigham and Women's Hospital, Boston, MA)
Dr. Bruce Robinson, Moderator (Arbor Research Collaborative for Health, Ann Arbor, MI)
Flannery Campbell, MS, Analyst (University of Michigan, Ann Arbor, MI)

Ad.2 If adapted, provide name of original measure:
Ad.3-5 If adapted, provide original specifications URL or attachment

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.6 Year the measure was first released:
Ad.7 Month and Year of most recent revision:
Ad.8 What is your frequency for review/update of this measure? 3 years
Ad.9 When is the next scheduled review/update for this measure? 2013

Ad.10 Copyright statement/disclaimers:

Ad.11 -13 Additional Information web page URL or attachment:

Date of Submission (*MM/DD/YY*)**:** 09/27/2010

Page 3: [1] Comment [k5] **Karen Pace** **10/5/2009 8:59:00 AM**

4 Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multi-step process, the step with the greatest effect on the desired outcome should be selected as the focus of measurement. For example, although assessment of immunization status and recommending immunization are necessary steps, they are not sufficient to achieve the desired impact on health status – patients must be vaccinated to achieve immunity. This does not preclude consideration of measures of preventive screening interventions where there is a strong link with desired outcomes (e.g., mammography) or measures for multiple care processes that affect a single outcome.

Page 7: [2] Comment [KP14] **Karen Pace** **10/5/2009 8:59:00 AM**

2d. Clinically necessary measure exclusions are identified and must be:

- supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion;

AND

- a clinically appropriate exception (e.g., contraindication) to eligibility for the measure focus;

AND

- precisely defined and specified:
  - if there is substantial variability in exclusions across providers, the measure is specified so that exclusions are computable and the effect on the measure is transparent (i.e., impact clearly delineated, such as number of cases excluded, exclusion rates by type of exclusion);

if patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that it strongly impacts performance on the measure and 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).