NATIONAL QUALITY FORUM

Measure Submission and Evaluation Worksheet 5.0

This form contains the information submitted by measure developers/stewards, organized according to NQF's measure evaluation criteria and process. The evaluation criteria, evaluation guidance documents, and a blank online submission form are available on the submitting standards web page.

NQF #: 0378 NQF Project: Cancer Project

(for Endorsement Maintenance Review)
Original Endorsement Date: Jul 31, 2008 Most Recent Endorsement Date: Jul 31, 2008 Last Updated Date: Mar 06, 2012

BRIEF MEASURE INFORMATION

De.1 Measure Title: MDS: Documentation of Iron Stores in Patients Receiving Erythropoietin Therapy

Co.1.1 Measure Steward: American Medical Association - Physician Consortium for Performance Improvement

De.2 Brief Description of Measure: Percentage of patients aged 18 years and older with a diagnosis of MDS who are receiving erythropoietin therapy with documentation of iron stores prior to initiating erythropoietin therapy

2a1.1 Numerator Statement: Patients with documentation* of iron stores prior to initiating erythropoietin therapy

*Documentation includes either: bone marrow examination including iron stain OR serum iron measurement by ferritin or serum iron and TIBC

2a1.4 Denominator Statement: All patients aged 18 years and older with a diagnosis of MDS who are receiving erythropoietin therapy

2a1.8 Denominator Exclusions: Documentation of system reason(s) for not documenting iron stores prior to initiating erythropoietin therapy

1.1 Measure Type: Process
2a1. 25-26 Data Source: Administrative claims, Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Electronic Clinical Data : Laboratory
2a1.33 Level of Analysis: Clinician : Group/Practice, Clinician : Individual, Clinician : Team

1.2-1.4 Is this measure paired with another measure? No

De.3 If included in a composite, please identify the composite measure (*title and NQF number if endorsed*):

STAFF NOTES (*issues or questions regarding any criteria*)

Comments on Conditions for Consideration:

Is the measure untested? Yes☐ No☐ If untested, explain how it meets criteria for consideration for time-limited endorsement:

1a. Specific national health goal/priority identified by DHHS or NPP addressed by the measure (*check De.5*):
5. Similar/related endorsed or submitted measures (*check 5.1*):
Other Criteria:

Staff Reviewer Name(s):

1. IMPACT, OPPORTUITY, EVIDENCE - IMPORTANCE TO MEASURE AND REPORT

Importance to Measure and Report is a threshold criterion that must be met in order to recommend a measure for endorsement. All three subcriteria must be met to pass this criterion. See guidance on evidence.
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: H☐ M☐ L☐ I ☐
(*The measure directly addresses a specific national health goal/priority identified by DHHS or NPP, or some other high impact aspect of healthcare.*)

De.4 Subject/Topic Areas *(Check all the areas that apply)*: Cancer, Cancer : Hematologic
De.5 Cross Cutting Areas (*Check all the areas that apply)*:

1a.1 Demonstrated High Impact Aspect of Healthcare: Patient/societal consequences of poor quality, Severity of illness

1a.2 If "Other," please describe:

1a.3 Summary of Evidence of High Impact *(Provide epidemiologic or resource use data)*:
The American Cancer Society´s most recent estimates for myelodysplastic syndrome
(MDS) in the United States are about 12,000 new cases each year. The number of new cases seems to be going up as the average age of the population has increased. About 80% to 90% of all patients with MDS are older than 60 years. It is rare in young adults.

The risk of developing MDS increases with age. In one study, the annual incidence per 100,000 was estimated to be 0.5, 5.3, 15, 49, and 89 for individuals <50 years of age, 50 to 59, 60 to 69; 70 to 79; and >80 years, respectively.

1a.4 Citations for Evidence of High Impact cited in 1a.3: The American Cancer Society. Last Accessed: January, 12, 2012. Available at: http://www.cancer.org/Cancer/MyelodysplasticSyndrome/OverviewGuide/myelodysplastic-syndromes-overview-key-statistics

Williamson PJ, Kruger AR, Reynolds PJ, et al. Establishing the incidence of myelodysplastic syndrome. Br J Haematol 1994; 87:743.

1b. Opportunity for Improvement: H☐ M☐ L☐ I ☐
(*There is a demonstrated performance gap - variability or overall less than optimal performance*)

1b.1 Briefly explain the benefits (improvements in quality) envisioned by use of this measure:
To be effective erythropoietin therapy requires that adequate iron stores be present due to iron's importance in red-blood-cell synthesis. By promoting the documentation of adequate iron stores in MDS patients requiring EPO therapy, the efficacy of the treatment will be enhanced.

1b.2 Summary of Data Demonstrating Performance Gap *(Variation or overall less than optimal performance across providers)*:
[*For Maintenance – Descriptive statistics for performance results for this measure - distribution of scores for measured entities by quartile/decile, mean, median, SD, min, max, etc.*]
CMS Physician Quality Reporting Initiative
This measure was used in the 2007-2010 CMS Physician Quality Reporting Initiative Claims and Registry options.

There is a gap in care as shown by this 2008 data; 58.00% of patients reported on did not meet the measure.

10th percentile: 0.00%
25th percentile: 6.91%
50th percentile: 30.22%
75th percentile: 66.67%
90th percentile: 97.44%

The mean performance rate for PQRS 2009 was reported as 97.69% with a total of 1,269 eligible professionals submitting. demonstrating an opportunity for improvement. Unfortunately, data regarding the variability in performance rates across reporting eligible professionals for PQRS 2009 is not available at this time.

PQRS is a voluntary reporting program and performance rates may not be representative of all physicians treating patients with MDS.

1b.3 Citations for Data on Performance Gap: [*For Maintenance – Description of the data or sample for measure results reported in 1b.2 including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included*]
Confidential CMS PQRI 2008 and 2009 Performance Information by Measure. Jan-Sept TAP file.

1b.4 Summary of Data on Disparities by Population Group: [*For Maintenance –Descriptive statistics for performance results for this measure by population group*]
We are not aware of any publications/evidence outlining disparities for the documentation of iron stores in patients receiving erythropoietin therapy.

1b.5 Citations for Data on Disparities Cited in 1b.4: [*For Maintenance – Description of the data or sample for measure results reported in 1b.4 including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included*]

1c. Evidence (*Measure focus is a health outcome OR meets the criteria for quantity, quality, consistency of the body of evidence.*)
Is the measure focus a health outcome? Yes☐ No☐ If not a health outcome, rate the body of evidence.

Quantity: H☐ M☐ L☐ I ☐ Quality: H☐ M☐ L☐ I ☐ Consistency: H☐ M☐ L☐ I ☐

Quantity	Quality	Consistency	Does the measure pass subcriterion1c?
M-H	M-H	M-H	Yes☐
L	M-H	M	Yes☐ IF additional research unlikely to change conclusion that benefits to patients outweigh harms: otherwise No☐
M-H	L	M-H	Yes☐ IF potential benefits to patients clearly outweigh potential harms: otherwise No☐
L-M-H	L-M-H	L	No ☐

Health outcome – rationale supports relationship to at least one healthcare structure, process, intervention, or service	Does the measure pass subcriterion1c? Yes☐ IF rationale supports relationship

1c.1 Structure-Process-Outcome Relationship *(Briefly state the measure focus, e.g., health outcome, intermediate clinical outcome, process, structure; then identify the appropriate links, e.g., structure-process-health outcome; process- health outcome; intermediate clinical outcome-health outcome)*:
Erythropoietin therapy offers a safer alternative than red blood cell transfusion for a wide range of anemic patients with MDS. To be effective erythropoietin requires that adequate iron stores be present due to iron's importance in red-blood-cell synthesis. Iron deficiency presents a major limitation to the efficacy of erythropoietin therapy.

1c.2-3 Type of Evidence *(Check all that apply)*:
Clinical Practice Guideline

1c.4 Directness of Evidence to the Specified Measure *(State the central topic, population, and outcomes addressed in the body of evidence and identify any differences from the measure focus and measure target population)*:
Both our measure and the NCCN guidelines state that iron repletion be verified before instituting Epo or darbepoetin therapy.

1c.5 Quantity of Studies in the Body of Evidence *(Total number of studies, not articles)*: The description of the evidence review in the guideline did not address the overall quantity of studies in the body of evidence. However NCCN guidelines for MDS reference 160 articles.

1c.6 Quality of Body of Evidence (*Summarize the certainty or confidence in the estimates of benefits and harms to patients*

across studies in the body of evidence resulting from study factors. Please address: a) study design/flaws; b) directness/indirectness of the evidence to this measure (e.g., interventions, comparisons, outcomes assessed, population included in the evidence); and c) imprecision/wide confidence intervals due to few patients or events): The quality of the body of evidence supporting the guideline recommendation is summarized according to the NCCN categories of evidence and consensus as being based on "lower-level evidence". Lower-level evidence is later described as evidence that may include non-randomized trials; case series; or when other data are lacking, the clinical experience of expert physicians.

1c.7 Consistency of Results across Studies *(Summarize the consistency of the magnitude and direction of the effect):* Although there is no explicit statement regarding the overall consistency of results across studies in the guidelines supporting the measure, the recommendation received uniform NCCN consensus that the intervention is appropriate.

In March 2007 and 2008, the FDA announced alerts and strengthened safety warnings for the use of Erythropoiesis-Stimulating Agents (ESAs).

1c.8 Net Benefit *(Provide estimates of effect for benefit/outcome; identify harms addressed and estimates of effect; and net benefit - benefit over harms)*:
To be effective erythropoietin requires that adequate iron stores be present due to iron's importance in red-blood-cell synthesis. Iron deficiency presents a major limitation to the efficacy of erythropoietin therapy.

1c.9 Grading of Strength/Quality of the Body of Evidence. Has the body of evidence been graded? Yes

1c.10 If body of evidence graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: A panel of experts with members from each of the NCCN Member Institutions develops the NCCN Guidelines. Specialties that must be included on a particular panel are identified before that panel is convened but also evolve as the standard of care changes over time. This multidisciplinary representation varies from panel to panel. The NCCN Guidelines Panel Chairs are charged with ensuring that representatives of all treatment strategies are included. Many of the panels also include a patient representative, especially when issues of long-term care and patient preference are paramount in the panel´s considerations. NCCN publishes individual disclosures of potential conflicts of interest for panel members, NCCN Guidelines staff, and NCCN senior management. Relationships disclosed include research funding, participation in advisory groups, participation in speakers' bureaus, employment, and equity or patent ownership. Beginning in 2010, the NCCN Board of Directors has directed that panel members compensation from external sources be less than published thresholds. These thresholds are <= $20,000 from a single entity and <= $50,000 in aggregate from any source.

The ASCO Clinical Practice Guidelines Committee convened the ASCO/ASH Update Committee to lead the 2010 update. The Update Committee met via a series of teleconferences to review evidence collected from the systematic review and make revisions to the guideline recommendations as warranted.The guideline was reviewed and approved by the entire Update Committee, ASCO's Clinical Practice Guidelines Committee, ASH's Committee on Practice, ASH's Subcommittee on Quality of Care, the ASCO Board of Directors, and the ASH Executive Committee.

1c.11 System Used for Grading the Body of Evidence: Other

1c.12 If other, identify and describe the grading scale with definitions: NCCN Categories of Evidence and Consensus Panel members identify the level of evidence supporting each recommendation. These categories are:

•Category 1: Based upon high-level evidence, there is uniform NCCN consensus that the intervention is appropriate.
•Category 2A: Based upon lower-level evidence, there is uniform NCCN consensus that the intervention is appropriate.
•Category 2B: Based upon lower-level evidence, there is NCCN consensus that the intervention is appropriate.
•Category 3: Based upon any level of evidence, there is major NCCN disagreement that the intervention is appropriate.

1c.13 Grade Assigned to the Body of Evidence: Category 2A

1c.14 Summary of Controversy/Contradictory Evidence: No controversy or contradictory evidence with regard to the importance of identifying documentation of iron stores in patients with MDS.

1c.15 Citations for Evidence other than Guidelines*(Guidelines addressed below)*:

1c.16 Quote verbatim, the specific guideline recommendation *(Including guideline # and/or page #)*:
Anemia related to MDS generally presents as a hypoproductive macrocytic anemia, often associated with suboptimal elevation of serum Epo levels. To determined FAB subtype, iron status, and the level of ring sideroblasts, bone marrow aspiration with iron stain, biopsy, and cytogenetics should be examines. Patients should also be considered for HLA-DR15 typing as indicated above. Iron repletion needs to be verified before instituting Epo or darbepoetintherapy.

2010 recommendation by American Society of Hematology: This recommendation remains the same as in 2007. Baseline and periodic monitoring of iron, total iron-binding capacity, transferrin saturation, or ferritin levels and instituting iron repletion when indicated may help to reduce the need for ESAs, maximize symptomatic improvement for patients, and determine the reason for failure to respond adequately to ESA therapy.

1c.17 Clinical Practice Guideline Citation: National Comprehensive Cancer Network (NCCN). Clinical Practice Guidelines in Oncology: Myelodysplastic syndromes. Version 1, 2012. Available at: http://www.nccn.org/professionals/physician_gls/pdf/mds.pdf

J. Douglas Rizzo, Melissa Brouwers, Patricia Hurley, Jerome Seidenfeld, Murat O. Arcasoy, Jerry L. Spivak, Charles L. Bennett, Julia Bohlius, Darren Evanchuk, Matthew J. Goode, Ann A. Jakubowski, David H. Regan and Mark R. Somerfield. Approved by the American Society of Clinical Oncology Board of Directors on July 7, 2010. Approved by the Executive Committee of the American Society of Hematology on July 14, 2010. Available here: http://www.hematology.org/Practice/Guidelines/2934.aspx

1c.18 National Guideline Clearinghouse or other URL: www.nccn.org

1c.19 Grading of Strength of Guideline Recommendation. Has the recommendation been graded? Yes

1c.20 If guideline recommendation graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: See 1c.10 above

1c.21 System Used for Grading the Strength of Guideline Recommendation: Other

1c.22 If other, identify and describe the grading scale with definitions: NCCN Categories of Evidence and Consensus Panel members identify the level of evidence supporting each recommendation. These categories are:

•Category 1: Based upon high-level evidence, there is uniform NCCN consensus that the intervention is appropriate.
•Category 2A: Based upon lower-level evidence, there is uniform NCCN consensus that the intervention is appropriate.
•Category 2B: Based upon lower-level evidence, there is NCCN consensus that the intervention is appropriate.
•Category 3: Based upon any level of evidence, there is major NCCN disagreement that the intervention is appropriate.

1c.23 Grade Assigned to the Recommendation: Category 2A

1c.24 Rationale for Using this Guideline Over Others: It is the PCPI policy to use guidelines, which are evidence-based, applicable to physicians and other health-care providers, and developed by a national specialty organization or government agency. In addition, the PCPI has now expanded what is acceptable as the evidence base for measures to include documented quality improvement (QI) initiatives or implementation projects that have demonstrated improvement in quality of care.

Based on the NQF descriptions for rating the evidence, what was the developer's assessment of the quantity, quality, and consistency of the body of evidence?
1c.25 Quantity: Moderate 1c.26 Quality: Moderate1c.27 Consistency: Moderate

Was the threshold criterion, *Importance to Measure and Report*, met?
(*1a & 1b must be rated moderate or high and 1c yes*) Yes☐ No☐
Provide rationale based on specific subcriteria:

For a new measure if the Committee votes NO, then STOP.
For a measure undergoing endorsement maintenance, if the Committee votes NO because of 1b. (no opportunity for improvement), it may be considered for continued endorsement and all criteria need to be evaluated.

2. RELIABILITY & VALIDITY - 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)
Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate field. Supplemental materials may be referenced or attached in item 2.1. See guidance on measure testing.

S.1 Measure Web Page *(In the future, NQF will require measure stewards to provide a URL link to a web page where current detailed specifications can be obtained*). Do you have a web page where current detailed specifications for this measure can be obtained? Yes

S.2 If yes, provide web page URL: www.physicianconsortium.org

2a. RELIABILITY. Precise Specifications and Reliability Testing: H☐ M☐ L☐ I☐

2a1. Precise Measure Specifications. (*The measure specifications precise and unambiguous.*)

2a1.1 Numerator Statement *(Brief, narrative description of the measure focus or what is being measured about the target population, e.g., cases from the target population with the target process, condition, event, or outcome)*:
Patients with documentation* of iron stores prior to initiating erythropoietin therapy

*Documentation includes either: bone marrow examination including iron stain OR serum iron measurement by ferritin or serum iron and TIBC

2a1.2 Numerator Time Window *(The time period in which the target process, condition, event, or outcome is eligible for inclusion)*:
At least once during measurement period

2a1.3 Numerator Details *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, codes with descriptors, and/or specific data collection items/responses*:
Definition: *Documentation of Iron Stores: Includes either: bone marrow examination including iron stain OR serum iron measurement by ferritin or serum iron and TIBC.

Definition: Erythropoietin Therapy: Includes the following medications: epoetin and darbepoetin for the purpose of this measure.

For EHR: especification currently under development. Data elements (using Quality Data Model) required for the measure attached.

Administrative claims:

CPT Category II code: 3160F: Documentation of iron stores prior to initiating erythropoietin therapy

2a1.4 Denominator Statement *(Brief, narrative description of the target population being measured)*:
All patients aged 18 years and older with a diagnosis of MDS who are receiving erythropoietin therapy

2a1.5 Target Population Category *(Check all the populations for which the measure is specified and tested if any)*: Adult/Elderly Care

2a1.6 Denominator Time Window *(The time period in which cases are eligible for inclusion)*:
12 consecutive months

2a1.7 Denominator Details (*All information required to identify and calculate the target population/denominator such as definitions, codes with descriptors, and/or specific data collection items/responses)*:
For EHR: especification currently under development. Data elements (using Quality Data Model) required for the measure attached.

Administrative claims:

AGE: >= 18 years and older

ICD-9-CM diagnosis codes: 238.72, 238.73, 238.74, 238.75
ICD-10-CM diagnosis codes: D46.0, D46.1, D46.20, D46.21, D46.22, D46.4, D46.9, D46.a, D46.b, D46.c, D46.z

Diagnosis: MDS

AND

CPT codes: 99201, 99202, 99203, 99204, 99205, 99212, 99213, 99214, 99215, 99241, 99242, 99243, 99244, 99245

AND

CPT category II 4090F: Patient receiving erythropoietin therapy

2a1.8 Denominator Exclusions *(Brief narrative description of exclusions from the target population)*:
Documentation of system reason(s) for not documenting iron stores prior to initiating erythropoietin therapy

2a1.9 Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, codes with descriptors, and/or specific data collection items/responses)*:
The PCPI methodology uses three categories of reasons for which a patient may be excluded from the denominator of an individual measure. These measure exception categories are not uniformly relevant across all measures; for each measure, there must be a clear rationale to permit an exception for a medical, patient, or system reason. Examples are provided in the measure exception language of instances that may constitute an exception and are intended to serve as a guide to clinicians. For this measure, exceptions may include system reasons, e.g. for not documenting iron stores prior to initiating erythropoietin therapy. Where examples of exceptions are included in the measure language, these examples are coded and included in the eSpecifications. Although this methodology does not require the external reporting of more detailed exception data, the PCPI recommends that physicians document the specific reasons for exception in patients' medical records for purposes of optimal patient management and audit-readiness. The PCPI also advocates the systematic review and analysis of each physician's exceptions data to identify practice patterns and opportunities for quality improvement. For example, it is possible for implementers to calculate the percentage of patients that physicians have identified as meeting the criteria for exception. Additional details by data source are as follows:

For EHR: especification currently under development. Data elements (using Quality Data Model) required for the measure attached.

Administrative claims:

Denominator Exceptions:
Documentation of system reason(s) for not documenting iron stores prior to initiating erythropoietin therapy
Append modifier to CPT Category II code: 3160F-3P

2a1.10 Stratification Details/Variables *(All information required to stratify the measure results including the stratification variables, codes with descriptors, definitions, and/or specific data collection items/responses)*:
We encourage the results of this measure to be stratified by race, ethnicity, gender, and primary language, and have included these variables as recommended data elements to be collected.

2a1.11 Risk Adjustment Type *(Select type. Provide specifications for risk stratification in 2a1.10 and for statistical model in 2a1.13)*: No risk adjustment or risk stratification 2a1.12 If "Other," please describe:

2a1.13 Statistical Risk Model and Variables *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development should be addressed in 2b4.)*:

2a1.14-16 Detailed Risk Model Available at Web page URL (or attachment). Include coefficients, equations, codes with descriptors, definitions, and/or specific data collection items/responses. Attach documents only if they are not available on a webpage and keep attached file to 5 MB or less. NQF strongly prefers you make documents available at a Web page URL. Please supply login/password if needed:

2a1.17-18. Type of Score: Rate/proportion

2a1.19 Interpretation of Score *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*: Better quality = Higher score

2a1.20 Calculation Algorithm/Measure Logic*(Describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; etc.)*:
To calculate performance rates:
1) Find the patients who meet the initial patient population (ie, the general group of patients that the performance measure is designed to address).
2) From the patients within the initial patient population criteria, find the patients who qualify for the denominator (ie, the specific group of patients for inclusion in a specific performance measure based on defined criteria). Note: in some cases the initial patient population and denominator are identical.
3) From the patients within the denominator, find the patients who qualify for the Numerator (ie, the group of patients in the denominator for whom a process or outcome of care occurs). Validate that the number of patients in the numerator is less than or equal to the number of patients in the denominator
4) From the patients who did not meet the numerator criteria, determine if the physician has documented that the patient meets any criteria for denominator exception when exceptions have been specified [for this measure: or system reason(s) (eg, for not documenting iron stores prior to initiating erythropoietin therapy)]. If the patient meets any exception criteria, they should be removed from the denominator for performance calculation. --Although the exception cases are removed from the denominator population for the performance calculation, the number of patients with valid exceptions should be calculated and reported along with performance rates to track variations in care and highlight possible areas of focus for QI.

If the patient does not meet the numerator and a valid exception is not present, this case represents a quality failure.

See attached calculation algorithm in 2a1.21.

2a1.21-23 Calculation Algorithm/Measure Logic Diagram URL or attachment:
Attachment
AMA-PCPI_Measure Calculation-Standard Measures-634631931846113738.pdf

2a1.24 Sampling (Survey) Methodology. If measure is based on a sample (or survey), provide instructions for obtaining the sample, conducting the survey and guidance on minimum sample size (response rate):
Not applicable. The measure does not require sampling or a survey.

2a1.25 Data Source *(Check all the sources for which the measure is specified and tested*). If other, please describe:
Administrative claims, Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Electronic Clinical Data : Laboratory

2a1.26 Data Source/Data Collection Instrument (*Identify the specific data source/data collection instrument, e.g. name of database, clinical registry, collection instrument, etc.*): Not Applicable

2a1.27-29 Data Source/data Collection Instrument Reference Web Page URL or Attachment:

2a1.30-32 Data Dictionary/Code Table Web Page URL or Attachment:
Attachment
0378 MDS_Iron Stores Data Elements_FINAL.pdf

2a1.33 Level of Analysis (*Check the levels of analysis for which the measure is specified and tested)*: Clinician : Group/Practice, Clinician : Individual, Clinician : Team

2a1.34-35 Care Setting *(Check all the settings for which the measure is specified and tested)*: Ambulatory Care : Clinic/Urgent Care, Ambulatory Care : Clinician Office, Laboratory

2a2. Reliability Testing. (*Reliability testing was conducted with appropriate method, scope, and adequate demonstration of reliability*.)

2a2.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
PCPI Testing Project
- Two hematology practice sites representing various types, locations and sizes were identified to participate in testing the measures
- Site A was a hematology group practice with eight physicians that cared for hematology patients. Site B was a large multi-specialty group clinic with 13 physicians that cared for hematology patients.
- Site A had a document retrieval system rather than a full-fledged EHR where data was scanned in and required searching. Site B had a fully functional EHR.
- Both sites were located in urban/suburban regions
- Hematology patient visit volume was 150 per day at site A and 120-150 per day at site B.
- Both sites were instructed to select 120 patient records (20 with acute leukemias and 35 for each of the following diagnoses: MDS, multiple myeloma and CLL).
- At site A the number of patients in practice in 2009 by specialty area was as follows:
  - Myelodysplastic Syndrome (MDS): 145 patients
- At site B the number of patients in practice in 2009 by specialty area was as follows:
  - Myelodysplastic Syndrome (MDS): 15 patients
- For this measure, the sample size included 41 abstracted patient charts. Site B did not have 30 patients in 2008 and 2009 so only 11 patients were included in the sample for this measure.
- The measurement period (data collected from patients seen) was between 1/1/2009 through 12/31/2009. Due to an inability to obtain the required number of patient records for acute leukemia and MDS during the specified measurement period, site B also included patients from 2008.
- Chart auditing was performed between 5/17/2010 and 7/15/2010
- Data auditing was performed between 8/2/2010 and 9/14/2010

2a2.2 Analytic Method *(Describe method of reliability testing & rationale)*:
Data abstracted from patient records were used to calculate inter-rater reliability for the measure.
41 MDS patient records were reviewed.

Data analysis included:
- Percent agreement
- Kappa statistic to adjust for chance agreement

2a2.3 Testing Results *(Reliability statistics, assessment of adequacy in the context of norms for the test conducted)*:
N, % Agreement, Kappa (95% Confidence Interval)
Overall Reliability: 41, 90.2%, 0.5470 (0.1578 – 0.9362)
Denominator Reliability: 41, 100.0%, Kappa is non-calculable*
Numerator Reliability: 41, 90.2%, 0.5470, (0.1578 – 0.9362)
Exceptions Reliability: 41, 100.0%, Kappa is non-calculable*

This measure demonstrates moderately reliable, as shown in results from the above analysis.

*Kappa Statistics cannot be calculated because of complete agreement. Confidence intervals cannot be calculated because to do so would involve dividing by zero which cannot be done.

2b. VALIDITY. Validity, Testing, including all Threats to Validity: H☐ M☐ L☐ I ☐

2b1.1 Describe how the measure specifications *(measure focus, target population, and exclusions)* are consistent with the evidence cited in support of the measure focus (*criterion 1c)* and identify any differences from the evidence:
Our measure recommends only that iron repletion be verified before Epo therapy and the NCCN evidence states that iron repletion be verified before instituting Epo or darbepoetintherapy.

2b2. Validity Testing. (*Validity testing was conducted with appropriate method, scope, and adequate demonstration of validity.)*

2b2.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
Both our measure and the NCCN guidelines state that iron repletion be verified before instituting Epo or darbepoetintherapy.

The expert panel consisted of the following 10 members(with specialties listed):

Steven L. Allen, MD (Co-Chair) (hematology/oncology)
William E. Golden, MD (Co-Chair) (internal medicine (IM))
Kenneth Adler, MD (hematology/IM)
Daniel Halevy, MD (nephrology)
Stuart Henochowicz, MD, MBA (IM)
Timothy Miley, MD (hematopathology)
David Morris, MD (radiation oncology)
John M. Rainey, MD (medical oncology)
Samuel M. Silver, MD, PhD (hematology/oncology)
Lawrence Solberg, Jr., MD, PhD (hematology/IM)

2b2.2 Analytic Method *(Describe method of validity testing and rationale; if face validity, describe systematic assessment)*:
All PCPI performance measures are assessed for content validity by expert Work Group members during the development process. Additional input on the content validity of draft measures is obtained through a 30-day public comment period and by also soliciting comments from a panel of consumer, purchaser, and patient representatives convened by the PCPI specifically for this purpose. All comments received are reviewed by the expert Work Group and the measures adjusted as needed. Other external review groups (i.e. focus groups) may be convened if there are any remaining concerns related to the content validity of the measures.

Face validity has been quantitatively assessed for this measure. Specifically, the work group members were asked to empirically assess face validity of the measure. The work group/expert panel consists of 10 members, whose specialties include oncology, internal medicine, and clinical pathology.

Face validity of the measure score as an indicator of quality was systematically assessed as follows:

After the measure was fully specified, the expert panel was asked to rate their agreement with the following statement:

The scores obtained from the measure as specified will provide an accurate reflection of quality and can be used to distinguish good and poor quality.

The scale 1-5, where 1=Disagree; 3=Neither Disagree nor Agree; 5=Agree.

2b2.3 Testing Results *(Statistical results, assessment of adequacy in the context of norms for the test conducted; if face validity, describe results of systematic assessment)*:
The results of the expert panel rating of the validity statement were as follows: N = 8; Mean rating = 4.75.

Percentage in the top two categories (4 and 5): 100%

Frequency Distribution of Ratings
1 - 0
2 - 0
3 - 0
4 - 2
5 – 6

POTENTIAL THREATS TO VALIDITY. (*All potential threats to validity were appropriately tested with adequate results.*)

2b3. Measure Exclusions. (*Exclusions were supported by the clinical evidence in 1c or appropriately tested with results demonstrating the need to specify them.*)

2b3.1 Data/Sample for analysis of exclusions *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
PCPI Testing Project

- 41 Myelodysplastic Syndrome patient records were reviewed for this measure.
- The measurement period (data collected from patients seen) was between 1/1/2009 through 12/31/2009.
- Chart auditing was performed between 5/17/2010 and 7/15/2010.
- Data auditing was performed between 8/2/2010 and 9/14/2010.

2b3.2 Analytic Method (*Describe type of analysis and rationale for examining exclusions, including exclusion related to patient preference)*:
Exceptions were analyzed for frequency and variability across providers.

2b3.3 Results *(Provide statistical results for analysis of exclusions, e.g., frequency, variability, sensitivity analyses):*
N, % Agreement, Kappa (95% Confidence Interval)
Exceptions Reliability: 41, 100.0%, Kappa is non-calculable*

This measure demonstrates perfect reliability, as shown in results from the above analysis.

The exception rate for this measure was 2.4%.

*Kappa Statistics cannot be calculated because of complete agreement. Confidence intervals cannot be calculated because to do so would involve dividing by zero which cannot be done.

2b4. Risk Adjustment Strategy. (*For outcome measures, adjustment for differences in case mix (severity) across measured entities was appropriately tested with adequate results.*)

2b4.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
This measure is not risk adjusted.

2b4.2 Analytic Method (*Describe methods and rationale for development and testing of risk model or risk stratification including*

selection of factors/variables):
This measure is not risk adjusted.

2b4.3 Testing Results *(Statistical risk model: Provide quantitative assessment of relative contribution of model risk factors; risk model performance metrics including cross-validation discrimination and calibration statistics, calibration curve and risk decile plot, and assessment of adequacy in the context of norms for risk models. Risk stratification: Provide quantitative assessment of relationship of risk factors to the outcome and differences in outcomes among the strata)*:
Not applicable.

2b4.4 If outcome or resource use measure is not risk adjusted, provide rationale and analyses to justify lack of adjustment:

2b5. Identification of Meaningful Differences in Performance. (*The performance measure scores were appropriately analyzed and discriminated meaningful differences in quality.*)

2b5.1 Data/Sample *(Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
CMS Physician Quality Reporting Initiative:
Clinical Condition and Measure: #68
11,494 patients were reported on for the 2008 program, the most recent year for which data are available.

In 2009 the following was reported for this measure:
Eligible Professionals: 21,607
Professionals Reporting >=1 Valid QDC: 1,235
% Professionals Reporting >=1 Valid QDC: 5.72%
# Professionals Satisfactorily Reporting: 452
% Professionals Satisfactorily Reporting: 36.60%

2b5.2 Analytic Method *(Describe methods and rationale to identify statistically significant and practically/meaningfully differences in performance)*:
CMS Physician Quality Reporting Initiative:
The inter-quartile range (IQR) was calculated, which provides a measure of the dispersion of performance.

2b5.3 Results *(Provide measure performance results/scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance)*:
CMS Physician Quality Reporting Initiative
This measure was used in the 2007-2011 CMS Physician Quality Reporting Initiative claims and registry options and group reporting option available in 2011.

There is a gap in care as shown by this 2008 data, the only year for which distribution by quartile/decile is available.

58.00% of patients reported on did not meet the measure.

10th percentile: 0.00%
25th percentile: 6.91%
50th percentile: 30.22%
75th percentile: 66.67%
90th percentile: 97.44%

The inter-quartile range (IQR) provides a measure of the dispersion of performance. The IQR is 59.76, and indicates that 50% of physicians have performance on this measure ranging from 6.91% and 66.67%. A quarter of reporting physicians have performance on this measure which is greater than 66.67%, while a quarter have performance on this measure less than 6.91%.

2b6. Comparability of Multiple Data Sources/Methods. (*If specified for more than one data source, the various approaches*

result in comparable scores.)

2b6.1 Data/Sample *(Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
This measure has not been compared across data sources.

2b6.2 Analytic Method *(Describe methods and rationale for testing comparability of scores produced by the different data sources specified in the measure)*:
This measure has not been compared across data sources

2b6.3 Testing Results *(Provide statistical results, e.g., correlation statistics, comparison of rankings; assessment of adequacy in the context of norms for the test conducted)*:
This measure has not been compared across data sources.

2c. Disparities in Care: H☐ M☐ L☐ I ☐ NA☐ (*If applicable, the measure specifications allow identification of disparities.*)

2c.1 If measure is stratified for disparities, provide stratified results *(Scores by stratified categories/cohorts)*: We encourage the results of this measure to be stratified by race, ethnicity, gender, and primary language, and have included these variables as recommended data elements to be collected.

2c.2 If disparities have been reported/identified (e.g., in 1b), but measure is not specified to detect disparities, please explain:
The PCPI advocates that performance measure data should, where possible, be stratified by race, ethnicity, and primary language to assess disparities and initiate subsequent quality improvement activities addressing identified disparities, consistent with recent national efforts to standardize the collection of race and ethnicity data. A 2008 NQF report endorsed 45 practices including stratification by the aforementioned variables.(1) A 2009 IOM report "recommends collection of the existing Office of Management and Budget (OMB) race and Hispanic ethnicity categories as well as more fine-grained categories of ethnicity(referred to as granular ethnicity and based on one's ancestry) and language need (a rating of spoken English language proficiency of less than very well and one's preferred language for health-related encounters)."(2)

References:
(1)National Quality Forum Issue Brief (No.10). Closing the Disparities Gap in Healthcare Quality with Performance Measurement and Public Reporting. Washington, DC: NQF, August 2008.

(2)Race, Ethnicity, and Language Data: Standardization for Health Care Quality Improvement. March 2010. AHRQ Publication No. 10-0058-EF. Agency for Healthcare Research and Quality, Rockville, MD. Available at: http://www.ahrq.gov/research/iomracereport. Accessed May 25, 2010.

2.1-2.3 Supplemental Testing Methodology Information:

Steering Committee: Overall, was the criterion, *Scientific Acceptability of Measure Properties*, met?
(*Reliability and Validity must be rated moderate or high*) Yes☐ No☐
Provide rationale based on specific subcriteria:

If the Committee votes No, STOP

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)

C.1 Intended Purpose/ Use *(Check all the purposes and/or uses for which the measure is intended)*: Professional Certification or Recognition Program, Public Reporting, Quality Improvement (Internal to the specific organization)

3.1 Current Use *(Check all that apply; for any that are checked, provide the specific program information in the following questions)*: Public Reporting, Professional Certification or Recognition Program, Quality Improvement (Internal to the specific organization)

3a. Usefulness for Public Reporting: H☐ M☐ L☐ I ☐
(*The measure is meaningful, understandable and useful for public reporting.*)

3a.1. Use in Public Reporting - disclosure of performance results to the public at large *(If used in a public reporting program, provide name of program(s), locations, Web page URL(s)).* If not publicly reported in a national or community program, state the reason AND plans to achieve public reporting, potential reporting programs or commitments, and timeline, e.g., within 3 years of endorsement: [*For Maintenance – If not publicly reported, describe progress made toward achieving disclosure of performance results to the public at large and expected date for public reporting; provide rationale why continued endorsement should be considered.*]
The measure has been in use in CMS PQRS program since 2007. The mean performance rate for 2009 was reported as 97.69% with a total of 1,269 eligible professionals submitting. demonstrating an opportunity for improvement. Unfortunately, data regarding the variability in performance rates across reporting eligible professionals for PQRS 2009 is not available at this time.

The PCPI believes that the reporting of participation information is a beneficial first step on a trajectory toward the public reporting of performance results, which is appropriate since the measure has been tested and the reliability of the performance data has been validated. Continued NQF endorsement will facilitate our ongoing progress toward this public reporting objective.

3a.2.Provide a rationale for why the measure performance results are meaningful, understandable, and useful for public reporting. If usefulness was demonstrated (e.g., focus group, cognitive testing), describe the data, method, and results: The PCPI believes that the reporting of participation information is a beneficial first step on a trajectory toward the public reporting of performance results, which is appropriate since the measure has been tested and the reliability of the performance data has been validated. Continued NQF endorsement will facilitate our ongoing progress toward this public reporting objective.

3.2 Use for other Accountability Functions (payment, certification, accreditation). If used in a public accountability program, provide name of program(s), locations, Web page URL(s): This hematology quality measure is used in the ASH Chart Abstraction tools/Performance Improvement Modules (PIMs), which is used by ASH members to complete the ABIM's Self Directed PIM for Maintenance of Certification, Part IV. The measure is used as described below:

MDS Measure 68: Percentage of patients aged 18 years and older with a diagnosis of MDS who are receiving erythropoietin therapy with documentation of iron stores prior to initiating erythropoietin therapy.

MDS Chart Abstraction Tool Question 9: Did you determine that your patient had adequate iron stores for treating MDS with erythropoietin/darbepoietin?

Adequate iron stores is determined by EITHER bone marrow examination, including iron stain OR peripheral blood iron showing ferritin >100 ng/ml and/or the ratio of serum iron to TIBC >20%.

3b. Usefulness for Quality Improvement: H☐ M☐ L☐ I ☐
(*The measure is meaningful, understandable and useful for quality improvement.*)

3b.1. Use in QI. If used in quality improvement program, provide name of program(s), locations, Web page URL(s):
[*For Maintenance – If not used for QI, indicate the reasons and describe progress toward using performance results for improvement*].
All PCPI measures are suitable for use in quality improvement initiatives and are made freely available on the PCPI website and through the implementation efforts of medical specialty societies and other PCPI members. The PCPI strongly encourages the use of its measures in QI initiatives and seeks to provide information on such initiatives to PCPI members.

3b.2. Provide rationale for why the measure performance results are meaningful, understandable, and useful for quality improvement. If usefulness was demonstrated *(e.g., QI initiative),* describe the data, method and results:

Overall, to what extent was the criterion, *Usability*, met? H☐ M☐ L☐ I ☐
Provide rationale based on specific subcriteria:

4. FEASIBILITY

Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. (evaluation criteria)

4a. Data Generated as a Byproduct of Care Processes: H☐ M☐ L☐ I ☐

4a.1-2 How are the data elements needed to compute measure scores generated? *(Check all that apply).*
Data used in the measure are:
generated by and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition

4b. Electronic Sources: H☐ M☐ L☐ I ☐

4b.1 Are the data elements needed for the measure as specified available electronically *(Elements that are needed to compute measure scores are in defined, computer-readable fields)*: ALL data elements in electronic health records (EHRs)

4b.2 If ALL data elements are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources:

4c. Susceptibility to Inaccuracies, Errors, or Unintended Consequences: H☐ M☐ L☐ I ☐

4c.1 Identify susceptibility to inaccuracies, errors, or unintended consequences of the measurement identified during testing and/or operational use and strategies to prevent, minimize, or detect. If audited, provide results:
We are not aware of any unintended consequences related to this measurement.

4d. Data Collection Strategy/Implementation: H☐ M☐ L☐ I ☐

A.2 Please check if either of the following apply (*regarding proprietary measures*):
4d.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 and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues *(e.g., fees for use of proprietary measures)*:
This measure has not been compared across data sources.

Overall, to what extent was the criterion, *Feasibility*, met? H☐ M☐ L☐ I ☐
Provide rationale based on specific subcriteria:

OVERALL SUITABILITY FOR ENDORSEMENT

Does the measure meet all the NQF criteria for endorsement? Yes☐ No☐
Rationale:

If the Committee votes No, STOP.
If the Committee votes Yes, the final recommendation is contingent on comparison to related and competing measures.

5. COMPARISON TO RELATED AND COMPETING MEASURES

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

5.1 If there are related measures *(either same measure focus or target population)* or competing measures *(both the same measure focus and same target population*), list the NQF # and title of all related and/or competing measures:

5a. Harmonization
5a.1 If this measure has EITHER the same measure focus OR the same target population as NQF-endorsed measure(s): Are the measure specifications completely harmonized? 5a.2 If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden:
5b. Competing Measure(s)
5b.1 If this measure has both the same measure focus and the same target population as NQF-endorsed measure(s): Describe why this measure is superior to competing measures *(e.g., a more valid or efficient way to measure quality);* OR provide a rationale for the additive value of endorsing an additional measure. *(Provide analyses when possible)*: No related or competing measures.

CONTACT INFORMATION
Co.1 Measure Steward (Intellectual Property Owner): American Medical Association - Physician Consortium for Performance Improvement, 515 N State Street, Chicago, Illinois, 60654 Co.2 Point of Contact: Mark S., Antman, DDS, MBA, Director, Measure Development Operations Performance Improvement, mark.antman@ama-assn.org, 312-464-5056-
Co.3 Measure Developer if different from Measure Steward: American Medical Association - Physician Consortium for Performance Improvement, 515 N State Street, Chicago, Illinois, 60654 Co.4 Point of Contact: Mark, Antman, mark.antman@ama-assn.org, 312-464-5056-
Co.5 Submitter: Molly, Siegel, molly.siegel@ama-assn.org, 312-464-4901-, American Medical Association - Physician Consortium for Performance Improvement
Co.6 Additional organizations that sponsored/participated in measure development: American Society of Hematology
Co.7 Public Contact: Mark, Antman, mark.antman@ama-assn.org, 312-464-5056-, American Medical Association - Physician Consortium for Performance Improvement

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. Steven L. Allen, MD (Co-Chair) (hematology/oncology) William E. Golden, MD (Co-Chair) (internal medicine (IM)) Kenneth Adler, MD (hematology/IM) Daniel Halevy, MD (nephrology) Stuart Henochowicz, MD, MBA (IM) Timothy Miley, MD (hematopathology) David Morris, MD (radiation oncology) John M. Rainey, MD (medical oncology) Samuel M. Silver, MD, PhD (hematology/oncology) Lawrence Solberg, Jr., MD, PhD (hematology/IM) PCPI measures are developed through cross-specialty, multi-disciplinary work groups. All medical specialties and other health care professional disciplines participating in patient care for the clinical condition or topic under study must be equal contributors to the

measure development process. In addition, the PCPI strives to include on its work groups individuals representing the perspectives of patients, consumers, private health plans, and employers. This broad-based approach to measure development ensures buy-in on the measures from all stakeholders and minimizes bias toward any individual specialty or stakeholder group. All work groups have at least two co-chairs who have relevant clinical and/or measure development expertise and who are responsible for ensuring that consensus is achieved and that all perspectives are voiced.
Ad.2 If adapted, provide title of original measure, NQF # if endorsed, and measure steward. Briefly describe the reasons for adapting the original measure and any work with the original measure steward:
Measure Developer/Steward Updates and Ongoing Maintenance Ad.3 Year the measure was first released: 2007 Ad.4 Month and Year of most recent revision: 09, 2010 Ad.5 What is your frequency for review/update of this measure? Coding/Specifications updates occur annually. See additional information below: Ad.6 When is the next scheduled review/update for this measure? 12, 2012
Ad.7 Copyright statement:
Ad.8 Disclaimers:
Ad.9 Additional Information/Comments: Coding/Specifications updates occur annually. The PCPI has a formal measurement review process that stipulates regular (usually on a three-year cycle, when feasible) review of the measures. The process can also be activated if there is a major change in scientific evidence, results from testing or other issues are noted that materially affect the integrity of the measure.
Date of Submission (*MM/DD/YY*): 10/03/2011