TO: Peter Crooks, MD; Jeffrey Berns, MD; Michael Fischer, MD
Joseph Nally, MD; Andrew Narva, MD; Greg Miller, PhD
Debra Hain, PhD, PARN, ANP-BC, GNP-BC

FROM: Karen Pace, PhD.
Senior Director, Measures Maintenance

SUBJECT: Briefing Materials for Ad-hoc Review: Measurement of Serum Phosphorus Concentration (0255)

DATE: September 23, 2012

We are grateful to each of you for your willingness to serve as the expert panel for the ad hoc review of **NQF Measure 0255**-Measurement of Serum Phosphorus Concentration. This measure was submitted and evaluated in the 2011 Renal Endorsement Maintenance Project.

The purpose of this memo is to provide the charge to this Expert Panel and all materials related to the ad hoc review of measure 0255. This ad hoc review has a Project Page on our Web site to keep the public and NQF membership informed as the review moves forward.

My colleagues, Elisa Munthali & Ashley Morsell, and I are available to answer your questions. We can be reached at measuremaintenance@qualityforum.org or 202.783.1300.

Actions Needed

- Review this briefing memo, key questions, and all materials.
- Identify if you have any questions to be directed to either the requestor of the ad hoc review (KCP) or the measure steward (CMS) and notify NQF staff prior to the call.
- Notify NQF staff if you are aware of other sources of evidence or other materials that we should obtain for the Expert Panel's consideration.
- Participate in the conference call to make a recommendation regarding the requested change to measure 0255.

Materials Included in this Memo and Attachments

- NQF Ad Hoc Review Criteria and Process
- Related Clinical Practice Guidelines
- NQF Evidence Subcriterion and Guidance for Evaluation

Attachments

- KCP Request for Ad hoc Review and Accompanying Material
- CMS Technical Expert Panel Review and Recommendation
- 2011 Measure submission and evaluation of 0255

Ad Hoc Review

The Ad Hoc Review criteria and process are stated in NQF's Consensus Standards Maintenance and Endorsement Cycle Process.

Criteria

An ad hoc review may be conducted on an endorsed measure at any time if one or more of the following criteria are met:

- the evidence supporting the measure, practice or event has changed and it no longer reflects updated evidence,
- there is evidence that implementation of the measure or practice may result in unintended consequences:
  a. Use of the measure or practice may result in inappropriate or harmful care
  b. Measure performance scores may yield invalid conclusions about quality of care (e.g., misclassification or incorrect representation of quality)
- material changes have been made to a currently endorsed measure.

Process

- NQF receives a request for an ad hoc review. NQF staff conducts an initial review of the request to determine if a review is justified, including communication with the measure steward.
- A notice of ad hoc review is posted to the NQF web site. NQF will solicit technical experts from previously convened committees and if necessary a call for nominations for experts will be open for no less than 10 business days. The provisional slate of technical advisors will be posted for comment for no less than 10 business days.
- The selected technical advisors will review the evidence and provide input to the Consensus Standards Approval Committee (CSAC). The ad hoc review requestor and the measure steward are given the opportunity to provide information to the technical advisors and CSAC.
- Review and comment period for the draft recommendations from the technical advisors will be posted for no less than 10 business days.
- The information is forwarded to the CSAC (including the assessment of the technical advisors, public and member comments, and input from the measure steward and requester) and CSAC makes a decision on endorsement status and/or specification changes.
- The CSAC decision is forwarded to the NQF Board of Directors for ratification.
- There is a 30-day appeals period.
- The measure, practice, or event is still subject to review in its designated maintenance cycle

Reason for this Request

Kidney Care Partners (KCP) has requested an ad hoc review, asserting that implementation of the measure will result in unintended consequences.

The measure as endorsed calls for monthly measurement of serum phosphorus and the request is to amend the measure specifications to allow for either serum or plasma phosphorus. KCP contends that testing plasma phosphorus concentration is an acceptable alternative to serum phosphorus therefore, performance scores of facilities using plasma testing could inappropriately indicate lower quality.

Although KCP cited unintended consequences as the reason for the request, their position is based on the assertion that plasma testing of phosphorus is an acceptable alternative to serum phosphorus. Therefore, the ad hoc review will necessarily focus on the evidence in regards to plasma testing of

phosphorus and NQF's measure evaluation criteria and guidance for evaluating the evidence for a performance measure will apply.

We are looking to you for an assessment of whether the evidence for the requested change to the measure to include plasma phosphorus testing meets NQF's Measure Evaluation Criteria.

Key Questions for Expert Panel

- Does the evidence for plasma testing of phosphorus meet NQF guidance for quantity, quality, and consistency? If not, should an exception to the evidence criterion be considered, and if so, why? (*see criteria and guidance below*)
- What is the evidence that supports the current guidelines that specify serum phosphorus testing? (*see next section to review guidelines and supporting evidence*)

Related Clinical Practice Guidelines

Measure 0255 is consistent with current clinical practice guidelines. Click on the links to view the full guidelines and supporting evidence.

The guidelines developed by the National Kidney Foundation-Kidney Disease Outcomes Quality Initiative, ***KDOQI Clinical Practice Guidelines for Bone Metabolism and Disease in Chronic Kidney Disease***, call for serum phosphorus testing**.**
1.1 **Serum** levels of calcium, phosphorus, and intact plasma parathyroid hormone (PTH) should be measured in all patients with CKD and GFR <60 mL/min/1.73 m^2. (EVIDENCE).

The international guidelines from Kidney Disease Improving Global Outcome, ***KDIGO Guideline for Chronic Kidney Disease - Mineral and Bone Disorder***, also calls for serum levels.
3.1.1. We recommend monitoring **serum** levels of calcium, phosphorus, PTH, and alkaline phosphatase activity beginning in CKD stage 3 (1C).

Key Questions

- Why do the guideline recommendations specify serum testing?
- Does the evidence for the guideline recommendations address serum vs. plasma testing?

NQF's Evidence Subcriterion and Guidance

Following are the evidence subcriterion and guidance for evaluating the evidence supporting a performance measure.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.

- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- **Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.**
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

5. 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 multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

Evaluation of Quantity, Quality, and Consistency of Body of Evidence for Structure, Process, and Intermediate Outcome Measures

DEFINITION/ RATING	**QUANTITY** OF BODY OF EVIDENCE	**QUALITY** OF BODY OF EVIDENCE	**CONSISTENCY** OF RESULTS OF BODY OF EVIDENCE
Definition	Total number of studies (not articles or papers)	Certainty or confidence in the estimates of benefits and harms to patients across studies in the body of evidence related to study factors[a] including: study design or flaws; directness/indirectness to the specific measure (regarding the population, intervention, comparators, outcomes); imprecision (wide confidence intervals due to few patients or events)	Stability in both the direction and magnitude of clinically/practically meaningful benefits and harms to patients (benefit over harms) across studies in the body of evidence
High	5+ studies[b]	Randomized controlled trials (RCTs) providing direct evidence for the specific measure focus, with adequate size to obtain precise estimates of effect, and without serious flaws that introduce bias	Estimates of clinically/practically meaningful benefits and harms to patients are consistent in direction and similar in magnitude across the preponderance of studies in the body of evidence

DEFINITION/ RATING	**QUANTITY** OF BODY OF EVIDENCE	**QUALITY** OF BODY OF EVIDENCE	**CONSISTENCY** OF RESULTS OF BODY OF EVIDENCE
Moderate	2-4 studies[b]	• Non-RCTs with control for confounders that could account for other plausible explanations, with large, precise estimate of effect OR • RCTs without serious flaws that introduce bias, but with either indirect evidence or imprecise estimate of effect	Estimates of clinically/practically meaningful benefits and harms to patients are consistent in direction across the preponderance of studies in the body of evidence, but may differ in magnitude If only one study, then the estimate of benefits greatly outweighs the estimate of potential harms to patients (one study cannot achieve high consistency rating)
Low	1 study[b]	• RCTs with flaws that introduce bias OR • Non-RCTs with small or imprecise estimate of effect, or without control for confounders that could account for other plausible explanations	• Estimates of clinically/practically meaningful benefits and harms to patients differ in both direction and magnitude across the preponderance of studies in the body of evidence OR • wide confidence intervals prevent estimating net benefit If only one study, then estimate of benefits do not greatly outweigh harms to patients
Insufficient to Evaluate *(See Table 3 for exceptions.)*	• No empirical evidence OR • Only selected studies from a larger body of evidence	• No empirical evidence OR • Only selected studies from a larger body of evidence	No assessment of magnitude and direction of benefits and harms to patients

[a]*Study designs* that affect certainty of confidence in estimates of effect include: randomized controlled trials (RCTs), which control for both observed and unobserved confounders, and non-RCTs (observational studies) with various levels of control for confounders.
Study flaws that may bias estimates of effect include: lack of allocation concealment; lack of blinding; large losses to follow-up; failure to adhere to intention to treat analysis; stopping early for benefit; and failure to report important outcomes.
Imprecision with wide confidence intervals around estimates of effects can occur in studies involving few patients and few events.
Indirectness of evidence includes: indirect comparisons (e.g., two drugs compared to placebos rather than head-to head); and differences between the population, intervention, comparator interventions, and outcome of interest and those included in the relevant studies.[15]
[b]The suggested number of studies for rating levels of quantity is considered a general guideline.

Evaluation of Subcriterion 1a Based on the Quantity, Quality, and Consistency of the Body of Evidence

QUANTITY OF BODY OF EVIDENCE	**QUALITY** OF BODY OF EVIDENCE	**CONSISTENCY** OF RESULTS OF BODY OF EVIDENCE	**PASS SUBCRITERION 1A**
Moderate-High	Moderate-High	Moderate-High	Yes
Low	Moderate-High	Moderate (if only one study, high consistency not possible)	Yes, but only if it is judged that additional research is unlikely to change conclusion that benefits to patients outweigh harms; otherwise, No
Moderate-High	Low	Moderate-High	Yes, but only if it is judged that potential benefits to patients clearly outweigh potential harms; otherwise, No
Low-Moderate-High	Low-Moderate-High	Low	No
Low	Low	Low	No
Exception to Empirical Body of Evidence for Health Outcome For a health outcome measure: A rationale supports the relationship of the health outcome to at least one healthcare structure, process, intervention, or service			Yes, if it is judged that the rationale supports the relationship of the health outcome to at least one healthcare structure, process, intervention, or service
Potential Exception to Empirical Body of Evidence for Other Types of Measures If there is no empirical evidence, expert opinion is systematically assessed with agreement that the benefits to patients greatly outweigh potential harms.			Yes, but only if it is judged that potential benefits to patients clearly outweigh potential harms; otherwise, No

Measure 0255

The measure evaluation summary and measure submission for measure #0255 are attached to this document. In 2011, the Steering Committee noted the lack of ***direct*** empirical evidence to support the performance measure, but also noted the lack of evidence to support a performance measure on intervention or intermediate clinical outcome of phosphorus levels. The Committee, however, did think it was important to monitor phosphorus levels. This would be consistent with considering an exception to the evidence criterion.

KIDNEY CARE PARTNERS

December 12, 2012

Gerald M. Shea
Interim Chief Executive Officer
National Quality Forum
1030 15th Street, NW
Suite 800
Washington, DC 20005

Helen Burstin, MD, MPH
Senior Vice President, Performance Measures
National Quality Forum
1030 15th Street, NW
Suite 800
Washington, DC 20005

Dear Mr. Shea and Dr. Burstin:

Kidney Care Partners (KCP) is an alliance of members of the kidney care community that serves as a forum for patient advocates, dialysis care professionals, providers, and manufacturers to advance policies that support the provision of high quality care for individuals with both chronic kidney disease and End-Stage Renal Disease (ESRD). An NQF member, we have been active participants in NQF's projects on ESRD quality and commend you for your work in this regard.

On the basis of Criterion 2, "implementation of the measure results in unintended consequences," KCP requests that NQF undertake an ad hoc review of *NQF 0255: Measurement of Serum Phosphorus Concentration*.

ID/Title/Steward	Description	Numerator	Denominator	Exclusions
NQF 0255 **Measurement of Serum Phosphorus Concentration** Steward: CMS Level: Facility	Percentage of all adult (≥ 18 years old) HD and PD patients with serum phosphorus measured at least once within the month.	Number of adult (≥ 18 years old) dialysis patients included in the denominator with serum phosphorus measured at least once within month.	All adult PD and HD patients included in the sample for analysis. Adjustment: 1. Transient dialysis patients (in unit <30 days). 2. Pediatric patients. 3. Kidney transplant recipients with a functioning graft.	1. Transient dialysis patients (in unit <30 days). 2. Pediatric patients. 3. Kidney transplant recipients with a functioning graft.

Specifically, KCP requests that NQF seek a change to the measure specifications by the developer, the Centers for Medicare and Medicaid Services (CMS), to permit measurement of

serum *or plasma* phosphorus concentration as equivalent assays. Absent this modification, KCP requests the withdrawal of NQF endorsement of a serum-only measure.

Implementation of NQF 0255

On November 2, 2012, CMS issued its Final Rule related to the *End-Stage Renal Disease Prospective Payment System, Quality Incentive Program (QIP), and Bad Debt Reductions* (77 FR 67449). Under the Final Rule, CMS finalizes adoption of a bone/mineral metabolism measure for Payment Year (PY) 2014 (i.e., measurement/implementation year 2013) requiring facilities to *attest that they have monitored each of their Medicare patient's phosphorus and calcium levels monthly throughout the performance period*. The Final Rule further notes *the NQF has previously endorsed phosphorus and calcium monitoring measures (#0261 and #0255) upon which this measure is based. NQF has since withdrawn its endorsement of the calcium measure* [#0261].

Unintended Consequence of Implementation of NQF 0255 as Specified

Under section 153(c) of MIPPA, which amended section 1881(h) to the Social Security Act, the Secretary must establish an ESRD QIP by: (1) selecting measures; (2) establishing the performance standards that apply to the individual measures; (3) specifying a performance period with respect to a year; (4) developing a methodology for assessing the total performance of each facility based on the performance standards with respect to the measures for a performance period; and (5) **applying an appropriate payment reduction to facilities that do not meet or exceed the established Total Performance Score**.

In short, dialysis facilities are subject to the nation's first and only penalty-based quality performance program. As such, the implementation of NQF 0255 as a component of one of five measures for PY 2014 (measurement/implementation year CY 2013) could have significant, negative financial consequences if facilities are unable to attest to monthly *serum* phosphorus measurements (because these facilities use plasma as the substrate) and so fail to perform well on the measure.

Equivalency of Serum and Plasma Assays for Phosphorus Measurement

KCP posits that the specifications for NQF 0255 need to accommodate industry-accepted standard measurement of both serum and plasma; to accept only serum, as the measure specifies, means the regulations prefer a specific testing substrate upon implementation, despite evidence of the equivalency of serum and plasma for phosphorus measurement. Without the change, facilities that use laboratories that deploy plasma testing will be inappropriately and unfairly disadvantaged in their performance scores and will face an increased financial penalty as a consequence.

We are aware that at least one renal laboratory, Ascend Clinical, has been using plasma testing since 2006. Others (e.g., Spectra Laboratories) are considering it because it is more patient-centered, requiring less blood. Additionally, plasma is more stable and requires less manipulation should additional testing be required. Serum and plasma testing have been validated for most clinical chemistry analyzers, with both deemed acceptable by analyzer manufacturers.

As noted in the additional attached documentation provided by a KCP member (attachment A), there was virtually no difference between phosphorus measured in serum vs. plasma: a difference of 0.01 mg/dL; phosphorus values are reported to the nearest 0.1 mg/dL. And while the documentation acknowledges some reports of reported differences in serum phosphorus vs. plasma measurement, it also points out: i) such differences are within the College of American Pathologists total allowable error; and ii) such differences could not be replicated by two large experiments conducted by Spectra Laboratories.

Summary and Requested Action

We appreciate that NQF cannot compel CMS to change the specifications. Based on correspondence from Dr. Patrick Conway, CMS Chief Medical Officer, however, the Agency seems willing to make changes in the context of NQF review and endorsement, but points to a future annual maintenance cycle as the mechanism (attachment B). With implementation scheduled for 2013, however, KCP believes the consequences of waiting for such consideration are untenable and should therefore be conducted as an ad hoc review based on Criterion 2.

We look forward to working with you on this important matter. Please do not hesitate to contact Kathy Lester (klester@pattonboggs.com, 202.447.6562) and Lisa McGonigal, MD, MPH (lmcgon@msn.com, 203.530.9524) with any questions or concerns regarding this request.

Sincerely,

Linda DeRuvo-Keegan
Executive Director

cc: Heidi Bossley, MSN, MBA
NQF

Patrick Conway, MD, MSc
CMS

Jean Moody-Williams, RN, MPP
CMS

FRESENIUS MEDICAL CARE

Patrick Conway, M.D.
Director and Chief Medical Officer
Office of Clinical Standards and Quality
Centers for Medicare and Medicaid Services
U.S. Department of Health and Human Services
7500 Security Boulevard
Baltimore, MD 21244

October 8, 2012

Dear Dr. Conway:

We are sending this correspondence in follow up to the issue we discussed with you on September 25 regarding the need for certain components of the ESRD Prospective Payment System (PPS) and Quality Incentive Program (QIP) proposed rule that pertain to laboratory measurement for calcium and phosphorus to be modified in the final rule. Such changes are necessary to accommodate industry accepted standard measurements of both serum and plasma, and to ensure that the regulations do not indicate a preference for a particular testing method upon implementation. Currently, the language in the proposed rule (QIP section III.B.3) specifically states "*serum* calcium and *serum* phosphorus" (emphasis added).

This issue was first raised by members of the Kidney Care Council at a meeting with Jean, Teresa Casey, and others to discuss recommendations related to the ESRD QIP issues. That meeting was held a week before the Kidney Care Partners meeting with you, Jean, Dr. Shari Ling, and Brenda Gentle.

We are aware of at least one renal laboratory, Ascend Clinical (www.ascendclinical/about-us/milestone) that has been using plasma testing since 2006. We have begun to explore the use of plasma testing in our clinical laboratory, Spectra Laboratories, primarily because it requires less blood, it is more stable (i.e. there is no clotting required), and there are fewer manipulations on residual blood (e.g. spinning the sample), should there be a need to repeat a test or perform additional testing.

It is important to establish that both serum and plasma samples have been validated for most clinical chemistry analyzers, with both specified as acceptable by the analyzer manufacturers. As far as the measurement process is concerned, it makes no difference whether serum or plasma samples are used. When we studied the differences between serum and plasma samples for calcium and phosphorus testing on a group of ESRD patients, our findings were generally consistent with those of published studies. Specific results for calcium and phosphorus are included in the Appendix.

Some calcium is consumed in the clotting process (i.e., for serum). We observed a difference of about 0.13 mg/dL (plasma higher than serum). This difference is minor when compared to the Clinical Laboratory Improvement Act (CLIA) specification of 1.0 mg/dL allowable error for proficiency testing. This CLIA limit is the minimum standard for performance that is adhered to by instrument manufacturers, and sets the level of current practice in most laboratories.

Most instruments are not capable of distinguishing calcium differences of the order of 0.1 mg/dL, due to their inherent imprecision, without testing samples in large numbers of replicates. Calcium is customarily reported to the nearest 0.1 mg/dL, further minimizing the practical impact of this difference. To place it in the proper context, in the study we conducted comparing calcium testing results for the same pooled samples sent to 8 renal laboratories (copy included with this letter), the mean bias was 0.09 mg/dL and the range of the difference between the highest mean from a laboratory and the lowest mean was 0.51 mg/dL Since the difference, though small, is positive, the practical impact to the provider under the proposed rule could be a slight increase in the proportion of hypercalcemic patients. However, within the context of the requirement for a 3-month rolling average to determine hypercalcemia, the impact of this difference between serum and plasma measurements should be attenuated further.

For phosphorus, there were almost no measurable differences in Spectra's tests. The average difference was less than 0.01 mg/dL, which is negligible considering that phosphorus is customarily reported to the nearest 0.1 mg/dL. However, we have noted in some publications that authors have found serum values can be higher than plasma by up to 0.2-0.3 mg/dL, if phosphorus is released by cells during clotting. This difference is within the College of American Pathologists total allowable error (CLIA specifications do not address phosphorus except as three times the peer group standard deviation). In two large experiments (N=101 and N=129 paired samples), Spectra was unable to reproduce similar differences in dialysis patients. Furthermore, although not reproduced in Spectra's hands, the potential difference from the lab variability study comparing pooled ESRD serum tested by 8 renal laboratories for phosphorus indicated a mean bias of 0.09 mg/dL with the greatest potential range of the difference between the highest mean from a laboratory and the lowest mean was 0.39 mg/dL

We realize that the official comment period for the proposed rule for the Quality Incentive Program has lapsed. However, we believe that one value of working with the community of professionals and stakeholders is that we are able to bring to your attention issues such as this one, for your consideration in the final rule.

We further understand that NQF approved the measures for phosphorous and calcium linking them to serum, and CMS has adopted this NQF endorsed measure for the QIP PY2014 and beyond. In our review of the NQF process, if the measures are still endorsed by NQF, it would likely accept a change of this type (recognizing the switch in testing media that does not result in change in the values) as a non-material change. The measure owner, which we understand is CMS, would need to file an update to the measures and NQF would make that change. While the

community would support this change, we do not own the measure and cannot initiate the process. In terms of the QIP, we believe the issue requires a change through rulemaking because CMS adopted the mineral metabolism structural (reporting) measure as a combination of the NQF measures.

To resolve the problem, CMS should: (1) as the owner of the measures, file the update to the measures with NQF if these measures remain endorsed by NQF at this time; (2) state in the final QIP rule that the Agency has taken this step with the NQF and that it is finalizing the mineral metabolism reporting measure to allow facilities to report values from calcium and phosphorous tests that were done using either serum or plasma as the medium, and (3) that CMS use both plasma and serum, or other broad language, in future rulemaking and clinical performance measure development that would allow for future inclusion of new methods and technologies.

What appears to be a minor change, such as the measurement specifications for calcium and phosphorus, may actually have repercussions that resonate beyond the QIP and into CrownWeb, as well as outside the world of ESRD, because these measurements are also utilized interchangeably by laboratories that service physician offices, clinics and hospitals. Finally, the interchangeability of plasma and serum measurements actually extend to other chemistry analytes, including albumin, ferritin, and others, such that the implications can be fairly significant for future rulemaking.

In summary, we are bringing to your attention a specific urgent need to modify measurement requirements for calcium and phosphorus, to include both serum and plasma measurements, as the Agency finalizes the proposed rule for the payment year 2015 ESRD QIP. We offer further assistance in discussing specific details and/or more technical aspects of this issue with your team, including supplementary information about other analytes, as well as to discuss any issues or concerns you may have. We believe that this pre-emptive effort may avoid locking CMS into static techniques and methods, whereas the field has continued to evolve, and that addressing the issue at this time will avert future issues and problems.

Thank you once again for fostering this spirit of collaboration in ESRD care.

Sincerely,

Franklin W. Maddux, MD, FACP
Chief Medical Officer and EVP, Clinical & Scientific Affairs
Fresenius Medical Care, North America

Eduardo Lacson, Jr., MD, MPH, FACP, FASN
VP, Clinical Science, Epidemiology, and Research
Fresenius Medical Care, North America

Appendix

Additional specific information based on laboratory validation studies comparing matched plasma sample testing (PST) and serum sample testing (SST) performed by Spectra laboratories.

Phosphorus

- CLIA allowable total error = n/a
- CAP allowable error = 0.3 mg/dL or 10.7% (higher)
- Fraser "desirable bias" = 3.2%
- Target concentrations : 3.0, 3.5, 5.5 mg/dL
- Target biases: 0.10, 0.11, 0.18 mg/dL, respectively
- Analytical SD approx. 0.08 mg/dL @ target
- **Bias (PST-SST, two experiments)**
    - **+0.01 mg/dL @ patient mean of 4.65 mg/dL**
    - **-0.01 mg/dL @ patient mean of 5.49 mg/dL**

Calcium

- CLIA allowable total error = 1.0 mg/dL
- Fraser "desirable bias" = 0.8%
- Target concentrations : 8.5, 9.5, 10.2 mg/dL
- Target biases: 0.068, 0.076, 0.082 mg/dL, respectively (very small relative to CLIA allowable error)
- Analytical SD approx. 0.14 mg/dL @ target (Very difficult to detect target bias with analytical SD)
- **Bias (PST-SST) = +0.128 mg/dL @ patient mean of 9.09 mg/dL**

Note: "Fraser" desirable bias reference: "Desirable Specifications for Total Error, Imprecision, and Bias, derived from intra- and inter-individual biologic variation." In: Ricos C, Alvarez V, Cava F, Garcia-Lario JV, Hernandez A, Jimenez CV, Minchinela J, Perich C, Simon M. "Current databases on biologic variation: pros, cons and progress." Scand J Clin Lab Invest 1999;59:491-500. This database was most recently updated in 2010, from Westgard QC website, http://www.westgardqc.com/biodatabase1.htm#SCC, accessed October 21, 2010.

Study of Bias among Laboratories for Tests Used in Monitoring End Stage Renal Disease

June 2009

Prepared by Westgard QC, Inc.

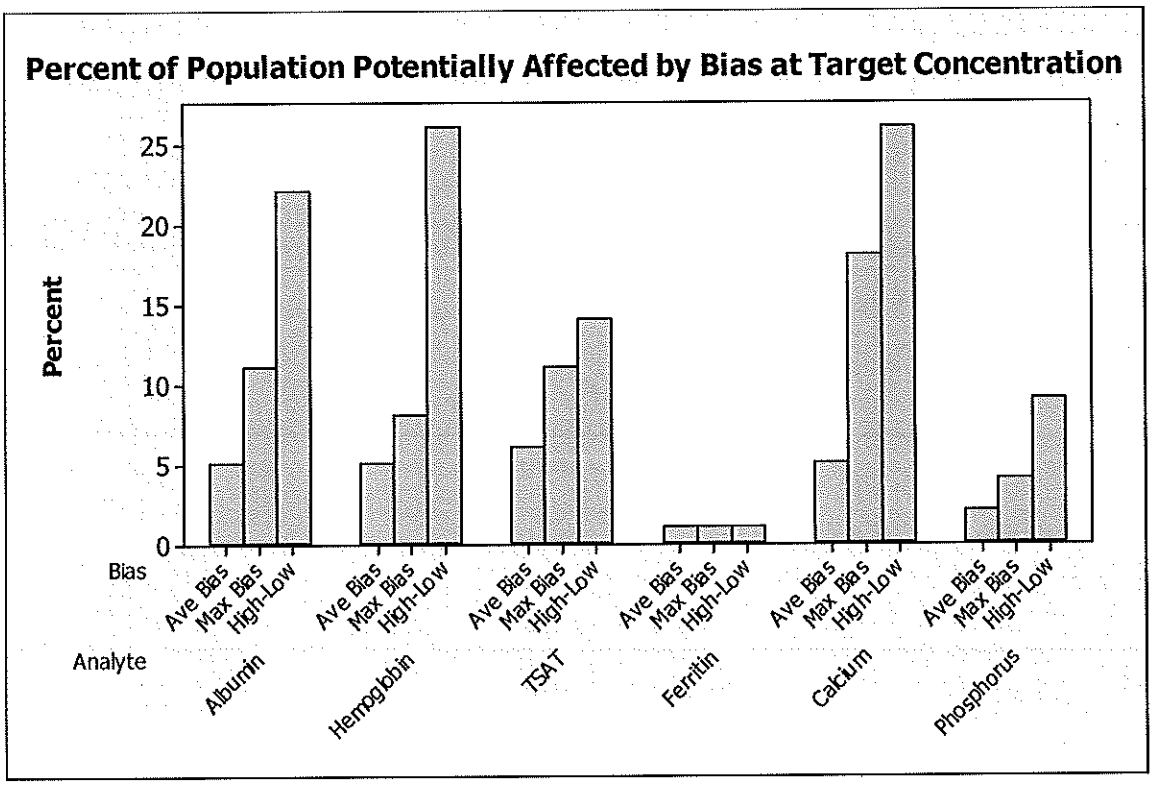
R. Neill Carey, Ph.D.
James O. Westgard, Ph.D.
Sten A. Westgard, MS

Prepared for the
Participating Laboratories

Executive Summary

Pay-for-Performance (P4P) incentives that involve laboratory tests as measures of quality assume that routine testing methods provide comparable test results. However, there are many different measurement principles and different method implementations found in different laboratories. Examination of proficiency testing results from a large number (1000 to 5000) of US labs involved in the College of American Pathologists survey showed significant biases among the different method subgroups for these tests. These biases were sufficiently large to compromise any P4P scheme utilizing laboratory test results to classify patients.

To assess whether similar biases existed in a smaller specialized group of ESRD testing labs, 50 patient specimen pools were collected and distributed to 8 different laboratories for analysis. Estimates of bias were assessed relative to the mean of results for all labs, as well as the maximum bias between the highest and lowest labs. The observed biases can cause the proportions of the ESRD population meeting target concentrations to be altered considerably, e.g., the biases between the highest and lowest laboratory can shift the proportion of the ESRD population meeting target concentrations by 20% or more for some of these analytes, as shown below.



These findings are consistent with the earlier study of comparability and again demonstrate the importance of establishing traceability and demonstrating comparability of any measures to be used in Pay-for-Performance incentives.

Introduction

Pay-for-Performance incentives are being considered to improve the quality of treatment of patients with end-stage renal disease (ESRD). Some of these incentives will involve financial rewards for successfully treating patients to achieve target values for laboratory tests used to monitor therapy for ESRD and its complications. It is anticipated that incentives will focus on measurements such as anemia management, dialysis adequacy, iron management, bone mineral metabolism, and vascular access [1]. This will involve laboratory tests for hemoglobin, urea, creatinine, iron, ferritin, transferrin saturation, calcium, phosphorus, and parathyroid hormone. It might also be expected that serum albumin will be involved because of its importance as an indicator of nutritional status and patient outcome. Laboratory test results would seem to be attractive as Pay-for-Performance indicators of the quality of treatment of patients with ESRD, as there are widely accepted practice guidelines from the National Kidney Foundation (KDOQI) that set outcome goals for key lab tests for monitoring the status of patients with chronic renal failure. Specific goals have been recommended for laboratory tests used to monitor nutrition [2], anemia [3,4], and bone metabolism [5,6]. These goals are in the form of target concentrations.

In a previous study, proficiency testing data were examined for these tests used to monitor patients with ESRD [7]. Significant biases were observed among different methods for these different laboratory tests. These biases among laboratories arise from their lack of traceability from primary reference methods and primary reference materials down to the working methods and manufacturers' calibrators used for routine patient testing. These biases were sufficiently large that the proportion of the ESRD population meeting a target concentration can be significantly altered simply by the choice of analytical method used for testing. The lack of comparability due to the lack of traceability for these methods has the potential to defeat the purpose of any Pay-for-Performance scheme utilizing laboratory test results to classify patients. While calcium, creatinine, and urea have the potential for complete traceability, there are distinct limitations to traceability for albumin, hemoglobin, ferritin, transferrin, iron, phosphorus, and parathyroid hormone.

This current study was undertaken to determine the levels of biases among a group of eight large laboratories for tests used to monitor patients with ESRD. These laboratories specialize in testing samples from ESRD patients. Several of them use the same methods from the same manufacturers for most of the analytes studied. The purpose of this study was to quantify the biases among these laboratories, which should perform similarly, and to assess the potential impact of these biases on the proportions of test results from ESRD patient populations that meet target specifications.

Fifty pooled serum samples were distributed to eight laboratories for analysis for albumin, hemoglobin, transferrin saturation, total iron, ferritin, calcium, phosphorus, urea nitrogen, and creatinine. Results were examined for bias among the laboratories, and biases were assessed vs. specifications for allowable error and for their impact on

proportions of the ESRD patient population meeting National Kidney Foundation KDOQI recommended target concentrations.

Materials and Methods

Participants and Methods

Eight large laboratories specializing in testing of specimens from ESRD patients participated in the study: DaVita Laboratory Services, Inc., Ft. Lauderdale, FL; DaVita Laboratory Services, Inc., DeLand, FL; Dialysis Clinic, Inc., Nashville, TN; Nationwide Laboratory Services, Ft. Lauderdale, FL; Renalab, Inc., Jackson, MS; Satellite Laboratory Services, Redwood City, CA; Spectra East, Inc, Northvale, NJ; and Spectra West, Inc., Milpitas, CA. Laboratory identities were randomly coded from the following names: Neptune, Commodore, Admiral, Kestrel, Eagle, Mariner, Tern, and Gull.

Analytical methods used for each test in each laboratory were as follows:
Albumin: Olympus 5400 with Olympus BCG; Admiral, Eagle, Mariner, Tern, Gull, Olympus 5431 with Olympus BCG; Kestrel, Roche Modular with DCL BCG; Commodore, Roche Modular with Genzyme BCG reagent; Neptune.
Hemoglobin: Siemens Advia 2120 with Siemens reagent; Neptune, Kestrel, Eagle, Mariner, Tern, Gull, Sysmex XE 21000; Commodore, Admiral.
Transferrin: Olympus 5400 with Olympus reagent; Eagle, Mariner, Tern, Gull, Olympus 5431 with Kamiya reagent; Admiral, Roche Modular with Roche reagent; Neptune, Commodore.
Iron: Olympus 5400 with Olympus reagent; Admiral, Eagle, Mariner, Tern, Gull, Olympus 5431 with Olympus reagent; Kestrel, Roche Modular with Roche reagent; Neptune, Commodore.
Ferritin: Olympus 5400 with Kamiya reagent; Admiral, Roche Modular with Roche reagent; Neptune, Commodore, Siemens Advia Centaur with Siemens reagent; Kestrel,
Calcium: Olympus 5400 with Olympus reagent; Admiral, Eagle, Mariner, Tern, Gull, Olympus 5431 with Olympus reagent; Kestrel, Roche Modular with Roche reagent; Neptune, Commodore.
Phosphorus: Olympus 5400 with Olympus reagent; Admiral, Eagle, Mariner, Tern, Gull, Olympus 5431 with Olympus reagent; Kestrel, Roche Modular with Roche reagent; Neptune, Commodore.
Urea: Olympus 5400 with Olympus reagent; Admiral, Eagle, Mariner, Tern, Gull, Olympus 5431 with Olympus reagent; Kestrel, Roche Modular with Roche reagent; Neptune, Commodore.
Creatinine: Olympus 5400 with Olympus reagent; Admiral, Eagle, Mariner, Tern, Gull, Olympus 5431 with Olympus reagent; Kestrel, Roche Modular with Roche reagent; Neptune, Commodore.

Specimens

Specimens were prepared by DCI. Specimen pools of hemoglobin were combined in pools of approximately 21 mL in 50 mL polypropylene containers using EDTA whole blood, less than 72 hours after collection. The pooled specimens were mixed by inversion twenty times; frozen at -70 °C; thawed, and centrifuged to remove cellular debris. Then the supernatant was removed to another 50 mL polypropylene container.

The hemolyslate/supernatant was mixed by inversion twenty times; dispensed into ten aliquots of 2 ml each in small polypropylene screw-capped tubes; labeled with the year, date of preparation, and pool number (1 through 50). These aliquots were stored at -70 °C until distribution. Two frozen aliquots were retained until study completion.

For chemistry analytes, serum specimen pools were combined in pools of approximately 21 mL in 50 mL polypropylene containers less than 72 hours after collection. The pooled specimens were mixed by inversion twenty times and centrifuged to remove any debris. Then the supernatant was removed to another 50 mL polypropylene container.

The supernatant was mixed by inversion twenty times; dispensed into ten aliquots of 2 mL each in small polypropylene screw-capped tubes; labeled with the year, date of preparation, and pool number (1 through 50). These aliquots were stored at -70 °C until distribution. Two frozen aliquots were retained until study completion.

Specimens were shipped frozen overnight to the participating laboratories. All laboratories tested all specimens on May 14 or May 15, 2009. Results were submitted on spreadsheets directly to the authors for analysis.

Assessment of Bias

Biases were calculated for each laboratory for each analyte as bias vs. the group mean in concentration units and percentages, mean absolute bias from the group mean, maximum absolute bias from the group mean, and high-low bias (difference between the highest laboratory mean and lowest laboratory mean). The effects of the biases on outcomes measures using KDOQI target concentrations were assessed by their impacts on populations of ESRD test results from one of the participating laboratories. For high-low bias, when the means of the patient test result populations were close to the KDOQI target concentrations, the simple bias between the two laboratory mean concentrations was taken as the bias. When the target concentrations were more distant from the means of the patient test result populations, linear regression was used to estimate the high-low bias at the target concentrations. Ordinary least squares linear regression and Passing-Bablock regression [8] techniques were used. For some analytes, there are no KDOQI target concentrations; for these analytes bias was assessed at the mean of the data.

Observed biases were compared to specifications for allowable total error and allowable bias. For some analytes of interest in ESRD, allowable total analytic error has been defined the Federal CLIA legislation [9]. For those tests without regulated CLIA allowable error, allowable total errors have been specified by the College of American Pathologists proficiency testing programs (CAP). For all of the analytes of interest in ESRD testing without CLIA limits, CAP has specified allowable error as $3 \cdot SD_{PT\ Peer\ Group}$, which is not applicable here because different instruments are used in these laboratories. A European group has specified "desirable" limits of bias based on biologic variation [10], with the intention of limiting the percentage increase in test result variability due to analytical imprecision to 12%. Target concentrations and allowable error and bias specifications used to assess the observed biases are summarized in the table below.

Analyte	Target Concentration(s)	Allowable Total Error (Source)	European Desirable Bias Limit (%)
Albumin	4.0 g/dl	10% (CLIA)	1.3
Hemoglobin	11.0 – 12.0 g/dL	10% (CLIA)	1.8
Transferrin Saturation	20%	$3 \cdot SD_{PT\ Peer\ Group}$ (CAP)	N/A (iron; 8.8, transferrin; 1.3)
Total Iron	N/A	20% (CLIA)	8.8
Ferritin	200 ng/mL	$3 \cdot SD_{PT\ Peer\ Group}$ (CAP)	5.2
Calcium	8.4 – 9.5 mg/dL	1.0 mg/dL (CLIA)	0.8
Phosphate	3.5 – 5.5 mg/dL	$3 \cdot SD_{PT\ Peer\ Group}$ (CAP)	3.2
Urea Nitrogen	N/A	9.0% (CLIA)	4.9
Creatinine	N/A	15.0% (CLIA)	3.8

Results

Results are summarized in Tables 1 and 2. Bias charts, results of statistical tests, and comparison plots are given in the Appendix.

Albumin

The KDOQI target concentration is 4.0 g/dL. Allowable error under Federal CLIA regulations is 10%. The European specification for desirable bias is 1.3%.

The grand mean for all samples in all laboratories was 3.76 g/dL. Means ranged from 3.69 g/dL, for laboratory Kestrel, to 3.86 g/dL for laboratory Eagle. The mean absolute bias from the grand mean was 0.05 g/dL (1.3%), and the maximum absolute bias from the grand mean was 0.10 g/dL (2.7%). The CLIA allowable error is 0.38 g/dL (10%).

The difference between laboratories Eagle and Kestrel (high-low bias) is representative of the maximum bias to be expected between two laboratories, and was 0.17 g/dL or 4.6% ($p = 0.000$). This simple bias between these two laboratories is a good estimate of the high-low bias near the target albumin concentration of 4.0 g/dL because the mean albumin concentrations for these two laboratories were close to the target concentration.

Estimates of bias between laboratories Eagle and Kestrel obtained from ordinary linear regression and Passing-Bablock linear regression were very close to the 0.17 g/dL simple bias (0.16 g/dL and 0.20 g/dL, respectively, see Appendix).

Hemoglobin

The KDOQI target concentrations are 11.0 and 12.0 g/dL. Allowable error under Federal CLIA regulation is 10%. The European specification for desirable bias is 1.8%.

The grand mean for all samples in all laboratories was 11.67 g/dL. Means ranged from 11.27 g/dL, for laboratory Kestrel, to 12.00 g/dL for laboratory Commodore. The mean absolute bias from the grand mean was 0.17 g/dL (1.5%), and the maximum absolute bias from the grand mean was 0.40 g/dL (3.4%). The CLIA 10% allowable error is 1.17 g/dL.

The difference between laboratories Neptune and Commodore (high-low bias) of 0.73 g/dL hemoglobin (6.3%) is representative of the maximum differences to expect between two laboratories, and is statistically significant at $p = 0.000$. The simple bias between laboratories Neptune and Commodore is a good estimate of the bias near the target hemoglobin concentration of 11.0 g/dL because the mean hemoglobin concentrations are close to the target concentration.

Bias between laboratories Neptune and Commodore at 12.0 g/dL was calculated by linear regression to be 0.8 g/dL (see Appendix).

Transferrin Saturation (TSAT)

The KDOQI target is 20.0%. There is no Federal CLIA specification of allowable error. CAP allowable error is $3 \cdot SD_{PT\ Peer\ Group}$. There is no European specification for desirable bias for TSAT.

The grand mean for all samples in all laboratories was 26.30%. Means ranged from 22.58%, for laboratory Neptune, to 28.32% for laboratory Tern. The mean absolute bias from the grand mean was 1.66% (6.3%), and the maximum absolute bias from the grand mean was 3.72% (14.1%). The CAP allowable error would depend on the peer group standard deviation. Since more than one method group is represented, allowable error is ambiguous.

The lowest mean is from laboratory Neptune, with a mean TSAT of 22.58%, and the highest mean is from laboratory Tern, with a mean of 28.32%. The high-low bias of 5.74% TSAT (21.8% of the grand mean) between laboratories Neptune and Tern is statistically significant at $p = 0.000$.

The bias to be expected between laboratories Neptune and Tern at the target concentration 20% TSAT was calculated by linear regression to be 5.0% (see Appendix).

Iron

There is not a separate KDOQI target concentration for iron. Iron is a factor in the calculation of transferrin saturation. The CLIA allowable error for iron is 20%. The European specification for desirable bias for iron is 8.8%.

The grand mean for all samples in all laboratories was 57.2 µg/dL. The mean absolute bias from the grand mean was 1.9 µg/dL (3.4%), and the maximum absolute bias from the grand mean was 3.3 µg/dL (5.7%).

The lowest mean is from laboratory Admiral, with a mean iron concentration of 54.0 µg/dL, and the highest mean is from laboratory Mariner, with a mean iron concentration of 60.1 µg/dL. The high-low bias of 6.1 µg/dL (10.7% of the grand mean) between laboratories Mariner and Admiral is statistically significant at $p = 0.000$.

A comparison graph of iron concentrations from laboratories Mariner and Admiral is shown in the Appendix. Bias was not estimated by regression because there is not a target concentration.

Ferritin

The KDOQI target concentration is 200 ng/mL. There is no Federal CLIA specification of allowable error. CAP allowable error is $3 \cdot SD_{PT\ Peer\ Group}$. The European specification for desirable bias is 5.2%.

The grand mean for all samples in all laboratories was 699.9 ng/mL. The mean absolute bias from the grand mean was 38.4 ng/mL (5.5%), and the maximum absolute bias from the grand mean was 81.0 ng/mL (11.6%). The CAP allowable error would depend on the peer group standard deviation. Since more than one method group is represented, allowable error is ambiguous.

The lowest mean is from laboratory Gull, with a mean ferritin concentration of 635.1 ng/mL, and the highest mean is from laboratory Commodore, with a mean ferritin concentration of 780.9 ng/mL. The high-low bias of 145.8 ng/mL (20.8% of the grand mean) between laboratories Gull and Commodore is statistically significant at $p = 0.000$.

These means are far from the target value of 200 ng/mL; thus the simple bias is not an accurate indicator of bias at 200 ng/mL. The bias between laboratories Gull and Commodore at 200 ng/mL, estimated by Passing-Bablock regression, is approximately 15 ng/mL (see Appendix).

Calcium

The KDOQI target concentrations for calcium are 8.4 and 9.5 mg/dL. CLIA allowable error is 1.0 mg/dL. The European specification for desirable bias is 0.8%.

The grand mean for all samples in all laboratories was 8.92 mg/dL. The mean absolute bias from the grand mean was 0.09 mg/dL (1.0%), and the maximum absolute bias from the grand mean was 0.36 mg/dL (4.0%).

The lowest mean is from laboratory Admiral, with a mean calcium concentration of 8.77 mg/dL, and the highest mean is from laboratory Neptune, with a mean calcium concentration of 9.28 mg/dL. Neptune's mean is an outlier at $p < 0.02$ [11], although all laboratories' means are well within the CLIA limit of allowable error, 1.0 mg/dL. The high-low bias of 0.51 mg/dL (5.7%) between laboratories Neptune and Admiral is statistically significant at $p = 0.000$.

The bias between laboratories Neptune and Admiral, calculated by linear regression, at 8.4 mg/dL and 9.5 mg/dL calcium concentrations is approximately 0.5 mg/dL (see Appendix), in good agreement with the simple bias of 0.51 mg/dL.

Phosphorus

The KDOQI target concentrations for phosphorus are 3.5 and 5.5 mg/dL. There is no Federal CLIA specification of allowable error. CAP allowable error is $3 \cdot SD_{PT\ Peer\ Group}$. The European specification for desirable bias is 3.2%.

The grand mean for all samples in all laboratories was 5.59 mg/dL. The mean absolute bias from the grand mean was 0.09 mg/dL (1.6%), and the maximum absolute bias from the grand mean was 0.20 mg/dL (3.6%).

The lowest mean is from laboratory Admiral, with a mean phosphorus concentration of 5.40 mg/dL, and the highest mean is from laboratory Neptune, with a mean calcium concentration of 5.79 mg/dL. The high-low bias of 0.39 mg/dL (7.0% of the grand mean) between laboratories Neptune and Admiral is statistically significant at $p = 0.000$.

The bias, calculated by Passing-Bablock regression, at phosphorus concentrations of 3.5 mg/dL and 5.5 mg/dL, is approximately 0.4 mg/dL (see Appendix), in good agreement with the simple bias of 0.39 mg/dL.

Blood Urea Nitrogen (BUN)

There is no KDOQI target concentration for urea nitrogen. CLIA allowable error is 9%. The European specification for desirable bias is 4.9%. Plasma and urine urea concentrations are factors in calculating dialysis efficiency. The impact of bias in measuring urea clearance may be reduced, because urea appears in the numerator and denominator of clearance equations. The impact of bias would be reduced if the urine and plasma urea measurements had the same bias.

The grand mean for all samples in all laboratories was 57.1 mg/dL. The mean absolute bias from the grand mean was 1.25 mg/dL (2.2%), and the maximum absolute bias from the grand mean was 1.9 mg/dL (3.3%).

The lowest mean is from laboratory Tern, with a mean urea concentration of 55.5 mg/dL, and the highest mean is from laboratory Mariner, with a mean urea concentration of 58.9 mg/dL. The high-low bias of 3.4 mg/dL (6.0% of the grand mean) between laboratories Mariner and Tern is statistically significant at $p = 0.000$.

A comparison graph of urea concentrations from laboratories Mariner and Tern is shown in the Appendix. Bias was not estimated by regression because there is no KDOQI target concentration for urea.

Creatinine

There is no KDOQI target concentration for creatinine. Plasma and urine creatinine concentrations are factors in calculating dialysis efficiency. The CLIA allowable error is 15%. The European specification for desirable bias is 3.8%. Similarly to urea, the clearance calculation may mitigate the effects of bias.

The grand mean for all samples in all laboratories was 9.49 mg/dL. The mean absolute bias from the grand mean was 0.14 mg/dL (1.5%), and the maximum absolute bias from the grand mean was 0.31 mg/dL (3.3%).

The lowest creatinine mean is from laboratory Admiral, with a mean concentration of 9.27 mg/dL, and the highest mean is from laboratory Eagle, with a mean creatinine concentration of 9.80 mg/dL. The high-low bias of 0.53 mg/dL (5.6% of the grand mean) between laboratories Eagle and Admiral is statistically significant at $p = 0.000$.

A comparison graph of creatinine concentrations from laboratories Eagle and Admiral is shown in the Appendix. Bias was not estimated by regression because there is no KDOQI target concentration for creatinine.

Table 1. Summary of Comparison Results.

	Neptune	Commodore	Admiral	Kestrel	Eagle	Mariner	Tern	Gull
Albumin Mean	3.74	3.82	3.70	3.69	3.86	3.74	3.74	3.78
Albumin Bias	-0.02	0.06	-0.06	-0.07	0.10	-0.02	-0.02	0.02
% Albumin Bias	-0.55	1.69	-1.50	-1.93	2.71	-0.49	-0.49	0.57
Hemoglobin Mean	11.27	12.00	11.90	11.55	11.50	11.68	11.67	11.81
Hemoglobin Bias	-0.40	0.33	0.23	-0.13	-0.17	0.01	0.00	0.13
% Hemoglobin Bias	-3.45	2.83	1.94	-1.07	-1.47	0.07	0.01	1.14
TSAT Mean	22.58	24.78	25.24	27.30	27.94	28.24	28.32	25.96
TSAT Bias	-3.72	-1.52	-1.06	1.00	1.64	1.94	2.02	-0.34
% TSAT Bias	-14.13	-5.76	-4.01	3.82	6.26	7.40	7.70	-1.27
Iron Mean	57.56	55.34	53.98	57.88	59.78	60.12	58.66	54.64
Iron Bias	0.32	-1.90	-3.26	0.64	2.54	2.88	1.42	-2.60
% Iron Bias	0.55	-3.33	-5.70	1.11	4.43	5.02	2.47	-4.55
Ferritin Mean	736.70	780.90	638.40	697.60	722.40	674.90	713.20	635.10
Ferritin Bias	36.80	81.00	-61.50	-2.30	22.50	-25.00	13.30	-64.80
% Ferritin Bias	5.26	11.57	-8.79	-0.33	3.21	-3.57	1.90	-9.26
Calcium Mean	9.28	8.84	8.77	8.91	8.93	8.87	8.84	8.91
Calcium Bias	0.36	-0.08	-0.15	-0.01	0.01	-0.05	-0.08	-0.01
% Calcium Bias	4.07	-0.89	-1.63	-0.11	0.06	-0.55	-0.85	-0.11
Phosphorus Mean	5.79	5.71	5.40	5.48	5.65	5.56	5.60	5.58
Phosphorus Bias	0.19	0.11	-0.20	-0.11	0.05	-0.04	0.00	-0.02
% Phosphorus Bias	3.48	2.05	-3.53	-2.02	0.95	-0.67	0.05	-0.31
Urea Mean	56.76	55.72	56.14	58.64	58.64	58.92	55.50	56.18
Urea Bias	-0.30	-1.34	-0.92	1.58	1.58	1.86	-1.56	-0.88
% Urea Bias	-0.53	-2.35	-1.62	2.76	2.76	3.26	-2.74	-1.55
Creatinine Mean	9.60	9.36	9.27	9.66	9.80	9.48	9.39	9.39
Creatinine Bias	0.10	-0.13	-0.22	0.17	0.31	-0.02	-0.10	-0.10
% Creatinine Bias	1.09	-1.39	-2.36	1.75	3.23	-0.17	-1.10	-1.05

Table 2. Summary of Comparison Statistics

Analyte	Mean	Maximum Bias	Percent Maximum Bias	Average Bias	Percent Average Bias	High - Low	Percent High - Low
Albumin, g/dL	3.8	0.10	2.7	0.05	1.2	0.17	4.6
Hemoglobin, g/dL	11.7	0.40	3.5	0.17	1.5	0.73	6.3
TSAT, %	26.3	3.7	14.1	1.66	6.3	5.74	21.8
Iron, µg/dL	57.2	3.26	5.7	1.94	3.4	6.14	10.7
Ferritin, ng/mL	699.9	81.0	11.6	38.4	5.5	145.80	20.8
Calcium, mg/dL	8.9	0.36	4.1	0.09	1.0	0.51	5.7
Phosphorus, mg/dL	5.6	0.20	3.5	0.09	1.6	0.39	7.0
BUN, mg/dL	57.1	1.86	3.3	1.25	2.2	3.42	6.0
Creatinine, mg/dL	9.5	0.31	3.2	0.14	1.5	0.53	5.6

Discussion

The average absolute biases and absolute maximum biases from the grand means, and the differences between the highest and lowest laboratory have been estimated for each analyte. These data are summarized as percentages in Figure 1. Average absolute percent biases ranged from 1.0% for calcium to 6.3% for transferrin saturation. Maximum absolute percent biases ranged from 2.7% for albumin to 14.1% for transferrin saturation. High-low differences ranged from 4.6% for albumin to 21.9% for transferrin saturation.

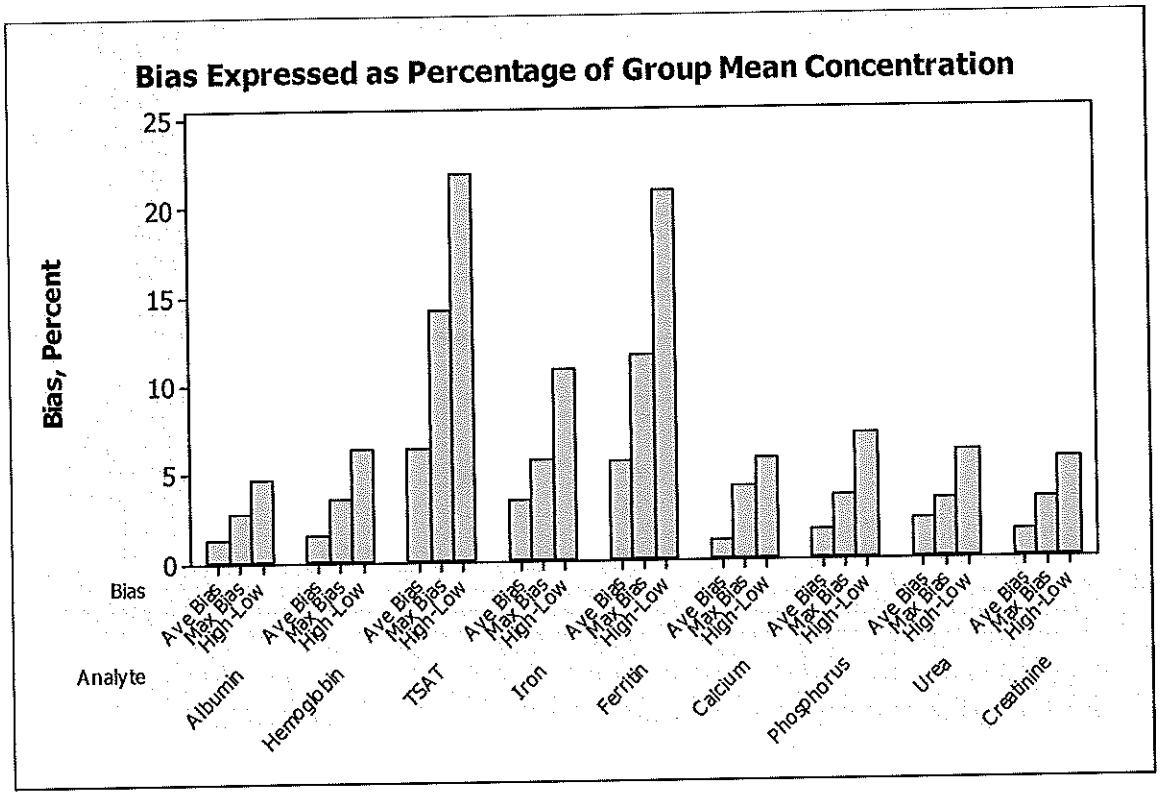


Figure 1. Sizes of absolute average bias (Ave Bias), maximum absolute bias (Max Bias), and high-low absolute bias (High-Low), expressed as percentages of the grand mean concentrations of the 50 samples tested for each analyte.

These same biases are shown in comparison to CLIA total allowable error and European specifications for allowable bias in Figure 2. High-low bias is not included since total allowable error is measured vs. the group mean. CLIA has not specified quantitative limits for error for ferritin and phosphorus. This figure shows that the observed biases are well within the CLIA limits of allowable error.

Even the mean absolute biases for albumin, hemoglobin, ferritin, and calcium are close to or exceed the European specifications for bias based on biologic variation (desirable specification). For albumin and calcium, six of the eight laboratories use equipment and

reagents from the same manufacturer; five of the six use the same model analyzer. The maximum absolute bias for phosphorus exceeds the European desirable specification, and the maximum absolute bias for creatinine is close to the European desirable specification. If more laboratories were included in this study, it is to be expected that the maximum absolute bias would be larger and would exceed the limit for creatinine.

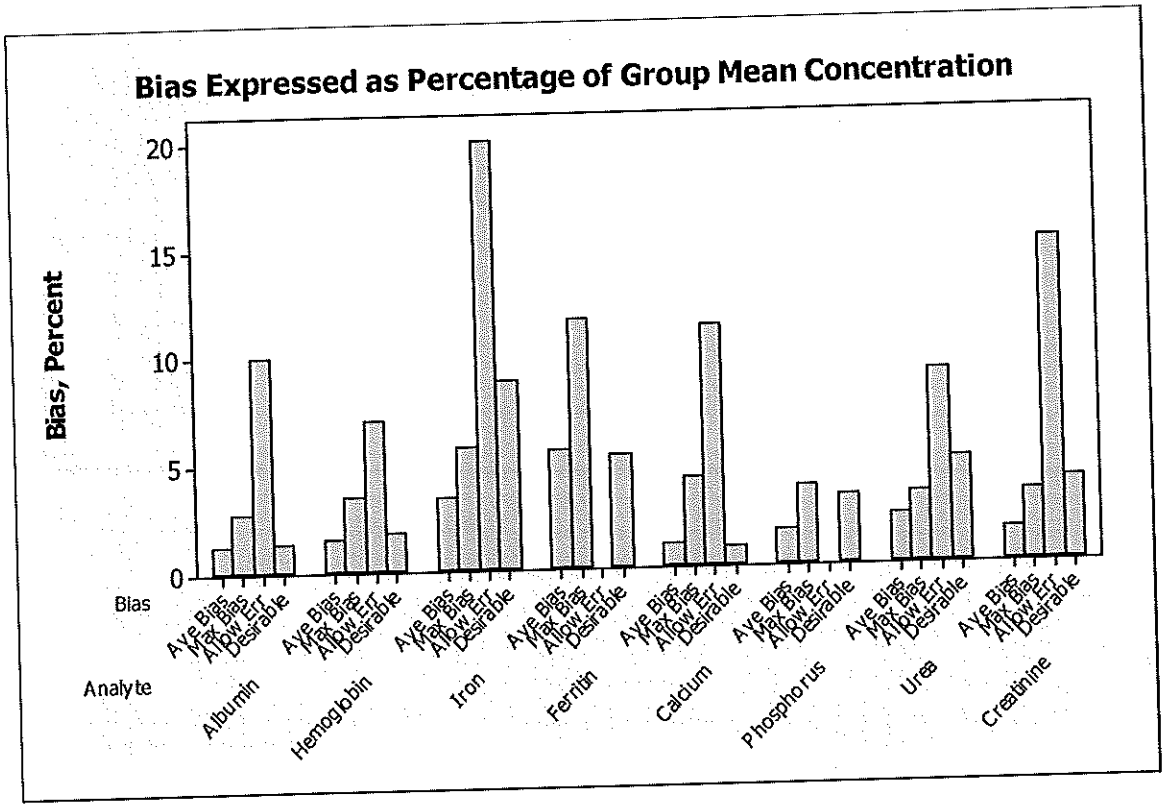


Figure 2. Sizes of absolute average bias (Ave Bias), maximum absolute bias (Max Bias), CLIA allowable error, and European desirable limits for bias, expressed as percentages of the grand mean concentrations of the 50 samples tested for each analyte.

In our previous study of comparability of laboratory tests of interest in ESRD, we examined biases among peer groups of laboratories in proficiency testing [7]. In this study samples were distributed to a group of eight very large laboratories specializing in testing specimens from ESRD patients. Biases among the laboratories were calculated without regard to peer group, as they would be observed in Pay-for-Performance from this group of laboratories which use only a few different methods for each analyte. Figure 3 shows a comparison of the biases observed in the present study vs. those of our previous proficiency testing study, where "Ave Bias PT" and "Max Bias PT" are the mean absolute bias and maximum absolute bias, respectively, from the grand mean over all method peer groups. Absolute and maximum biases are worse in the present study than the PT study for hemoglobin, and nearly identical to those in the PT study for iron, calcium, and urea. The present study demonstrates that biases among these eight specialty laboratories are similar to those seen in proficiency testing.

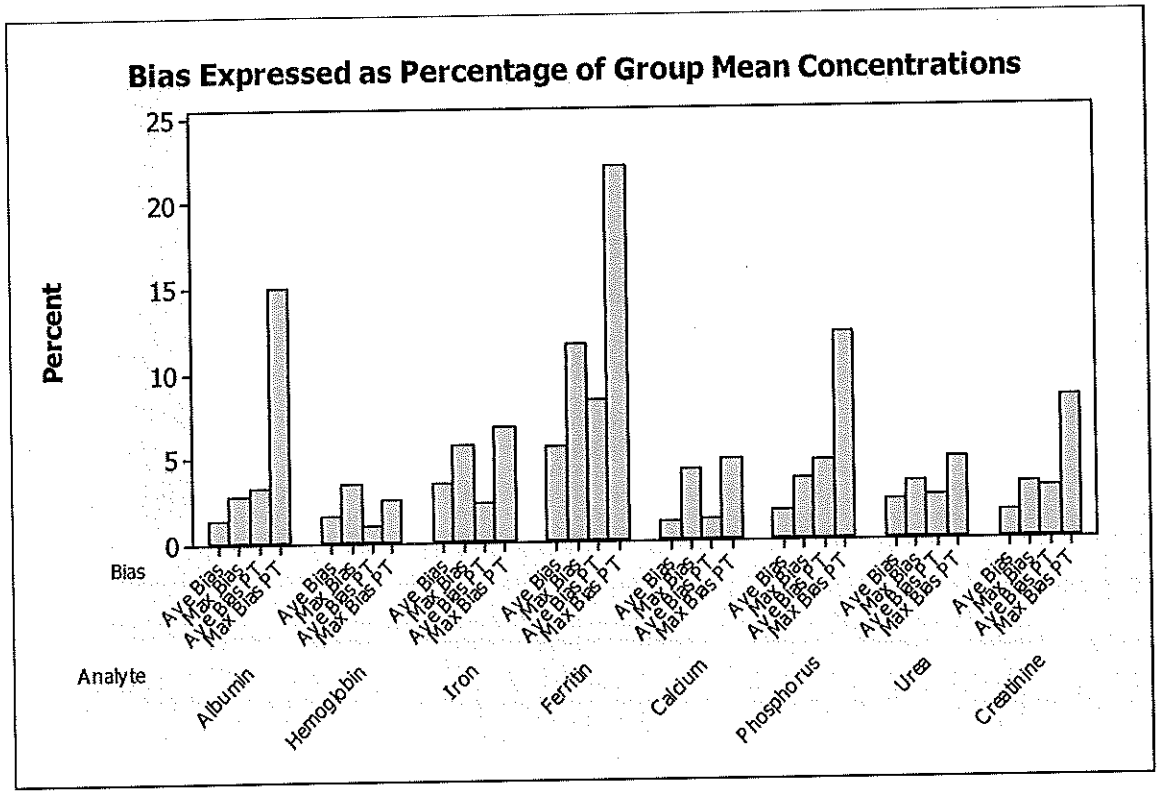


Figure 3. Sizes of absolute average absolute bias and maximum absolute bias from the current study and average absolute bias (Ave Bias PT) and maximum absolute bias (Max Bias PT) from the previous study utilizing PT specimens.

Effects of Bias on Proportions of ESRD Patient Populations Meeting Target Concentrations

Albumin

The average bias was 0.05 g/dL, the maximum bias from the grand mean was 0.1 g/dL, and the bias between laboratories Kestrel and Eagle was nearly 0.2 g/dL.

The distribution of patient albumin concentrations from one of the participating laboratories is shown below. The tallest bar, which includes about 22% of the population, is centered at 4.0 g/dL, the target concentration. The maximum bias from the grand mean, 0.1 g/dL, is half the width of this bar, representing a potential shift of about 11% of the patient population above or below the target concentration.

The bias of nearly 0.2 g/dL between laboratories Kestrel and Eagle could shift nearly 22% of the patient population from one side of the target concentration of 4.0 g/dL to the other. The white arrow shows the effect of the bias between the highest and lowest laboratories. In samples measured at laboratory Eagle, as many as 22% more of the

patient population could be above the 4.0 g/dL outcome goal than those measured at laboratory Kestrel. The impact of shifts on patient concentrations meeting or exceeding the target concentration is so significant for albumin because the target concentration is near the densest part of the histogram.

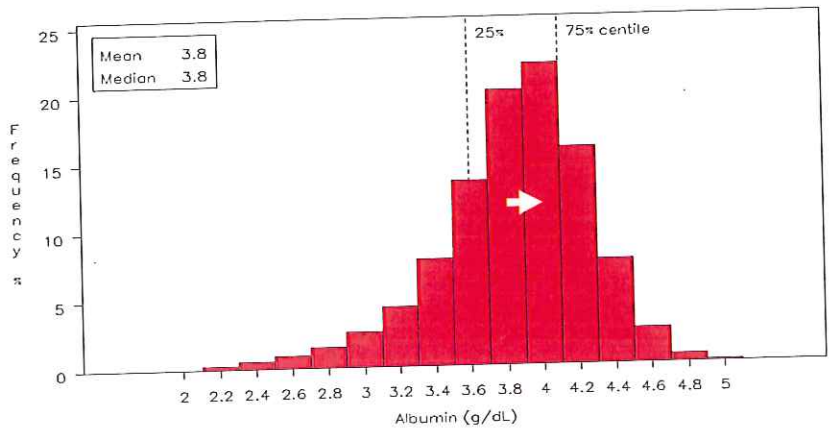


Figure 4. Distribution of albumin test results in ESRD patient population.

Laboratories Eagle (Olympus 5400 and Olympus BCG reagent) and Kestrel (Olympus 5431 and Olympus BCG reagent) use different models of the same instrument from the same instrument vendor and the same reagent for albumin testing. The bias observed between the labs may be due entirely due to small differences in performance of reagent lots and calibrator lots from the same manufacturer. All of the laboratories included in the study use the BCG albumin method. Biases between laboratories using the BCG method and laboratories using the BCP method would be significantly larger [12].

Hemoglobin

The maximum absolute bias from the grand mean was 0.40 g/dL, and the difference between laboratories Neptune and Commodore was 0.73 g/dL.

The distribution of hemoglobin values among dialysis patients in one of the participating laboratories is shown below. Each bar is 0.2 g/dL. The average bias of 0.17 g/dL is nearly the width of one bar at 12.0 g/dL, nearly 7%. The maximum absolute bias from the mean of 0.40 g/dL is the width of two bars, and represents approximately a potential shift of approximately 13% of the population up to the 12.0 g/dL target.

Bias at 12.0 g/dL between laboratories Neptune and Commodore was calculated by linear regression to be 0.8 g/dL. This bias would bring albumin concentrations as low as 11.2 g/dL in laboratory Neptune up to 12.0 g/dL in laboratory Commodore. Approximately 26% of the population would be shifted by the bias if the target is 12.0 g/dL. The white arrow beginning at 11.2 g/dL on the right side of the chart represents the impact of this shift.

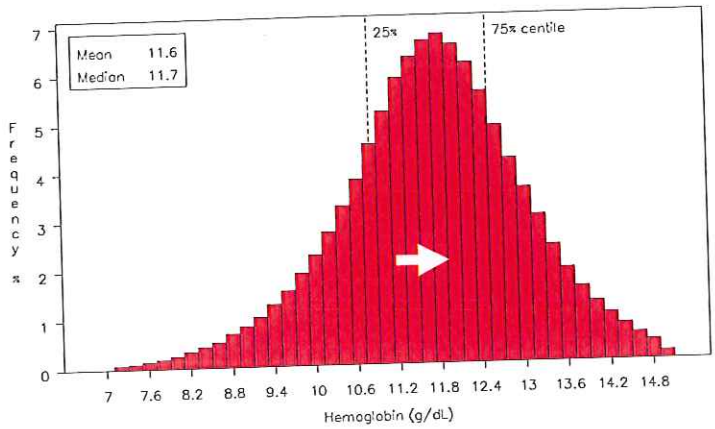


Figure 5. Distribution of hemoglobin test results in ESRD patient population.

Transferrin Saturation

The mean absolute bias from the grand mean was 1.66% (6.3%), and the maximum absolute bias from the grand mean was 3.72% (14.1%). The bias to be expected between laboratories Neptune and Tern at the target concentration 20% TSAT was 5.0%.

The distribution of percent transferrin saturation at one of the participating laboratories is shown below. Each bar is 1% transferrin saturation. In the region of the target 20% saturation, each bar is approximately 4% of the population. A bias of 1.7% could shift 6% of the population over the target. Bias of 3.7% could shift 11%, and bias of 5.0% between laboratories Neptune and Tern could shift 14% of the population over the target. This is shown by the arrow below.

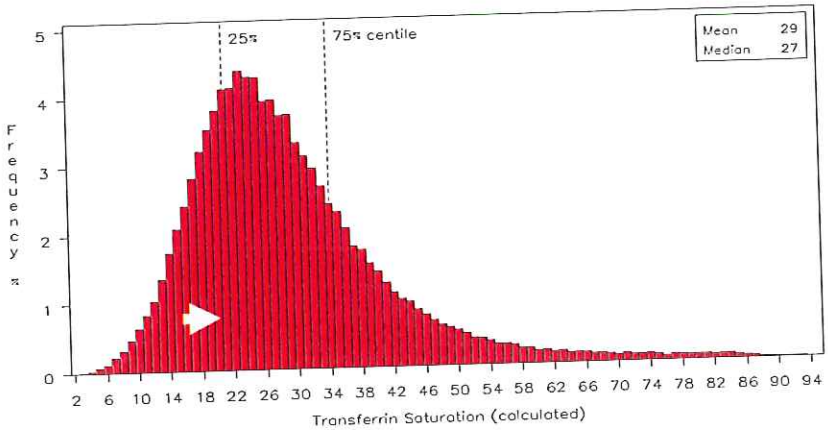


Figure 6. Distribution of transferrin saturation test results in ESRD patient population.

Some of the variation observed in transferrin saturation was due to the diversity of methods used by the participating laboratories. Laboratory Neptune calculated the transferrin saturation from iron and transferrin concentrations measured with the methods on Roche Modular instrument. Laboratory Tern calculated the transferrin saturation from iron and unbound iron binding capacity values obtained using the methods on the Olympus 5400 instrument.

Total Iron

The mean absolute bias from the grand mean was 1.9 µg/dL (3.4%), and the maximum absolute bias from the grand mean was 3.3 µg/dL (5.7%). These biases are small relative to those of transferrin saturation, suggesting that biases in transferrin cause most of the biases in transferrin saturation.

Ferritin

The mean absolute bias from the grand mean was 38.4 ng/mL, and the maximum absolute bias from the grand mean was 81.0 ng/mL. The estimated bias between laboratories Gull and Commodore at 200 ng/mL was approximately 15 ng/mL. The histogram below shows the distribution of ferritin concentrations at one of the participating laboratories. In the region of the 200 ng/mL target, each bar is 50 ng/mL wide, and represents approximately 4% of the population. Approximately 1% of the population in the region of 200 ng/mL could be affected by the estimated 15 ng/mL bias.

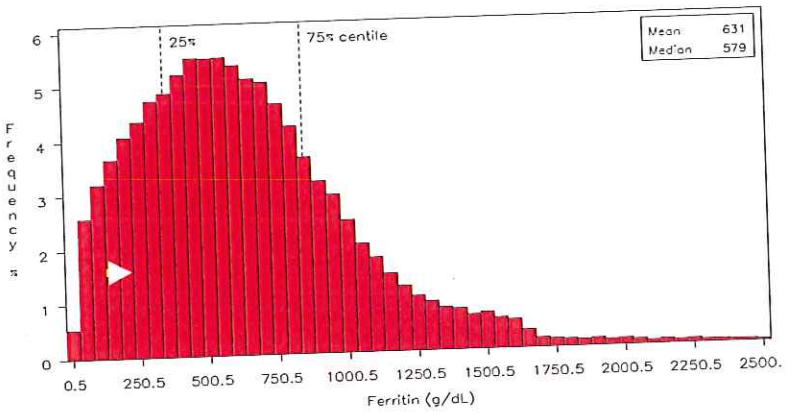


Figure 7. Distribution of ferritin test results in ESRD patient population.

Calcium

The mean absolute bias from the grand mean was 0.09 mg/dL, and the maximum absolute bias from the grand mean was 0.36 mg/dL. The difference between laboratories Neptune and Admiral was 0.51 mg/dL.

The distribution of calcium values among dialysis patients in one of the participating laboratories is shown below. Each bar is 0.2 mg/dl wide. At 9.5 mg/dL, the mean absolute bias of 0.09 mg/dL is equivalent to approximately ½ of a bar, or about 5%. The maximum absolute bias of 0.36 mg/dL is equivalent to approximately 1 ¾ bars, or about 18%. A bias of 0.5 mg/dL at 9.5 mg/dL includes the population in the bars from 9.0 mg/dL to 9.5 mg/dL. The calcium concentrations of approximately 26% of the population would be shifted from 8.4 mg/dL or less in laboratory Admiral to greater than 8.4 mg/dL in laboratory Neptune. This effect is shown by the arrow on the histogram below.

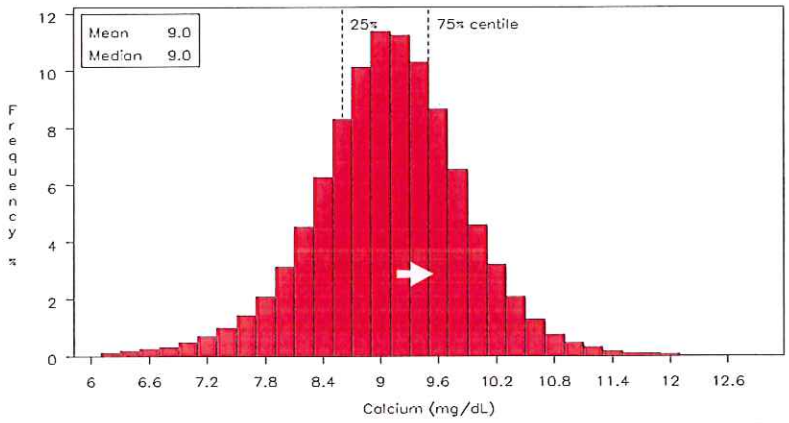


Figure 8. Distribution of calcium test results in ESRD patient population.

Phosphorus

The mean absolute bias from the grand mean was 0.09 mg/dL, and the maximum absolute bias from the grand mean was 0.20 mg/dL. The bias to be expected between laboratories Neptune and Admiral at the target concentrations of 3.5 mg/dL and 5.5 mg/dL was 0.4 mg/dL.

The distribution of phosphorus values among dialysis patients in one of the participating laboratories is shown below. The width of each bar is 0.5 mg/dL. A bias of 0.09 mg/dL represents approximately 1/5 of the bar centered at 5.5 mg/dL, or about 2% of the population. A bias of 0.20 mg/dL represents approximately 2/5 of the bar, or about 4% of the population. A bias of 0.4 mg/dL at 5.5 mg/dL includes approximately 4/5 of the population in the bar centered at 5.5, or approximately 9% of the population. The phosphorus concentrations of approximately 9% of the population could be shifted from 5.1 mg/dL or less in laboratory Admiral to 5.5 mg/dL or greater in laboratory Neptune. This effect is shown by the arrow on the histogram below.

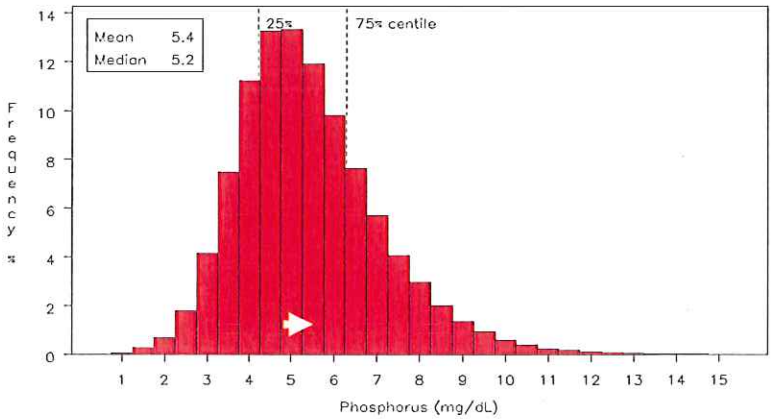


Figure 9. Distribution of phosphorus test results in ESRD patient population.

The proportions of the populations near target concentrations that are potentially affected near the target concentrations are summarized in Figure 10.

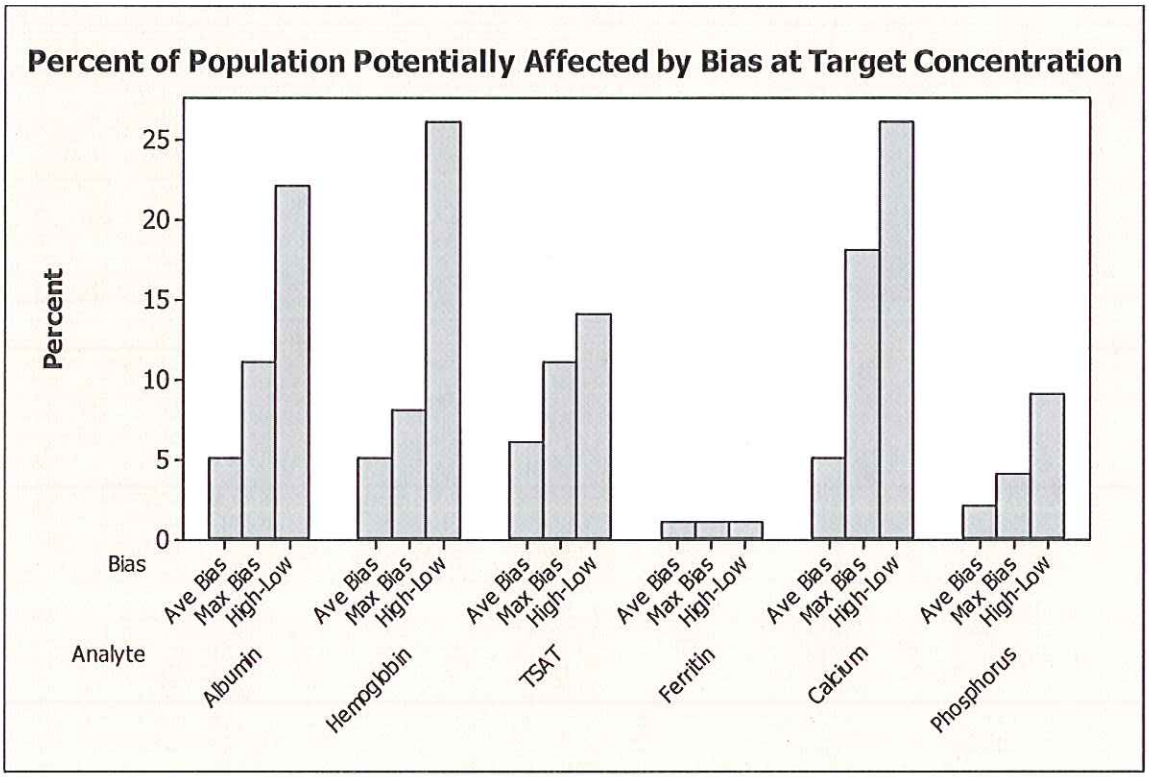


Figure 10. Percentages of patient result populations affected by bias at KDOQI target concentrations.

Conclusions

In our previous study [7], we concluded, "...biases observed among these laboratory tests are sufficiently large that the proportion of the ESRD population meeting a target concentration can be significantly altered simply by the choice of the laboratory method used for testing. The lack of comparability due to the lack of traceability for reference methods, materials, and calibrators has the potential to defeat the purpose of any Pay-for-Performance scheme utilizing laboratory test results to classify patients."

Clearly the current quality of testing for several key analytes used in monitoring the care of ESRD patients is subject to the effects of large biases, even among the best of laboratories. These biases between the highest and lowest of these eight major laboratories are large enough to impact on Pay-for-Performance by shifting 20% or more the proportion of the ESRD population meeting target values for albumin, hemoglobin, and calcium. The impact of bias on transferrin saturation, phosphorus, and ferritin was less, 14%, 9%, and 1% of the populations potentially affected, respectively. Clearly, these biases are sufficiently large that the proportions of the ESRD population meeting target concentrations for several tests can be altered dramatically by the biases among the laboratories.

Parathyroid hormone was not included in the study because its variability among laboratories is well known. Maximum biases of 30% to 50% were documented in our previous study [7]; even larger proportions of ESRD patients meeting target concentrations would be impacted by these huge biases.

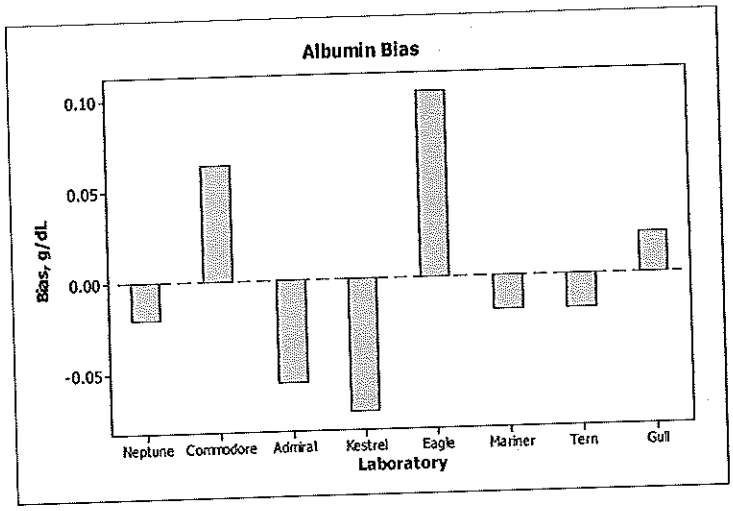
There are serious issues of traceability even in laboratories using very similar methods for the same analyte. For example, although calcium has the potential for complete traceability, as much as 26% of the ESRD population could be shifted by bias among these eight laboratories. These issues of traceability must be addressed before Pay-for-Performance is viable.

This study also demonstrates that the current specifications for allowable error under the CLIA regulations do not insure the quality necessary for Pay-for-Performance to be viable for albumin, hemoglobin, transferrin saturation and calcium. More clinically appropriate specifications for allowable error will drive improvement in traceability.

Appendix: Statistical Printouts, Correlation Graphs, and Calculations

Albumin

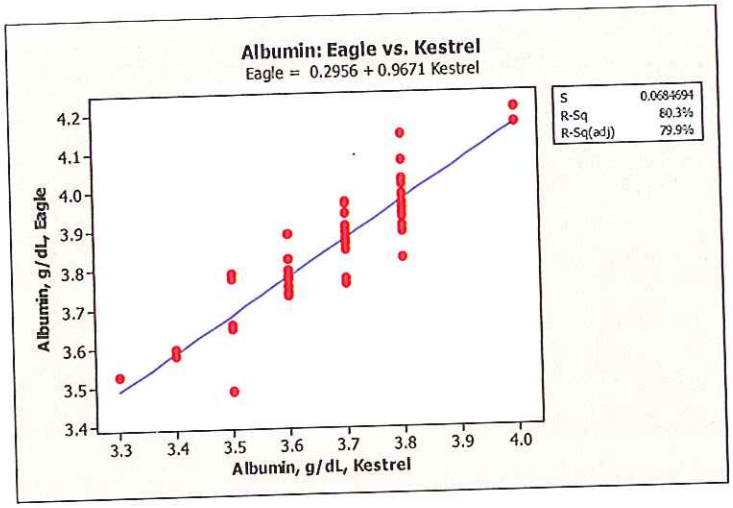
The mean biases of the laboratories from the grand mean for the 50 samples are shown in the chart below.



Paired T-Test and CI: Kestrel, Eagle

```
Paired T for Kestrel - Eagle

             N      Mean    StDev  SE Mean
Kestrel     50    3.6860   0.1414   0.0200
Eagle       50    3.8602   0.1526   0.0216
Difference  50  -0.17416  0.06793  0.00961

95% CI for mean difference: (-0.19346, -0.15486)
T-Test of mean difference = 0 (vs not = 0): T-Value = -18.13  P-Value = 0.000
```

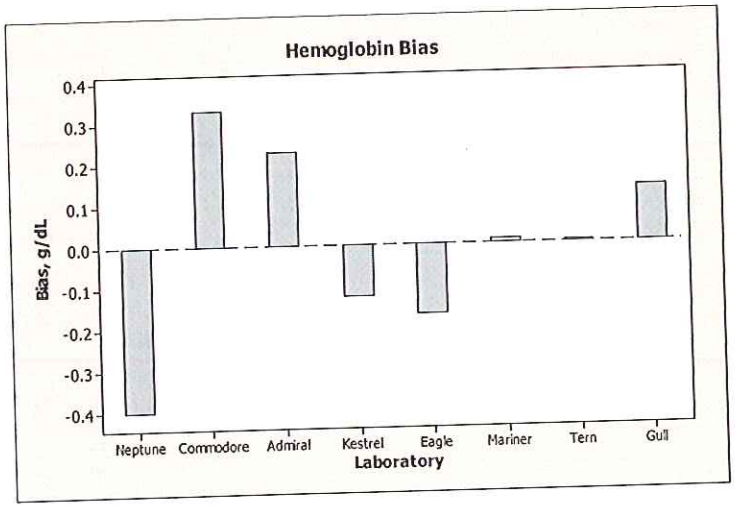
A plot of the albumin values from laboratory Eagle vs. those from laboratory Kestrel is shown below. Ordinary least squares linear regression statistics can be used to calculate the albumin result that laboratory Eagle would obtain from a sample whose albumin was measured as 4.0 g/dL in laboratory Kestrel. Using the slope of 0.967 and the y-intercept of 0.296 g/dL and a concentration of 4 g/dL, in laboratory Kestrel, laboratory Eagle would obtain an albumin concentration of 4.16 g/dL on average. The correlation coefficient, r is 0.896, which indicates that simple regression should not be to calculate bias at concentrations far from the center of the data.



Nonparametric Passing-Bablock regression analysis (not affected by the range of the data) gives a slope of 1.000 and the y-intercept is 0.20 g/dL. On average, if a sample's albumin concentration, measured in laboratory Kestrel, was at the target concentration of 4.0 g/dL, the result in laboratory Eagle would be 4.2 g/dL, according to Passing-Bablock regression. This bias of 0.2 g/dL from Passing-Bablock regression is very close to the simple bias of 0.174 g/dL calculated above by subtracting the simple means and the bias of 0.16 g/dL from simple linear regression.

Hemoglobin

The mean biases of the laboratories from the grand mean for the 50 samples are shown in the chart below.

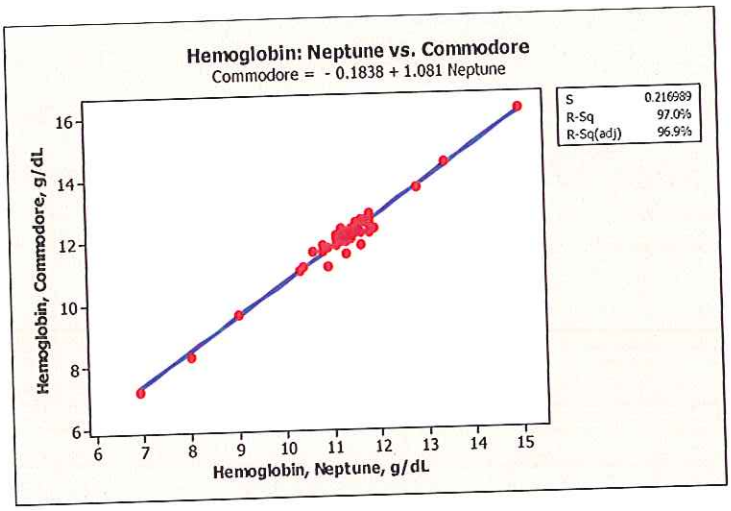


Paired T-Test and CI: Commodore, Neptune

```
Paired T for Commodore - Neptune

              N     Mean   StDev  SE Mean
Commodore    50  12.004   1.237    0.175
```

```
Neptune      50  11.270   1.126    0.159
Difference   50   0.7340  0.2335   0.0330

95% CI for mean difference: (0.6676, 0.8004)
T-Test of mean difference = 0 (vs not = 0): T-Value = 22.23  P-Value = 0.000
```

A plot of hemoglobin results by laboratory Neptune vs. hemoglobin results by laboratory Commodore is shown below.



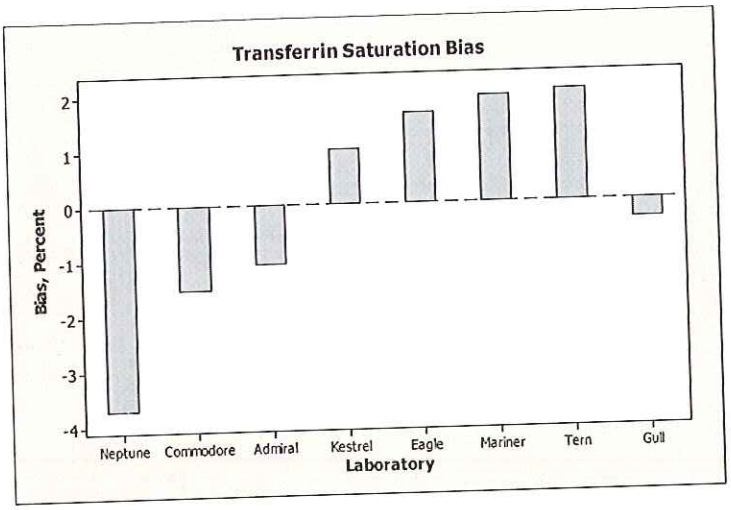
The ordinary least squares linear regression slope and intercept are 1.081 and -0.18 g/dL, respectively. The correlation coefficient is 0.985, indicating that ordinary least squares linear regression should not be used to estimate bias at concentrations far removed from the means of the data. The bias between laboratories Commodore and Neptune at 11.0 g/dL calculated by linear regression is 0.71 g/dL. Bias at 12.0 g/dL calculated by linear regression is 0.79 g/dL.

Nonparametric Passing-Bablock regression analysis is not affected by limited range of data. The slope obtained from Passing-Bablock regression analysis is 1.067 and the y-intercept is 0.02 g/dL. This bias between laboratories Commodore and Neptune calculated from these statistics is 0.76 g/dL, very close to that calculated with ordinary least squares linear regression, and very close to the simple bias calculated by subtracting the mean of laboratory Neptune from the mean of laboratory Commodore.

Bias at 12.0 g/dL calculated by ordinary least squares linear regression is 0.79 g/dL.
Bias at 12.0 g/dL calculated by Passing Bablock regression is 0.82 g/dL.

Transferrin Saturation

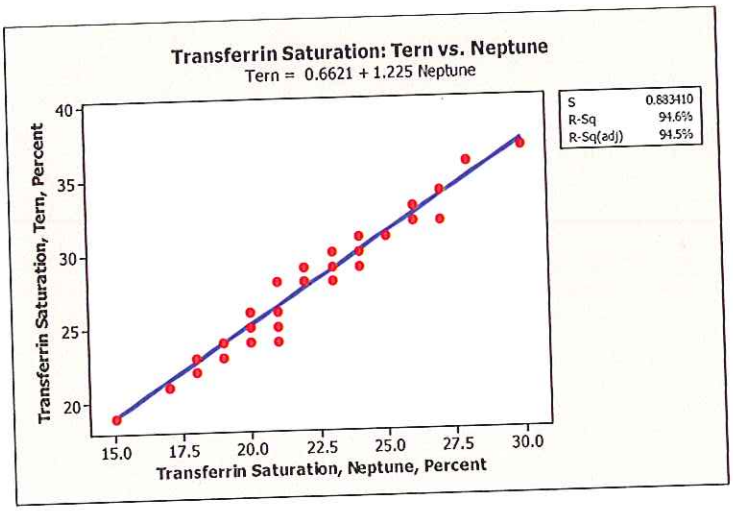
The mean biases of the laboratories from the grand mean for the 50 samples are shown in the chart below.



Paired T-Test and CI: Tern, Neptune

```
Paired T for Tern - Neptune

             N    Mean  StDev  SE Mean
Tern        50  28.320  3.766    0.533
Neptune     50  22.580  2.990    0.423
Difference  50   5.740  1.103    0.156

95% CI for mean difference: (5.427, 6.053)
T-Test of mean difference = 0 (vs not = 0): T-Value = 36.80  P-Value = 0.000
```

A plot of transferrin saturation percentages for laboratory Tern vs. laboratory Neptune is shown below.



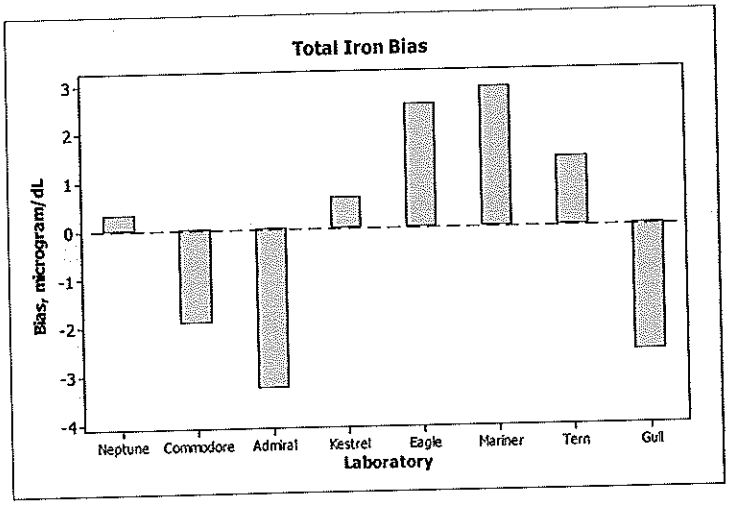
The ordinary least squares linear regression slope and intercept are 1.225 and 0.662 percent, respectively. The correlation coefficient is 0.973, indicating that ordinary least squares linear regression should not be used to estimate bias at concentrations far removed from the means of the data. The bias between laboratories Tern and Neptune calculated by linear regression at 20% TSAT is 4.5%.

The slope obtained from Passing-Bablock regression analysis is 1.250 and the y-intercept is 0.00%. This bias between laboratories Tern and Neptune calculated from these statistics is 5.0%, close to that calculated with ordinary least squares linear regression, and close to the simple bias of 5.7% calculated by subtracting the mean of laboratory Neptune from the mean of laboratory Tern.

Iron

Paired T-Test and CI: Mariner, Admiral

The mean biases of the laboratories from the grand mean for the 50 samples are shown in the chart below.

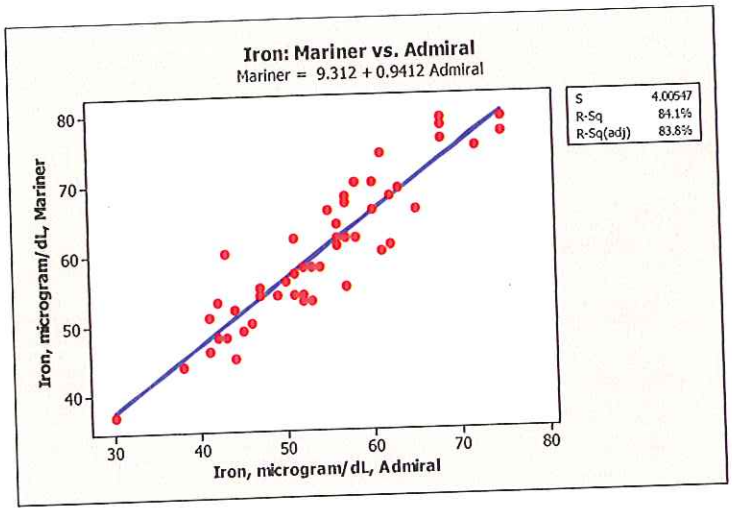


Paired T for Mariner - Admiral

	N	Mean	StDev	SE Mean
Mariner	50	60.12	9.95	1.41
Admiral	50	53.98	9.70	1.37
Difference	50	6.140	4.005	0.566

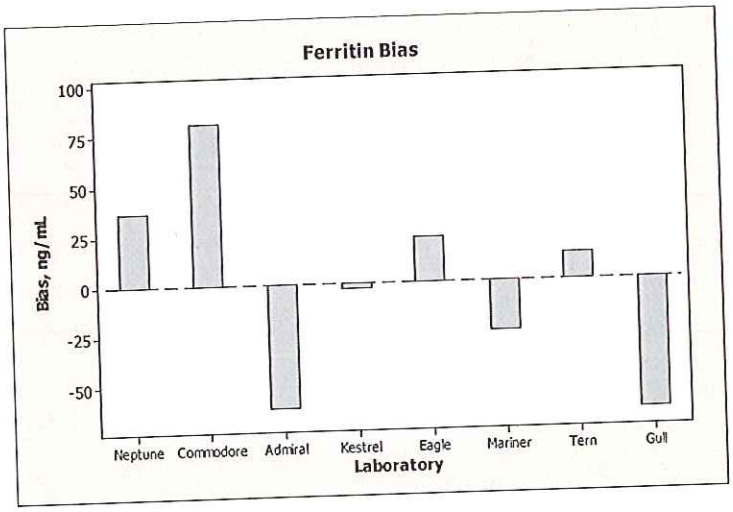
95% CI for mean difference: (5.002, 7.278)
T-Test of mean difference = 0 (vs not = 0): T-Value = 10.84 P-Value = 0.000

A plot of iron concentrations for laboratory Mariner vs. laboratory Admiral is shown below.



Ferritin

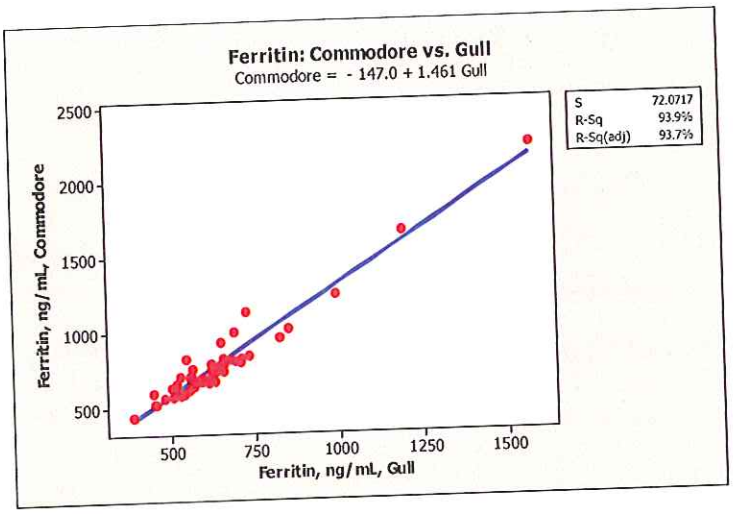
The mean biases of the laboratories from the grand mean for the 50 samples are shown in the chart below.



Paired T-Test and CI: Commodore, Gull

```
Paired T for Commodore - Gull

               N   Mean  StDev  SE Mean
Commodore     50  780.9  288.1     40.7
Gull          50  635.1  191.0     27.0
Difference    50  145.8  113.3     16.0

95% CI for mean difference: (113.6, 178.0)
T-Test of mean difference = 0 (vs not = 0): T-Value = 9.10  P-Value = 0.000
```

A plot of ferritin concentrations for laboratory Commodore vs. laboratory Gull is shown below.

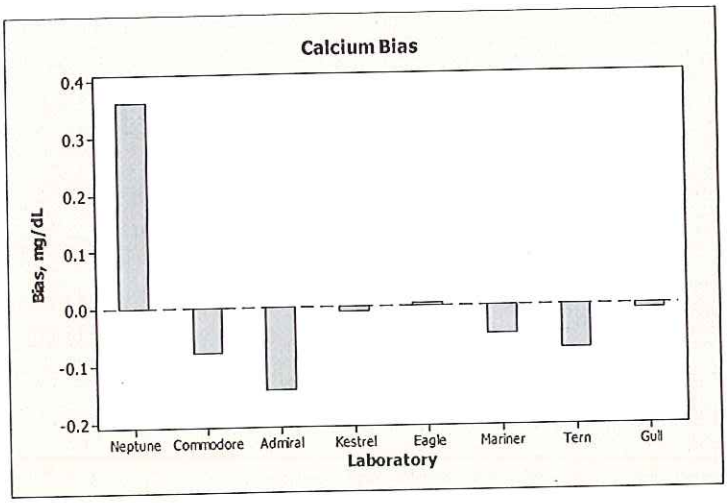


The ordinary least squares linear regression slope and intercept are 1.461 and (-)147.0 ng/mL, respectively. The correlation coefficient is 0.969, indicating that ordinary least squares linear regression should not be used to estimate bias at concentrations far removed from the means of the data. 200 ng/mL is far from the mean of the data. The bias estimated at 200 ng/mL is (-)54.8 ng/mL.

The slope obtained from Passing-Bablock regression analysis is 1.332 and the y-intercept is (-)81.8 ng/mL. This bias between laboratories Gull and Commodore calculated from these statistics is (-)15.4 ng/mL, far from the simple bias of (-)145.8 ng/mL, but close to the bias of (-)54.8 ng/mL calculated from ordinary least squares linear regression.

Calcium

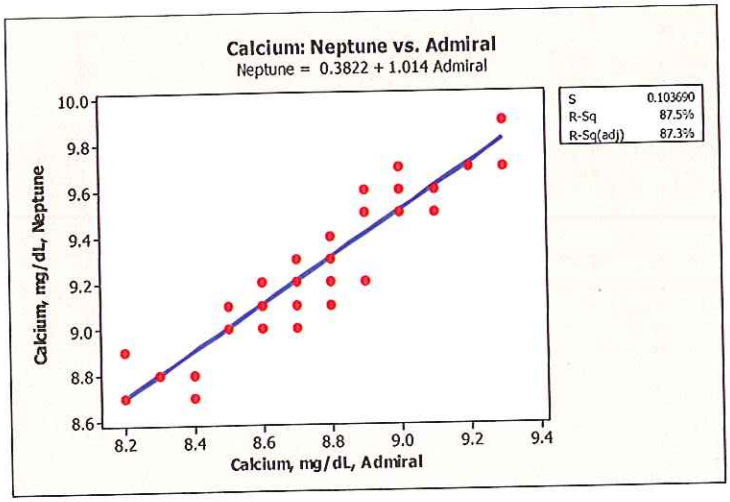
The mean biases of the laboratories from the grand mean for the 50 samples are shown in the chart below.



Paired T-Test and CI: Neptune, Admiral

```
Paired T for Neptune - Admiral

              N     Mean   StDev  SE Mean
Neptune      50   9.2820  0.2905   0.0411
Admiral      50   8.7740  0.2679   0.0379
Difference   50   0.5080  0.1027   0.0145

95% CI for mean difference: (0.4788, 0.5372)
T-Test of mean difference = 0 (vs not = 0): T-Value = 34.98  P-Value = 0.000
```

A plot of calcium concentrations for laboratory Neptune vs. laboratory Admiral is shown below.

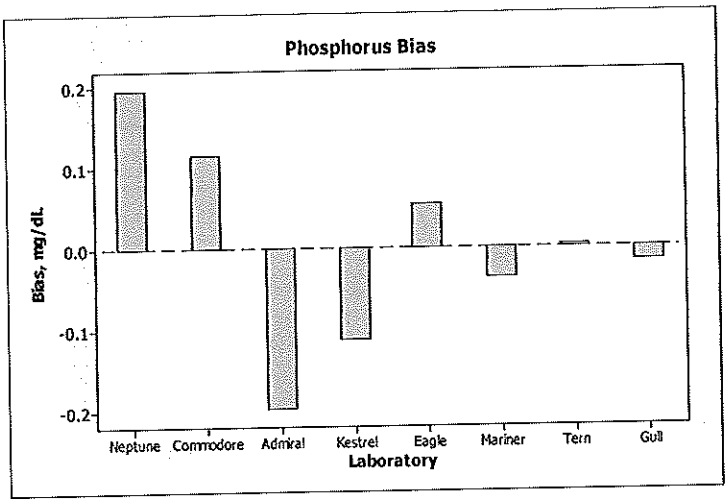


The ordinary least squares linear regression slope and intercept are 1.014 and 0.382 mg/dL, respectively. The correlation coefficient is 0.935, indicating that ordinary least squares linear regression should not be used to estimate bias at concentrations far removed from the means of the data. The bias estimated at 8.4 mg/dL is 0.50 mg/dL, and the bias estimated at 9.5 mg/dL is 0.52 mg/dL.

The slope obtained from Passing-Bablock regression analysis is 1.000 and the y-intercept is 0.5 mg/dL. This bias between laboratories Neptune and Admiral, calculated by Passing-Bablock regression, is 0.50 mg/dL at all concentrations.

Phosphorus

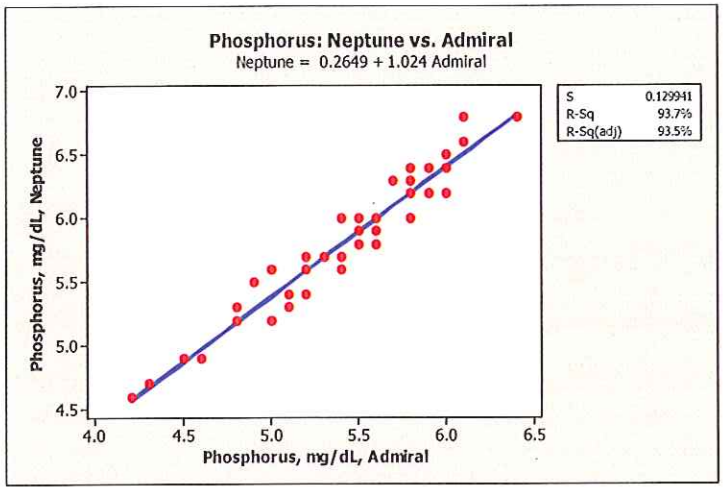
The mean biases of the laboratories from the grand mean for the 50 samples are shown in the chart below.



Paired T-Test and CI: Neptune, Admiral

```
Paired T for Neptune - Admiral

                N     Mean   StDev  SE Mean
Neptune        50   5.7900  0.5108   0.0722
Admiral        50   5.3980  0.4830   0.0683
Difference     50   0.3920  0.1291   0.0183

95% CI for mean difference: (0.3553, 0.4287)
T-Test of mean difference = 0 (vs not = 0): T-Value = 21.47  P-Value = 0.000
```

A plot of phosphorus concentrations for laboratory Neptune vs. laboratory Admiral is shown below.

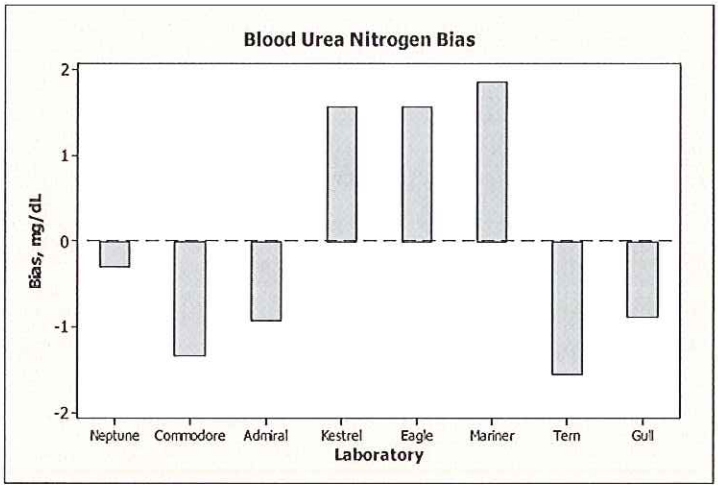


The ordinary least squares linear regression slope and intercept are 1.024 and 0.265 mg/dL, respectively. The correlation coefficient is 0.968, indicating that bias estimated using ordinary least squares linear regression may not be reliable at concentrations far removed from the means of the data. The bias estimated at 3.5 mg/dL is 0.35 mg/dL, and the bias estimated at 5.5 mg/dL is 0.40 mg/dL.

The slope obtained from Passing-Bablock regression analysis is 1.000 and the y-intercept is 0.4 mg/dL. The Passing-Bablock estimate of bias between laboratories Neptune and Admiral, calculated by Passing-Bablock regression, is 0.40 mg/dL at all concentrations.

Blood Urea Nitrogen (BUN)

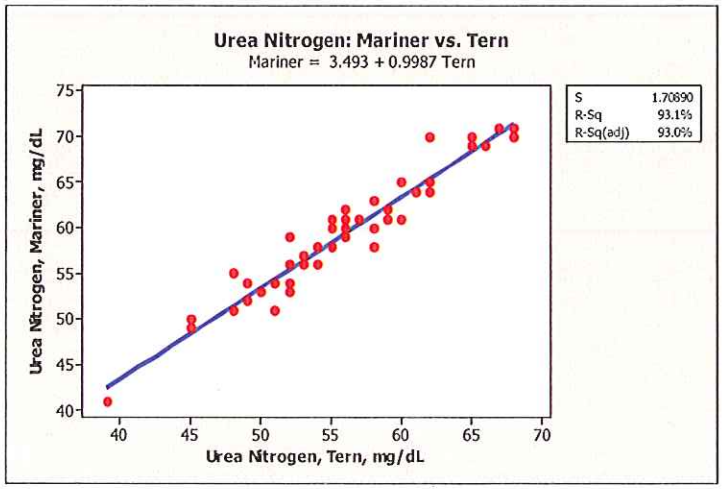
The mean biases of the laboratories from the grand mean for the 50 samples are shown in the chart below.



Paired T-Test and CI: Mariner, Tern

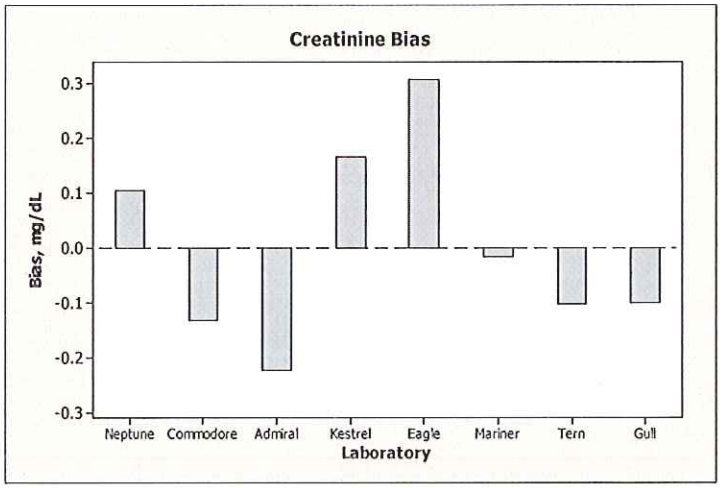
```
Paired T for Mariner - Tern

              N    Mean  StDev  SE Mean
Mariner      50  58.920  6.442    0.911
Tern         50  55.500  6.225    0.880
Difference   50   3.420  1.691    0.239


95% CI for mean difference: (2.939, 3.901)
T-Test of mean difference = 0 (vs not = 0): T-Value = 14.30  P-Value = 0.000
```

A plot of urea nitrogen concentrations from laboratory Mariner vs. laboratory Tern is shown below.



Creatinine

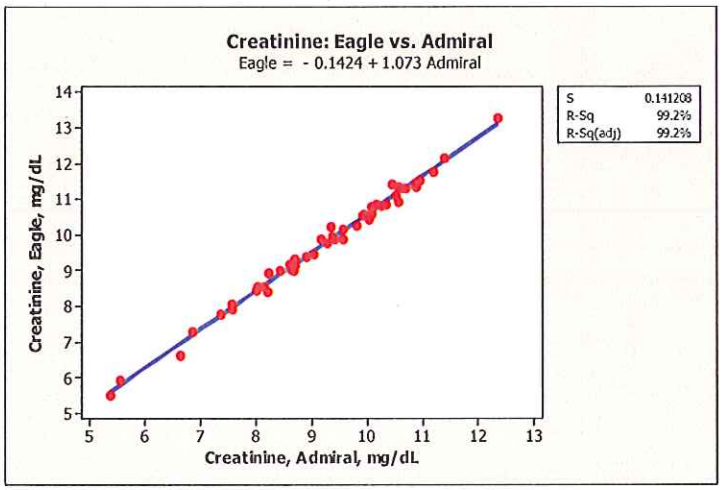
The mean biases of the laboratories from the grand mean for the 50 samples are shown in the chart below.



Paired T-Test and CI: Eagle, Admiral

```
Paired T for Eagle - Admiral

             N    Mean   StDev  SE Mean
Eagle       50   9.799   1.574    0.223
Admiral     50   9.268   1.462    0.207
Difference  50  0.5309  0.1755   0.0248


95% CI for mean difference: (0.4810, 0.5808)
T-Test of mean difference = 0 (vs not = 0): T-Value = 21.39  P-Value = 0.000
```

A plot of creatinine concentrations from laboratory Eagle vs. laboratory Admiral is shown below.



References

1. H.R. 6331: Medicare Improvement for Patients and Providers Act of 2008. http://www.govtrack.us/congress/bill.xpd?bill=h110-6331.

2. National Kidney Foundation Clinical Practice Guidelines for Nutrition in Chronic Renal Failure (NKF KDOQI) Guidelines 2000. Clinical Practice Guidelines for Nutrition in Chronic Renal Failure, Adult Guidelines, Maintenance Dialysis, Evaluation of Protein-Energy Nutritional Status. http://www.kidney.org/professionals/kdoqi/guidelines_updates/nut_a03.html (accessed 5/24/2009).

3. National Kidney Foundation KDOQI Clinical Practice Guideline and Clinical Practice Recommendations for Anemia in Chronic Kidney Disease: 2007 Update of Hemoglobin Target. http://www.kidney.org/professionals/kdoqi/guidelines_anemia/cpr21.htm (accessed May 24, 2009).

4. National Kidney Foundation KDOQI Clinical Practice Guidelines and Clinical Practice Recommendations for Anemia in Chronic Kidney Disease. II. Clinical Practice Guidelines and Clinical Practice Recommendations for Anemia in Chronic Kidney Disease in Adults. http://www.kidney.org/professionals/kdoqi/guidelines_anemia/cpr32.htm (accessed May 24, 2009).

5. National Kidney Foundation KDOQI Clinical Practice Guidelines for Bone Metabolism and Disease in Chronic Kidney Disease. Guideline 6. Serum Calcium and Calcium-Phosphorus Product. http://www.kidney.org/professionals/KDOQI/guidelines_bone/Guide6.htm. (accessed May 25, 2009).

6. National Kidney Foundation KDOQI Clinical Practice Guidelines and Clinical Practice Recommendations for Anemia in Chronic Kidney Disease. Guideline 3. Evaluation of Serum Phosphorus Levels. http://www.kidney.org/professionals/KDOQI/guidelines_bone/Guide3.htm. (accessed May 25, 2009).

7. Westgard JO, Carey RN, Westgard SA: Comparability of Laboratory Tests, Westgard QC, Madison, WI, prepared for Spectra Laboratories, 2008.

8. Passing H, Bablock W: Comparison of several regression procedures for method comparison studies and determination of sample sizes. J Clin Chem Clin Biochem 22:431, 1984.

9. Health Care Financing Administration (42 CFR Part 493, et al), the Public Health Service, U.S. Department of Health and Human Services: Clinical Laboratory Improvement Amendments of 1988, Final Rule, Federal Register 57:7003, 1992.

10. Ricos C, Alvarez V, Cava F, Garcia-Lario JV, Hernandez A, Jimenez CV, Minchinela J, Perich C, Simon M. "Current databases on biologic variation: pros, cons and progress." Scand J Clin Lab Invest 1999;59:491-500. Summarized at http://www.westgard.com/biodatabase1.htm.

11. Davies OL, Goldsmith PL. Statistical Methods in Research and Production, 4th Edition, New York, Hafner, 1972, 49-50.

12. National Kidney Foundation Kidney disease Outcomes Quality Indicators (NKF K/DOQI) Guidelines 2000. C. Appendices (Adult Guidelines); Appendix I. Methods for measuring serum albumin. http://www.kidney.org/professionals/kdoqi/guidelines_updates/nut_appx01a.html (accessed 5/24/2009).

ATTACHMENT B

DEPARTMENT OF HEALTH & HUMAN SERVICES
Centers for Medicare & Medicaid Services
7500 Security Boulevard, Mail Stop S3-02-01
Baltimore, Maryland 21244-1850

CMS
CENTERS for MEDICARE & MEDICAID SERVICES

Center for Clinical Standards and Quality

NOV 0 5 2012

Franklin W. Maddux
Chief Medical Officer and EVP, Clinical and Scientific Affairs
Fresenius Medical Care, North America
920 Winter Street
Waltham, MA 02451

Eduardo Lacson Jr.
VP, Clinical Science and Epidemiology Research
Fresenius Medical Care, North America
920 Winter Street
Waltham, MA 02451

Dear Drs. Maddux and Lacson,

Thank you for your letter regarding modification of laboratory measurement for calcium and phosphorus. We value your interest in and attention to CMS' policies and initiatives. We share your organization's relentless commitment to best serve our ESRD Medicare beneficiaries. We are in receipt of your concerns about the laboratory measurement for calcium and phosphorus levels. Specifically, you raised concerns about the need to accommodate industry accepted standard measurements of both serum and plasma, and to ensure that regulations do not indicate a preference for a particular testing method upon implementation.

We have spoken to our measure developers and clinicians, and tentatively concluded that plasma calcium and plasma phosphorus are an acceptable alternative to serum calcium and serum phosphorus. We intend to collaborate with the National Quality Forum (NQF) to independently review our conclusions before possible future incorporation into our measures. We believe that NQF's independent review provides a transparent process that incorporates input from the entire stakeholder community. The NQF conducts reviews of measure updates and measure maintenance each year, and we plan on bringing this issue to their attention for their review.

As valued partners in this work, we will consider your concerns as we move into the next payment year. Also, as we are actively engaged in development of the proposed rule for the 2016 payment year, we encourage you to use the public comment period that will be provided to raise this issue, should you have continued concerns.

As always, CMS continues to encourage the sharing of your expertise and insights regarding our important work with the ESRD community. If you have any additional concerns, please feel free to contact me.

Sincerely,

Patrick Conway, MD

Patrick Conway, M.D., M.Sc.
CMS Chief Medical Officer
Director, Center for Clinical Standards and Quality

From: Lisa McGonigal [mailto:lmcgon@msn.com]
Sent: Monday, March 18, 2013 12:31 PM
To: Karen Pace
Subject: Additional Plasma Phosphorus Info
Importance: High

Karen - here is some additional data shared by ASCEND Laboratories on the serum vs. plasma phosphorus comparability. Please see their email below, as well.

Let me know if you need anything else - otherwise, have a great Monday!

Best,

Lisa

Lisa McGonigal, MD, MPH
Healthcare Quality Consultant

Phone: 203-298-0567
Cell: 203-530-9524
Fax: 203-549-0796
Email: lmcgon@msn.com

Dr. Lacson,

Per our telephone conversation I have attached data gathered at our laboratory that was used to validate our switch from serum to plasma in 2005. Please feel free to share this information.

- The method comparison study was done on 54 ESRD patients that had samples collected over a two day period. Serum and plasma tubes were analyzed on the same instrument over a two day period. Plasma results had an average bias (-0.111 mg/dL) that was slightly lower than serum. This finding agrees with literature cited by Donald S. Young, MD, PhD in his book "Effects of Preanalytical Variables on Clinical Laboratory Tests" Second Edition 1997 AACC Press that indicates that serum phosphorus results are increased when compared to plasma due to release from cells with clotting.[1]

1. Lum G, Gambino SR. Serum versus plasma determinations in routine chemistry. *Clin Chem*, 18, 710 (1972)

- Later in 2010 we compared Phosphorus results from all patients tested by Ascend (formerly Satellite Labs) and compared that data to results from the 2008 ESRD CPM's.
    - The U.S. mean Phosphorus was 5.4 with 52% of the patients with a mean between 3.5 and 5.5
    - The Ascend mean Phosphorus was 5.2 with 49% of the patients with a mean between 3.5 and 5.5

- The method Ascend uses for Phosphorus testing is FDA approved for both serum and plasma

Hope this helps. Please let me know if you need additional information.

Marty

Martin Blair, CLS, MT (ASCP), MS, MBA
Vice President, Quality Assurance
O: 650.780.5521
M: 650.208.9455
F: 650.556.2271
blairm@ascendclinical.com

Ascend Clinical
1400 Industrial Way
Redwood City, CA 94063

(See attached file: Phosphorus Serum vs Plasma 12_30_2003.pdf)(See attached file: US ESRD Jan 2010_Phosphorus.docx)(See attached file: EN_INORGANIC PHOSPHORUS.pdf)

EP Evaluator

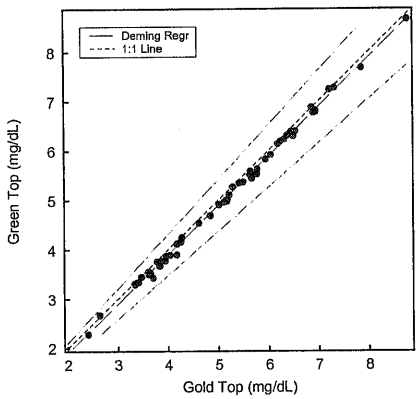
PHOSPHORUS

Clinical Laboratory -- Satellite Laboratory Services

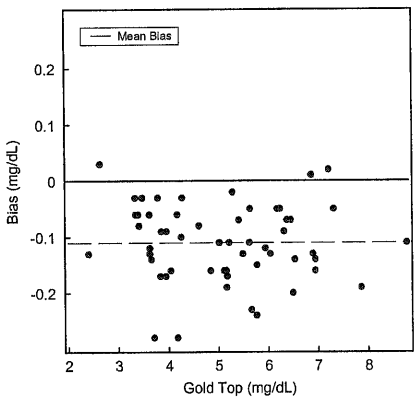
Alternate Method Comparison

X Method: Gold Top **Y Method: Green Top**

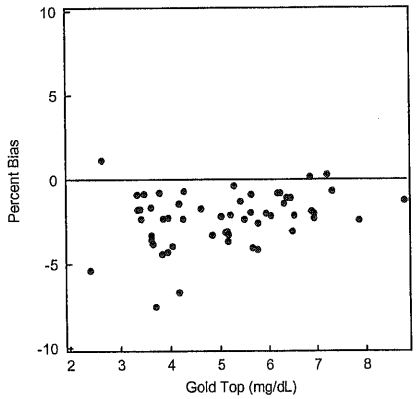
Scatter Plot



Bias



Percent Bias



Regression Analysis

	Deming	Regular
Slope:	0.998 (0.985 to 1.011)	0.997 (0.984 to 1.010)
Intercept:	-0.102 (-0.172 to -0.032)	-0.096 (-0.166 to -0.026)
Std Err Est:	0.069	0.069

95% Confidence Intervals are shown in parentheses

Supporting Statistics

Corr Coef (R):	0.9988	T Probability:	<0.001
Bias:	-0.111	Degrees Freedom:	54
XMean ± SD:	5.111 ± 1.410	SubRange Bounds:	None
YMean ± SD:	5.000 ± 1.408	Points (Plotted/Total):	56/56
Std Dev Diffs:	0.068	Outliers:	None
Paired T Test:	12.11	Scatter Plot Bounds:	Allowable Error

Experiment Description

	X Method	Y Method
ExptDate:	30 Dec 2003	30 Dec 2003
Rep SD:	1	1
Result Ranges:	2.41 to 8.78	2.28 to 8.67
Units:	mg/dL	mg/dL
Analyst:	Marty Blair	Marty Blair
Comment:		

Accepted by: _____ _____

Signature Date

Alternate Method Comparison

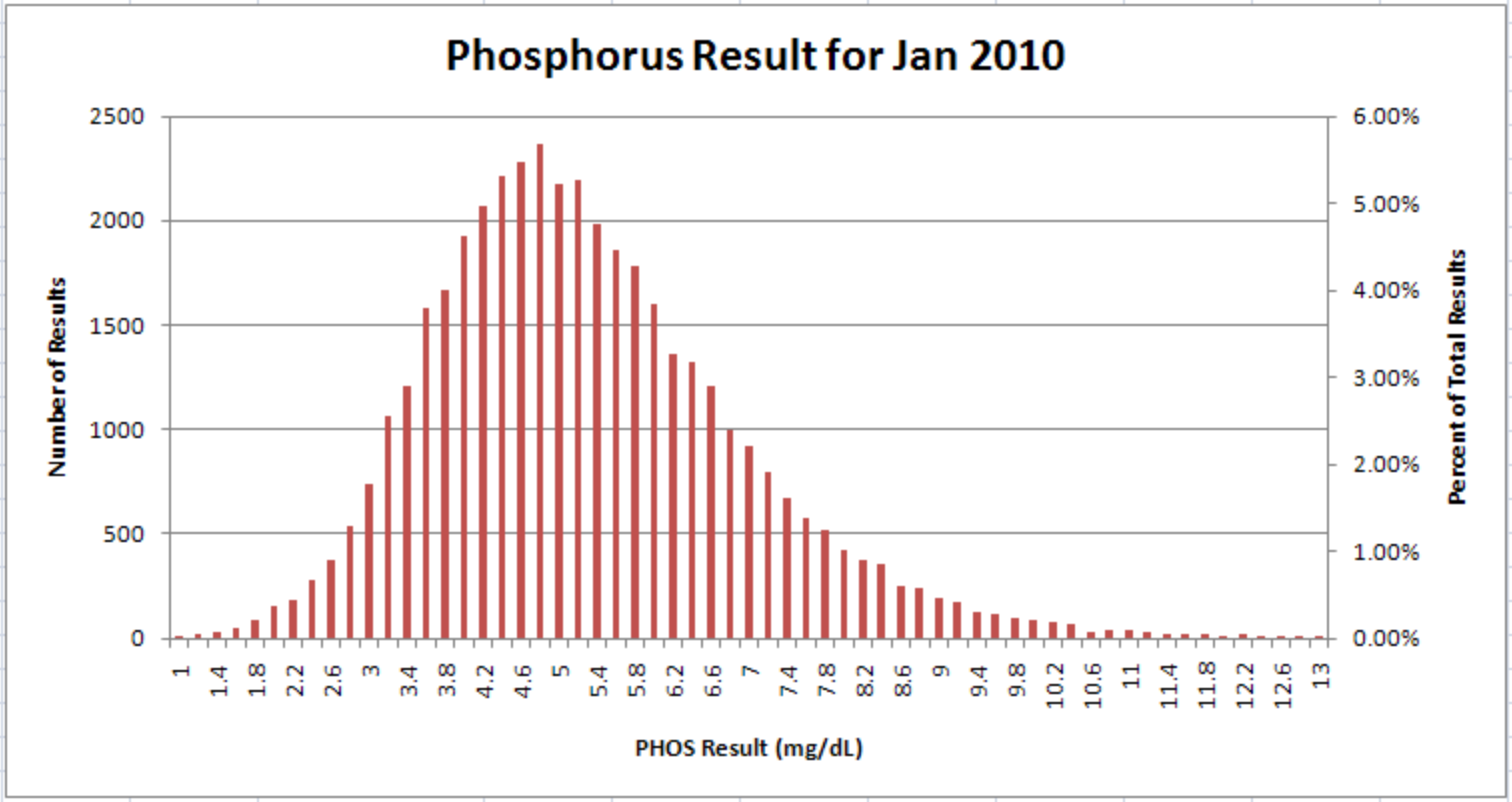
X Method: Gold Top **Y Method: Green Top**

Experimental Results

Specimen	X	Y	Bias	Specimen	X	Y	Bias	Specimen	X	Y	Bias
S00001	2.41	2.28	-0.13	S00020	6.54	6.4	-0.14	S00039	3.63	3.5	-0.13
S00002	3.63	3.51	-0.12	S00021	5.17	4.98	-0.19	S00040	5.63	5.52	-0.11
S00003	5.03	4.92	-0.11	S00022	5.77	5.53	-0.24	S00041	2.64	2.67	0.03
S00004	6.32	6.23	-0.09	S00023	5.78	5.63	-0.15	S00042	5.29	5.27	-0.02
S00005	4.85	4.69	-0.16	S00024	5.16	5	-0.16	S00043	4.18	3.9	-0.28
S00006	3.71	3.43	-0.28	S00025	5.22	5.11	-0.11	S00044	6.5	6.3	-0.2
S00007	4.62	4.54	-0.08	S00026	5.5	5.37	-0.13	S00045	6.46	6.39	-0.07
S00008	4.18	4.12	-0.06	S00027	4.28	4.25	-0.03	S00046	3.4	3.34	-0.06
S00009	5.02	4.91	-0.11	S00028	3.62	3.56	-0.06	S00047	5.95	5.83	-0.12
S00010	3.35	3.29	-0.06	S00029	3.84	3.67	-0.17	S00048	5.67	5.44	-0.23
S00011	3.95	3.78	-0.17	S00030	6.24	6.19	-0.05	S00049	3.96	3.87	-0.09
S00012	4.05	3.89	-0.16	S00031	3.34	3.31	-0.03	S00050	3.42	3.34	-0.08
S00013	7.87	7.68	-0.19	S00032	6.19	6.14	-0.05	S00051	8.78	8.67	-0.11
S00014	6.95	6.81	-0.14	S00033	6.87	6.88	0.01	S00052	6.9	6.77	-0.13
S00015	5.18	5.01	-0.17	S00034	5.12	4.96	-0.16	S00053	5.64	5.59	-0.05
S00016	3.48	3.45	-0.03	S00035	7.32	7.27	-0.05	S00054	6.05	5.92	-0.13
S00017	4.26	4.16	-0.10	S00036	6.38	6.31	-0.07	S00055	3.66	3.52	-0.14
S00018	3.86	3.77	-0.09	S00037	7.22	7.24	0.02	S00056	5.42	5.35	-0.07
S00019	3.79	3.76	-0.03	S00038	6.95	6.79	-0.16				

Values with an "X" were excluded from the calculations. Outliers "O" were also excluded.

All Phosphorus Results from U.S. ESRD Patients Tested by Satellite Labs



Month		
Count	39913	
Mean	5.2	
Median	5.0	
Std Dev	1.7	
25% Percentile	4.1	
75% Percentile	6.2	
	US netwo	SLS Population
%pt 3.5-5.5	52%	49%

INORGANIC PHOSPHOROUS

OSR6122	**4 x 15 mL**	**R1**
	4 x 15 mL	**R2**
OSR6222	**4 x 40 mL**	**R1**
	4 x 40 mL	**R2**

Intended Use
Photometric UV test for the quantitative determination of inorganic phosphorous in human serum, plasma and urine on Beckman Coulter analysers. For *in vitro* diagnostic use only.

Summary[1,2,3,4]
In plasma and serum the majority of phosphate exists in the inorganic form (Pi), approximately 15% bound to protein and the remainder in complexed and free forms. Serum phosphate concentrations are dependent on diet and variation in the secretion of hormones such as PTH. Intracellularly phosphate occurs primarily as organic phosphate however a small but extremely important fraction exists as inorganic phosphate which, because it is a substrate for oxidative phosphorylation, participates in reactions concerned with generation of metabolic energy. About 85% of extracellular phosphate occurs in the Pi form as hydroxyapatite thereby playing an important role in bone structure.
Hypophosphataemia (phosphate depletion) is relatively common in hospitalised patients and is found in up to 30% of surgical patients. Hypophosphataemia is caused by a decreased intake or absorption of phosphate such as occurs in Vit D deficiency, malabsorption, use of oral phosphate binders and primary PTH excess; increased excretion such as occurs in secondary PTH excess, post renal transplant and re-feeding starved patients; and from redistribution of phosphate e.g. hyperalimentation, recovery from diabetic ketoacidosis and respitatory alkalosis.
Hyperphosphataemia is caused by increased intake such as occurs in intravenous therapy and phosphate enemas; reduced excretion such as occurs in acute and chronic renal failure, low PTH or resistance to PTH and vitamin D toxicity; and redistribution of phosphate that occurs in tumour lysis, rhabdomyolysis and heat stroke.

Test Principle[5,6]
Inorganic phosphorous reacts with molybdate to form a heteropolyacid complex. The use of a surfactant eliminates the need to prepare a protein free filtrate. The absorbance at 340/380 nm is directly proportional to the inorganic phosphorous concentration in the sample.

Reaction principle

$$7\ H_3PO_4 + 12\ (Mo_7O_{24})^{6-} + 72\ H^+ \longrightarrow 7\ H_3PO_4(MoO_3)_{12} + 36\ H_2O$$

Contents, Reagent Composition in the Test
Final concentration of reactive ingredients:

Sulphuric acid	200 mmol/L
Ammoniumheptamolybdate	0.35 mmol/L
Glycine	50 mmol/L
Preservative	

Precautions and Warnings
Hazard Warnings and Risk Phrases:
Irritant. R36/38 Irritating to eyes and skin.

Safety Phrases:
S26, S37, S45, S60: In case of contact with eyes, rinse immediately with plenty of water and seek medical advice. Wear suitable gloves. In case of accident or if you feel unwell, seek medical advice immediately (show the label where possible). This material and its container must be disposed of as hazardous waste.

Dispose of all waste material in accordance with local guidelines.
Refer to Safety Data Sheets for further information.

Reagent Preparation
The reagents are ready for use and can be placed directly on board the instrument.

Storage and Stability
The reagents are stable, unopened, up to the stated expiry date when stored at 2...8°C. Once open, reagents stored on board the instrument are stable for 30 days.

Specimen
Serum and heparinised plasma: Stable in serum for 4 days when stored at 2…8°C and 1 day when stored at 15…25°C.[7]
Strongly haemolysed samples should be avoided.

Urine:[8] Acidified with 6M HCl. Collect timed 24-hour specimen using standard laboratory procedures. Store at 2...8°C.

Test Procedure
Refer to the appropriate User Guide and Setting Sheet for analyser-specific assay instructions for the sample type as listed in the Intended Use statement.

Calibration
Use System Calibrator Cat. No. 66300 for serum application and Urine Calibrator Cat. No. ODC0025 for urine application.
The inorganic phosphorous values of both calibrators are traceable to a Beckman Coulter Master Calibrator.
Recalibrate the assay every 30 days, or when the following occur:
Change in reagent lot or significant shift in control values;
Major preventative maintenance was performed on the analyser or a critical part was replaced.

Quality Control

Controls Cat. No. ODC0003 and ODC0004 or other control materials with values determined by this Beckman Coulter system may be used for the serum application.
Biorad Liquichek Urine Chemistry Controls Cat. No. 397 and 398 or other control materials with values determined by this Beckman Coulter system may be used for the urine application.
Each laboratory should establish its own control frequency however good laboratory practice suggests that controls be tested each day patient samples are tested and each time calibration/blanking is performed.
The results obtained by any individual laboratory may vary from the given mean value. It is therefore recommended that each laboratory generates analyte specific control target values and intervals based on multiple runs according to their requirements. These target values should fall within the corresponding acceptable ranges given in the relevant product literature.
If any trends or sudden shifts in values are detected, review all operating parameters.
Each laboratory should establish guidelines for corrective action to be taken if controls do not recover within the specified limits.

Calculation

The Beckman Coulter analysers automatically compute the inorganic phosphorous concentration of each sample.

Reference Intervals[2]

Serum	Adults	0.81 – 1.45 mmol/L (2.5 – 4.5 mg/dL)
	Children	1.29 – 2.26 mmol/L (4.0 – 7.0 mg/dL)
Urine	On non-restricted diet	12.9 – 42.0 mmol/d (0.4 – 1.3 g/day)

Expected values may vary with age, sex, sample type, diet and geographical location. Each laboratory should verify the transferability of the expected values to its own population, and if necessary determine its own reference interval according to good laboratory practice. For diagnostic purposes, results should always be assessed in conjunction with the patient's medical history, clinical examinations and other findings.

Specific performance characteristics

Data contained within this section is representative of performance on Beckman Coulter systems. Data obtained in your laboratory may differ from these values.

Linearity
The test is linear within a concentration range of 0.32 – 6.40 mmol/L (1 – 20 mg/dL) for serum.
The test is linear within a concentration range of 0 – 113 mmol/L (0 – 350 mg/dL) for urine.

Precision
The following data was obtained on an AU600 using 3 serum pools analysed over 10 days.

n = 60	Within Run		Total	
Mean mmol/L	SD	CV%	SD	CV%
0.96	0.01	1.03	0.01	1.55
1.64	0.01	0.63	0.02	1.33
3.36	0.02	0.61	0.04	1.23

The following data was obtained on an AU640 using 3 urine pools analysed over 20 days.

n = 80	Within Run		Total	
Mean mmol/L	SD	CV%	SD	CV%
9.54	0.13	1.41	0.28	2.99
32.96	0.23	0.71	0.51	1.55
87.35	0.62	0.71	1.14	1.30

Sensitivity
The lowest detectable level using serum settings on an AU600 analyser was estimated at 0.10 mmol/L.
The lowest detectable level using urine settings on an AU2700 analyser was estimated at 0.48 mmol/L.
The lowest detectable level represents the lowest measurable level of Inorganic phosphorous that can be distinguished from zero. It is calculated as the absolute mean plus three standard deviations of 20 replicates of an analyte free sample.

Method Comparison
Patient serum samples were used to compare this Inorganic phosphorous assay on the AU600 against another commercially available inorganic phosphorous assay. Results of linear regression analysis were as follows:

y = 0.968x – 0.055	r = 1.000	n = 118	Sample range = 0.44 – 6.41 mmol/L

Patient urine samples were used to compare this Inorganic phosphorous assay on the AU2700 against another commercially available inorganic phosphorous assay. Results of linear regression analysis were as follows:

y = 0.936x + 0.170	r = 0.999	n = 100	Sample range = 3.42 – 53.07 mmol/L

Interfering Substances
Results of serum studies conducted to evaluate the susceptibility of the method to interference were as follows:
Icterus: Interference less than 3% up to 40 mg/dL or 684 µmol/L bilirubin
Haemolysis: Interference less than 10% up to 3.5 g/L haemoglobin
Lipemia: Interference less than 10% up to 800 mg/dL Intralipid®

Results of urine studies conducted to evaluate the susceptibility of the method to interference were as follows:
Icterus: Interference less than 5% up to 40 mg/dL or 684 µmol/L bilirubin
Haemolysis: Interference less than 5% up to 5 g/L haemoglobin

In very rare cases gammopathy, especially monoclonal IgM (Waldenström's macroglobulinemia), may cause unreliable results.

Refer to Young[9] for further information on interfering substances.

Setting Sheet Footnotes

‡ Dilute samples and Urine Calibrator Cat. No. ODC0025 1:10 with purified H_2O.
# User defined ¤ Analyser default value
† System Calibrator Cat. No.: 66300/ Urine Calibrator Cat. No.: ODC0025
* Values set for working in SI units (mmol/L). To work in mg/dL multiply by 3.1.

BIBLIOGRAPHY

1. Smith AF, Beckett GJ, Walker SW, Rae PWH, eds. Lecture notes on clinical biochemisty, 6th ed.Oxford: Blackwell Science,1998:75pp.
2. Endres DB, Rude RK. Mineral and bone metabolism. In: Burtis CA, Ashwood ER, eds. Tietz textbook of clinical chemistry. Philadelphia: WB Saunders Company, 1999;1406-1441.
3. Fraser D, Jones G, Kooh SW, Radde IC. Calcium and phosphate metabolism. In: Tietz NW, ed. Fundamentals of clinical chemistry. Philadelphia:WB Saunders Company, 1987:706pp.
4. Thomas L. Phosphate. In: Thomas L, ed. Clinical laboratory diagnostics. Use and assessment of clinical laboratory results. Frankfurt/Main: TH-Books Verlagsgesellschaft, 1998:241-247.
5. Daly JA, Ertingshausen G. Direct method for determining inorganic phosphate in serum with the “CentrifiChem”. Clin Chem 1972;18(3):263-5.
6. Gamst O, Try K. Determination of serum-phosphate without deproteinization by ultraviolet spectrophotometry of the phosphomolybdic acid complex. Scand J Clin Lab Invest 1980;40(5):483-6.
7. Ehret W, Heil W, Schmitt Y, Töpfer G, Wisser H, Zawta B, et al. Use of anticoagulants in diagnostic laboratory investigations and stability of blood, plasma and serum samples. WHO/DIL/LAB/99.1 Rev.2:39pp.
8. NCCLS. Urinalysis and collection, transportation, and preservation of urine specimens; approved guideline. NCCLS Document GP16-A2, 2nd ed. Pennsylvania: NCCLS, 2001.
9. Young DS. Effects of drugs on clinical laboratory tests, 5th ed. AACC Press, 2000.

CE

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 #: 0255 NQF Project: Renal Endorsement Maintenance 2011

(for Endorsement Maintenance Review)
Original Endorsement Date: Nov 15, 2007 Most Recent Endorsement Date: Nov 15, 2007

BRIEF MEASURE INFORMATION

De.1 Measure Title: Measurement of Serum Phosphorus Concentration

Co.1.1 Measure Steward: Centers for Medicare & Medicaid Services

De.2 Brief Description of Measure: Percentage of all adult (>= 18 years of age) peritoneal dialysis and hemodialysis patients included in the sample for analysis with serum phosphorus measured at least once within month.

2a1.1 Numerator Statement: Number of adult (>= 18 years of age) dialysis patients included in denominator with serum phosphorus measured at least once within month

2a1.4 Denominator Statement: All adult peritoneal dialysis and hemodialysis patients included in the sample for analysis.

2a1.8 Denominator Exclusions: Transient dialysis patients (in unit < 30 days), pediatric patients and kidney transplant recipients with a functioning graft

1.1 Measure Type: Process
2a1. 25-26 Data Source: Electronic Clinical Data
2a1.33 Level of Analysis: Facility

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*):
N/A

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)*: Renal, Renal : End Stage Renal Disease (ESRD)
De.5 Cross Cutting Areas (*Check all the areas that apply)*: Population Health

1a.1 Demonstrated High Impact Aspect of Healthcare: Affects large numbers, Frequently performed procedure, High resource use, Patient/societal consequences of poor quality

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

1a.3 Summary of Evidence of High Impact *(Provide epidemiologic or resource use data)*:
In healthy individuals, the kidney occupies an integral, multi-faceted role in the maintenance of calcium-phosphorus homeostasis. It follows that abnormalities of calcium-phosphorus regulation are exceedingly common in patients with advanced chronic kidney disease, which, indeed, most data indicate that only 25-35% of dialysis patients are able to maintain calcium in the suggested target range of 8.4-9.5 mg/dL (KDOQI 2003). Numerous studies have demonstrated the impact of prolonged calcium-phosphorus dysregulation on patient morbidity and mortality (KDOQI 2003), which can lead to progressive bone weakness, bone pain and increased susceptibility to fractures, and severe arteriosclerosis that can precipitate strokes, heart attacks, and other adverse cardiac events. Unfortunately, overt symptoms can often remain unmanifested in many but the most extreme disordered states of calcium-phosphorus regulation, which is why routine blood tests are necessary to detect and monitor abnormal states of calcium and phosphorus balance in this especially vulnerable population.

1a.4 Citations for Evidence of High Impact cited in 1a.3: National Kidney Foundation. 2003. "K/DOQI Clinical Practice Guidelines for Bone Metabolism and Disease in Chronic Kidney Disease," American Journal of Kidney Disease, 42(Suppl 3): S17. Found at: http://www.kidney.org/professionals/kdoqi/guidelines_bone/index.htm

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:
Consistent monitoring of phosphorus levels helps ensure regulation of patient morbidity and mortality, including stabilization of bone density, decreased bone pain, fracture prevention and decreased rates of arteriosclerosis and related conditions (e.g., stroke, heart attack). Routine blood tests will also aid in detection of and monitoring for abnormal states phosphorus balance in this especially vulnerable population.

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.*]
The following statistics were generated from January 2010 CROWNWeb clinical data: mean(SD)=0.77(0.19); min=0.00; max=1.00; 25th percentile=0.71; 50th percentile=0.80; 75th percentile=0.88.

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*]
The data reported in 1b.4 were generated from January 2010 CROWNWeb clinical data (3,475 facilities and 293,223 patients).

1b.4 Summary of Data on Disparities by Population Group: [*For Maintenance –Descriptive statistics for performance results for this measure by population group*]
To our knowledge, disparity in care (with respect to measurement of serum phosphorus) is an issue that has neither been systematically explored nor developed. It is unlikely to play a major role since phosphorus measurements are typically included in the routine blood screening covered by Medicare.

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*]
N/A

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)*:
The measure focus is the facility´s process of measuring serum phosphorus each month for ESRD dialysis patients. This process leads to improvement in mortality as follows: Measure serum phosphorus--> Assess value-->Identify problem-->Identify treatment options-->Administer the appropriate treatment-->Patient experiences improvement in mortality.

1c.2-3 Type of Evidence *(Check all that apply)*:
Clinical Practice Guideline, Selected individual studies (rather than entire body of evidence)

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)*:
The body of evidence shows a relationship between prolonged calcium-phosphorus dysregulation and ESRD patient morbidity/mortality, which can lead to progressive bone weakness, bone pain and increased susceptibility to fractures, and severe arteriosclerosis that can precipitate strokes, heart attacks, and other adverse cardiac events.

1c.5 Quantity of Studies in the Body of Evidence *(Total number of studies, not articles)*: 6

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 submitting organization recognizes the opinion-based level of evidence supporting the KDIGO Clinical Practice Guidelines for measurement of serum concentration of phosphorus. As such, the overall quality of the body of evidence or the quality of individual studies is not rated in the KDIGO guidelines. Notwithstanding, research in many studies have observed that abnormalities of serum phosphorus concentration are common in the CKD population and that failure to monitor and correct such abnormalities are strongly associated with morbidity and mortality. Observational studies have shown a consistent adverse association of low serum phosphorus with all-cause mortality. Furthermore, the basic science supports a pathological role of low serum phosphorus and intracellular phosphorus depletion in disturbed cellular function.

1c.7 Consistency of Results across Studies *(Summarize the consistency of the magnitude and direction of the effect):* Serum phosphorus is consistently demonstrated to be an important biomarker, strongly associated with adverse cardiovascular outcomes. In addition, the data from in-vitro and in-vivo animal studies establish the biologic plausibility of the adverse effects of inappropriate levels of serum phosphorus on cardiovascular outcomes. Observational data consistently report an increased level of cardiovascular events and mortality when serum phosphorus rises above the normal range in patients with Stage 5 CKD.

1c.8 Net Benefit *(Provide estimates of effect for benefit/outcome; identify harms addressed and estimates of effect; and net benefit - benefit over harms)*:

Monitoring phosphorus levels in the ESRD population reduces the likelihood that this susceptible population develops hyper- or hypophosphatemia, conditions that the body of evidence shows are strongly linked to adverse cardiovascular outcomes. A meta-analysis of the available literature (47 cohort studies) showed an 18% increase in mortality for every 1-mg/dL increase in serum phosphorus (RR=1.18, 95% CI=1.12-1.25) [30].

1c.9 Grading of Strength/Quality of the Body of Evidence. Has the body of evidence been graded? No

1c.10 If body of evidence graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: N/A

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

1c.12 If other, identify and describe the grading scale with definitions: The body of evidence was not graded.

1c.13 Grade Assigned to the Body of Evidence: N/A

1c.14 Summary of Controversy/Contradictory Evidence: There are numerous observational studies that consistently demonstrate a (positive) correlation between mortality and phosphorus levels. However, to date, there are no randomized control trials that provide strong evidentiary support that would inform healthcare providers as to the best means of achieving appropriate phosphorus levels.

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

1) National Kidney Foundation: K/DOQI Clinical Practice Guidelines for Bone Metabolism and Disease in Chronic Kidney Disease. American Journal of Kidney Disease 2003 42:S1-S202 (suppl 3).
2) Kidney Disease: Improving Global Outcomes (KDIGO) CKD-MBD Work Group: KDIGO Clinical Practice Guideline for the Diagnosis, Evaluation, Prevention, and Treatment of Chronic Kidney Disease-Mineral and Bone Disorder (CKD-MBD). Kidney International 2009 76 (Suppl 113): S1-S130.
3) Block GA, Klassen PS, Lazarus JM, et al. Mineral metabolism, mortality, and morbidity in maintenance hemodialysis. Journal of the American Society of Nephrology : JASN 2004 15:2208-18.
4) Young EW, Albert JM, Satayathum S, et al. Predictors and consequences of altered mineral metabolism: the Dialysis Outcomes and Practice Patterns Study. Kidney international 2005 67:1179-87.
5) Kalantar-Zadeh K, Kuwae N, Regidor DL, et al. Survival predictability of time-varying indicators of bone disease in maintenance hemodialysis patients. Kidney international 2006 70:771-80.
6) Kimata N, Albert JM, Akiba T, et al. Association of mineral metabolism factors with all-cause and cardiovascular mortality in hemodialysis patients: the Japan dialysis outcomes and practice patterns study. Hemodialysis international. International Symposium on Home Hemodialysis 2007 11:340-8.
7) Tentori F, Blayney MJ, Albert JM, et al. Mortality risk for dialysis patients with different levels of serum calcium, phosphorus, and PTH: the Dialysis Outcomes and Practice Patterns Study (DOPPS). American journal of kidney diseases : the official journal of the National Kidney Foundation 2008 52:519-30.
8) Chertow G.M., Raggi P., Chasan-Taber S., Bommer J., Holzer H., Burke S.K. Determinants of progressive vascular calcification in haemodialysis patients. Nephrology Dialysis Transplantation 2004 19 (6), pp. 1489-1496.
9) Dhingra R, Sullivan LM, Fox CS, Wang TJ, D´Agostino RB Sr, Gaziano JM, Vasan RS: Relations of serum phosphorus and calcium levels to the incidence of cardiovascular disease in the community. Arch Intern Med 2007 167: 879–885.
10) Wang AY, Lam CW, Wang M, Chan IH, Lui SF, Sanderson JE. Is valvular calcification a part of the missing link between residual kidney function and cardiac hypertrophy in peritoneal dialysis patients? Clinical journal of the American Society of Nephrology 2009 4:1629-36.
11) Ketteler M, Schlieper G, Floege J. Calcification and cardiovascular health: new insights into an old phenomenon. Hypertension 2006 47:1027–1034.Giachelli CM. Vascular calcification mechanisms. Journal of the American Society of Nephrology : JASN 2004 15:2959–2964.
12) Yang H, Curinga G, Giachelli CM. Elevated extracellular calcium levels induce smooth muscle cell matrix mineralization in vitro. Kidney Int. 2004;66(6):2293–2299.
13) Gauci C, Moranne O, Fouqueray B et al: Pitfalls of measuring total blood calcium in patients with CKD. Journal of the American Society of Nephrology 2008;1592-1598.
14) Foley RN, Parfrey PS, Harnett JD, et al. Hypocalcemia, morbidity, and mortality in end-stage renal disease. American journal of

nephrology 1996 16:386-93.
15) Koch M, Lund R, Oldemeyer B, Meares AJ, Dunlay R. Refeeding hypophosphatemia in a chronically hyperphosphatemic hemodialysis patient. Nephron 2000;86(4):552.
16) Travis SF, Sugerman HJ, Ruberg RL, Dudrick SJ, Delivoria-Papadopoulos M, Miller LD, Oski FA. Alterations of red-cell glycolytic intermediates and oxygen transport as a consequence of hypophosphatemia in patients receiving intravenous hyperalimentation. N Engl J Med. 1971 Sep 30;285(14):763-8.
17) Knochel JP. The pathophysiology and clinical characteristics of severe hypophosphatemia. Arch Intern Med. 1977 Feb;137(2):203-20.
18) Marinella MA. The refeeding syndrome and hypophosphatemia. Nutr Rev. 2003 Sep;61(9):320-3.
19) Lindsay RM; Daily/Nocturnal Dialysis Study Group. The London, Ontario, Daily/Nocturnal Hemodialysis Study. Semin Dial. 2004 Mar-Apr;17(2):85-91.
20) Walsh M, Manns BJ, Klarenbach S, Tonelli M, Hemmelgarn B, Culleton B. The effects of nocturnal compared with conventional hemodialysis on mineral metabolism: A randomized-controlled trial. Hemodial Int. 2009 Dec 22.
21) Drechsler C, Krane V, Grootendorst DC, et al. The association between parathyroid hormone and mortality in dialysis patients is modified by wasting. Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association 2009 24:3151-7.
22) Gao P, D´Amour P: Evolution of the parathyroid hormone (PTH) assay--importance of circulating PTH immunoheterogeneity and of its regulation. Clinical Laboratory 51(1-2):21-9, 2005.
23) Souberbielle JC, Boutten A, Carlier MC et al. Inter-method variability in PTH measurement: implication for the care of CKD patients. Kidney International 70(2):345-50, 2006.
24) Souberbielle JC, Roth H, Fouque DP. Parathyroid hormone measurement in CKD. Kidney International 2010 Jan;77(2):93-100.
25) Glassock RJ, Pecoits-Filho R, Barberato SH. Left ventricular mass in chronic kidney disease and ESRD. Clin J Am Soc Nephrol. 2009 Dec;4 Suppl 1:S79-91.
26) Genovesi S, Pogliani D, Faini A, Valsecchi MG, Riva A, Stefani F, Acquistapace I, Stella A, Bonforte G, DeVecchi A, DeCristofaro V, Buccianti G, Vincenti A. Prevalence of atrial fibrillation and associated factors in a population of long-term hemodialysis patients. American Journal of Kidney Disease 2005 Nov;46(5):897-902.
27) Vazquez E, Sanchez-Perales C, Garcia-Garcia F, Castellano P, Garcia-Cortes MJ, Liebana A, Lozano C. Atrial fibrillation in incident dialysis patients. Kidney International 2009 Aug;76(3):324-30.
28) Goodman WG, Goldin J, Kuizon BD et al: Coronary-artery calcification in young adults with end-stage renal disease who are undergoing dialysis. New England Journal of Medicine 2000 342(20):1478-83.
29) Shroff RC, et al. Chronic mineral dysregulation promotes vascular smooth muscle cell adaptation and extracellular matrix calcification. Journal of the American Society of Nephrology : JASN 2010; 21:103-112.
30) Palmer SC, et al. Serum levels of phosphorus, parathyroid hormone, and calcium and risks of death and cardiovascular disease in individuals with chronic kidney disease: a systematic review and meta-analysis. Journal of the American Medical Association : JAMA 2011;305(11):1119-27.

1c.16 Quote verbatim, the specific guideline recommendation *(Including guideline # and/or page #)*:
"3.1.2 In patients with CKD stages 3-5D, it is reasonable to base the frequency of monitoring serum calcium, phosphorus, and PTH on the presence and magnitude of abnormalities, and the rate of progression of CKD. Reasonable monitoring intervals would be:

"...In CKD stages 5, including 5D: for serum calcium and phosphorus, every 1-3 months; and for PTH, every 3-6 months.

"In CKD patients receiving treatments for CKD-MBD, or in whom biochemical abnormalities are identified, it is reasonable to increase the frequency of measurements to monitor for trends and treatment efficacy and side-effects."

1c.17 Clinical Practice Guideline Citation: Kidney Disease: Improving Global Outcomes (KDOGI). KDIGO Clinical Practice Guideline for the Diagnosis, Evaluation, Prevention, and Treatment of Chronic Kidney Disease-Mineral and Bone Disorder (CKD-MBD). In Chapter 3.1: Diagnosis of CKD-MBD: biochemical abnormalities. Kidney International : 2009;76(Suppl 113);S22-S49.

1c.18 National Guideline Clearinghouse or other URL: http://www.kdigo.org/guidelines/mbd/guide3.html#chap31

1c.19 Grading of Strength of Guideline Recommendation. Has the recommendation been graded? No

1c.20 If guideline recommendation graded, identify the entity that graded the evidence including balance of representation

and any disclosures regarding bias:

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

1c.22 If other, identify and describe the grading scale with definitions: The guideline recommendation was not graded.

1c.23 Grade Assigned to the Recommendation: N/A

1c.24 Rationale for Using this Guideline Over Others: No other guidelines are available.

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: High 1c.26 Quality: High1c.27 Consistency: High

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: http://www.arborresearch.org/ESRD_QMS.aspx

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)*:
Number of adult (>= 18 years of age) dialysis patients included in denominator with serum phosphorus measured at least once within month

2a1.2 Numerator Time Window *(The time period in which the target process, condition, event, or outcome is eligible for inclusion)*:
One month

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*:
The numerator comprises all eligible patients who, during the 1-month study period, have a non-missing value in for the variable "Serum Phosphorus"

2a1.4 Denominator Statement *(Brief, narrative description of the target population being measured)*:
All adult peritoneal dialysis and hemodialysis patients included in the sample for analysis.

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)*:
One month

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)*:
The denominator comprises all patients who, during the 1 month study period, have an "Admit Date" prior or equal to the first day of the month; whose "Discharge Date" is blank or greater than or equal to the last day of the month; whose "Primary Type of Treatment" = ´Hemodialysis,´ ´CAPD´ or ´CCPD´ on the last day of the study period; and whose "Primary Dialysis Setting" = ´Dialysis Facility/Center´ on the last day of the Study Period

2a1.8 Denominator Exclusions *(Brief narrative description of exclusions from the target population)*:
Transient dialysis patients (in unit < 30 days), pediatric patients and kidney transplant recipients with a functioning graft

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)*:
We exclude records with an "Admit Date" later than the first day of the study month or with a "Discharge Date" less than the last day of the study month. We also exclude patients whose age is less than 18 years. For all CROWNWeb-collected measures, we make a global exclusion for patients not on either HD or PD, which includes kidney transplant recipients with a functioning graft.

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)*:
N/A

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.)*:
N/A

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.)*:
1. Using CROWNWeb-reported data (data stored as SAS files), identify the number of adult HD and PD patients under the care of a facility.
2. From this group, remove patients who were not in the facility for the entirety of the month (i.e., transient patients).
3. To form the denominator, remove from this group any kidney transplant recipients with a functioning graft.
4. To form the numerator, remove all denominator-eligible patients who do not have a serum phosphorus (variable name, "phosphorus") measurement for the study month.
5. Calculate the facility´s rate of serum phosphorus measurement by dividing the number calculated in Step 3 (the denominator) by

the number calculated in Step 4 (the numerator).

2a1.21-23 Calculation Algorithm/Measure Logic Diagram URL or attachment:
Attachment
Phos_Calculation_Flowchart.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):
N/A

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

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.*): CROWNWeb

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

2a1.30-32 Data Dictionary/Code Table Web Page URL or Attachment:
URL
http://projectcrownweb.org/crown/index.php?page=Public_Documents&subPage=Release_Documents

2a1.33 Level of Analysis (*Check the levels of analysis for which the measure is specified and tested)*: Facility

2a1.34-35 Care Setting *(Check all the settings for which the measure is specified and tested)*: Dialysis Facility

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)*:
We analyzed CROWNWeb data from July 2009 - October 2010. The number of facilities ranged from 3393 - 3581; the total number of patients per month ranged from 263,430 - 330,187.

2a2.2 Analytic Method *(Describe method of reliability testing & rationale)*:
We assessed reliability by calculating facility-level Pearson correlation coefficients between the current performance month and the preceding month for reporting months August 2009 - October 2010.

2a2.3 Testing Results *(Reliability statistics, assessment of adequacy in the context of norms for the test conducted)*:
Reliability of this measure has improved over time. Correlation coefficients ranged from 0.66 to 0.95. The lowest correlation was observed in the first reporting month (August 2009 compared with July 2009). In 2010, correlations from month-to-month were high (range: 0.74-0.95), indicating the data elements for this measure are reliable.

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:
The target population in the validity analysis comprised all adult, non-transient ESRD patients reported in CROWNWeb in 2009. The population and results from the validity analyses performed were consistent with the evidence provided. The validity analyses showed that relative to facilities with the highest performance scores, the Standardized Mortality Ratio (SMR) increased as performance scores decreased.

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)*:
We used July 2009 - October 2010 CROWNWeb data to calculate monthly performance scores, and 2009 Medicare-paid dialysis claims and the Medical Evidence Form (Form CMS-2728) to calculate the SMR. Documentation regarding the Medicare claims used to calculate the SMR is attached.

2b2.2 Analytic Method *(Describe method of validity testing and rationale; if face validity, describe systematic assessment)*:
We assessed validity using Poisson regresion models to measure the association between facility level quintiles of performance scores and the 2009 SMR (methodology on SMR calculations is attached). Facility-level performance scores were divided into quintiles, and the relative risk (RR) of mortality was calculated for each quintile. The highest quintile represented the reference group. Thus, a RR>1.0 for the lower performance score quintiles would indicate a higher relative risk of mortality.

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)*:
Quintiles of the performance scores were defined as follows:
Q1:0-<74%
Q2:74%-<81%
Q3:81%-<86%
Q4:86%-<90%
Q5:90%-100%
Results from the Poisson model indicated lower performance scores were significantly associated with SMR (p<0.001). Relative risks of mortality was highest in the lowest performance measure quintile (RR=1.17;95% CI: 1.13-1.21). The RR for Q2 was 1.13 (95% CI:1.09-1.17), for Q3 was 1.12 (95% CI:1.08-1.16) and for Q4 was 1.09 (95% CI:1.06-1.13).
These findings confirm the association between frequent (monthly) evaluation of hemodialysis adequacy and improved mortality.

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)*:
CROWNWeb data from July 2009 through October 2010 included up to 3581 facilities per month, with an average of 86 patients per facility. The total number of patients per month ranged from 263,430 to 330,187. We excluded patients who were not in the facility for the entirety of the reporting month.

2b3.2 Analytic Method (*Describe type of analysis and rationale for examining exclusions, including exclusion related to patient preference)*:
N/A

2b3.3 Results *(Provide statistical results for analysis of exclusions, e.g., frequency, variability, sensitivity analyses):*
N/A

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)*:
No risk adjustment is performed for this measure.

2b4.2 Analytic Method (*Describe methods and rationale for development and testing of risk model or risk stratification including selection of factors/variables)*:
N/A

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):
N/A

2b4.4 If outcome or resource use measure is not risk adjusted, provide rationale and analyses to justify lack of adjustment: We observed no disparities by population group (see results in Section 1b.4). Furthermore, there is no evidence suggesting this measure should be risk adjusted.

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)*:
We performed analyses using CROWNWeb data from January 2010. There were 3475 facilities and a total of 293,223 patients in this reporting month. Mean number of patients per facility was 84 (SD=52).

2b5.2 Analytic Method *(Describe methods and rationale to identify statistically significant and practically/meaningfully differences in performance)*:
We calculated facility-level rates of monthly serum phosphorus measurements as the number of patients within the facility with serum phosphorus reported divided by the total number of eligible patients in the facility. We also calculated the mean, SD and quartiles.

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)*:
Analysis of CROWNWeb data from January 2010 indicated the mean percentage of patients with a monthly serum phosphorus measurement was 77% (SD=19%). Distribution: Min=0%, Max=100%, 1st quartile=71%, median=80%, 3rd quartile=88%.
These results indicate that on average, facilties are not measuring serum phosphorus in 20% of patients. Furthermore, during this month some facilities measured none of their patients, and up to 25% of facilities measured serum phosphorus in only 71% of patients.

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)*:
We used only one data source (CROWNWeb).

2b6.2 Analytic Method *(Describe methods and rationale for testing comparability of scores produced by the different data sources specified in the measure)*:
N/A

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)*:
N/A

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)*: N/A

2c.2 If disparities have been reported/identified (e.g., in 1b), but measure is not specified to detect disparities, please explain:
N/A

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)*: Public Reporting, Quality Improvement (Internal to the specific organization), Quality Improvement with Benchmarking (external benchmarking to multiple organizations)

3.1 Current Use *(Check all that apply; for any that are checked, provide the specific program information in the following questions)*: Public Health/ Disease Surveillance, Regulatory and Accreditation Programs, Quality Improvement with Benchmarking (external benchmarking to multiple organizations), 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.*]
CROWNWeb national rollout is planned for early 2012. Quality measure results will then be evaluated for public reporting, potentially on Medicare's Dialysis Facility Compare website.

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:
Meaningful: Serum phosphorus monitoring in the ESRD population will help ensure reduced mortality and morbidity for these already susceptible patients, many of whom have several comorbidities.
Understandable: Both patients and healthcare providers understand the process of monitoring, as well as the fact that this mineral being out of a "normal" range can cause adverse outcomes. Furthermore, this measure has been reported in previous ESRD CPM Annual Reports (publicly available).

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): N/A

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*].
N/A

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:
Although this measure is not currently used in a quality improvement program, it has previously been included in ESRD CPM Annual Reports. The ESRD CPM Project was a national effort designed by CMS to assist dialysis providers to improve patiet care and outcomes.

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, Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims), Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

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:
There are no significant potential barriers to retrieving the needed data, and there are no data availability issues.

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)*:
Because this measure has been collected for several years as part of the CPM project, facilities are familiar with the data required for this measure, and data are readily available. It is unlikely that data elements will be susceptible to inaccuracies or errors.

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:
0261 : Measurement of Serum Calcium Concentration

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)*:

CONTACT INFORMATION
Co.1 Measure Steward (Intellectual Property Owner): Centers for Medicare & Medicaid Services, 7500 Security Boulevard , Mail Stop S3-01-02, Baltimore, Maryland, 21244-1850 Co.2 Point of Contact: Edward Q., Garcia III, MHS, Health Policy Analyst, MMSNQF@hsag.com, 410-786-6738-
Co.3 Measure Developer if different from Measure Steward: Arbor Research Collaborative for Health/University of Michigan Kidney Epidemiology & Cost Center, 340 East Huron Street, Ste 300, Ann Arbor, Michigan, 48104 Co.4 Point of Contact: Claudia, Dahlerus, claudia.dahlerus@arborresearch.org, 734-665-4108-
Co.5 Submitter: Claudia, Dahlerus, claudia.dahlerus@arborresearch.org, 734-665-4108-, Arbor Research Collaborative for Health/University of Michigan Kidney Epidemiology & Cost Center
Co.6 Additional organizations that sponsored/participated in measure development:
Co.7 Public Contact: ESRD Quality Measures, Help Desk, ESRD_Quality_Measures@ArborResearch.org, 877-665-1680-, Arbor Research Collaborative for Health

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. TEP convened for original measure submission in 2006: Matt Howard (ESRD Network 15) Raynel Kinney (ESRD Network 9) Chris Lovell (DCI) Norma Ofsthun (Fresenius)
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: N/A
Measure Developer/Steward Updates and Ongoing Maintenance Ad.3 Year the measure was first released: 2008 Ad.4 Month and Year of most recent revision: Ad.5 What is your frequency for review/update of this measure? Every 3 years Ad.6 When is the next scheduled review/update for this measure? 06, 2013
Ad.7 Copyright statement:

Ad.8 Disclaimers:
Ad.9 Additional Information/Comments: This form was revised on November 17, 2011. The items revised were 1c.6, 1c.7, and 3.1.
Date of Submission (*MM/DD/YY*): 06/23/2011

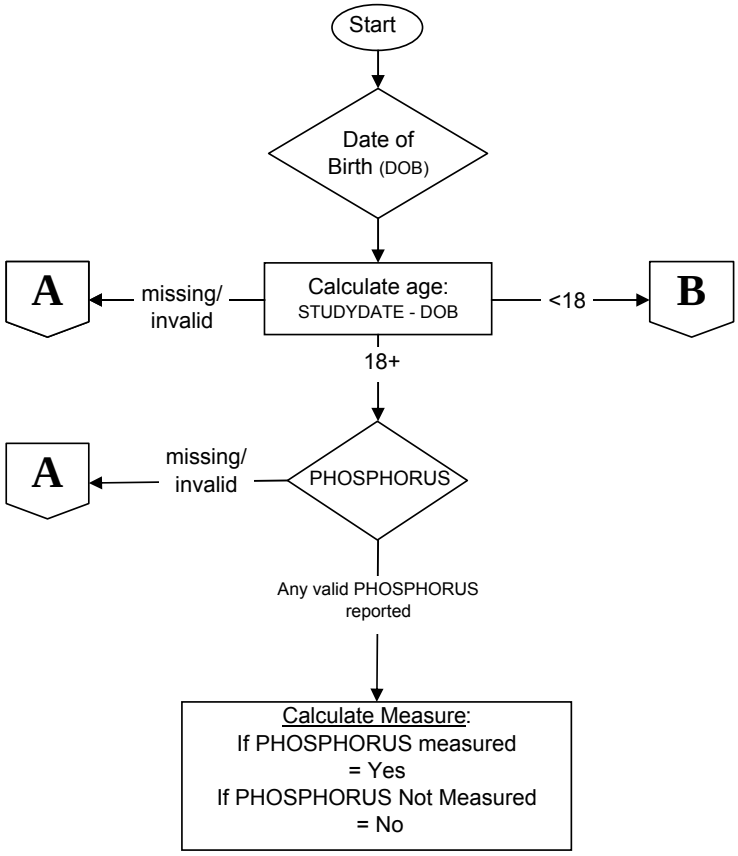
Mineral Metabolism

CPM I: Measurement of Serum Phosphorus

Numerator: Number of adult dialysis patients included in denominator with serum phosphorus measured at least once within the study measurement period. Study measurement periods are 1 month for in-unit HD for a total 3-month study period and 2 months for PD and home HD for a total 6-month study period.

Denominator: All adult (≥ 18 years old) peritoneal dialysis and hemodialysis patients included in the sample for analysis.

Exclusions: Transient dialysis patients (in this center < 30 days), acute HD, pediatric patients and kidney transplant patients.



A Excluded due to missing/invalid data

B Exclude for failing to meet inclusion criteria

Mineral Metabolism

0255 Measurement of Serum Phosphorus Concentration Specifications Submission
Description: Percentage of all adult (>= 18 years of age) peritoneal dialysis and hemodialysis patients included in the sample for analysis with serum phosphorus measured at least once within month. Numerator Statement: Number of adult (>= 18 years of age) dialysis patients included in denominator with serum phosphorus measured at least once within month Denominator Statement: All adult peritoneal dialysis and hemodialysis patients included in the sample for analysis. Exclusions: Transient dialysis patients (in unit < 30 days), pediatric patients and kidney transplant recipients with a functioning graft Adjustment/Stratification: No risk adjustment or risk stratification N/A N/A Level of Analysis: Facility Type of Measure: Process Data Source: Electronic Clinical Data Measure Steward: Centers for Medicare & Medicaid Services
1.Importance to Measure and Report (based on decision logic): Workgroup: Yes Steering Committee: Y-20; N-2 1a. Impact: Workgroup: H-7; M-1; L-1; I-0 1b. Performance Gap: Workgroup: H-0; M-4; L-4; I-1 Rationale: 1a. Impact - Serum phosphorus level has substantial associated clinical consequences. 1b. Performance Gap- The preliminary ratings were spread across all the rating categories. One member questioned whether the performance gap data indicating an average performance of 77% was accurate because most if not all inpatient dialysis facilities are already capturing phosphorus levels of those patients who are treated in the facility. After further discussion, the workgroup agreed that there is a performance gap for this measure. 1c. Evidence (based on decision logic): Workgroup: Yes Quantity: Workgroup: H-3; M-6; L-0; I-0 Quality: Workgroup: H-0; M-6; L-3-; I-0 Consistency: Workgroup: H-3; M-4; L-2; I-0 Rationale: The evidence is indirect, i.e., it is about the association between phosphorus and mortality rather than the frequency of assessment and there was no information submitted about any studies that show a decrease in phosphorus levels will lead to better mortality outcomes. A Committee member noted the inferiority of a measure simply of the frequency of assessment, given the recent NQF guidance on the evaluation criteria. However, the evidence does not support a measure of a specific phosphorus value (also noted by KDIGO). One member noted that the evidence of the association between phosphorus levels and mortality (18% increase in mortality for every 1 mg/dL increase in serum phosphorus) is much stronger than for the association with calcium or PTH. Additionally, the information presented in validity testing demonstrated an association between facility performance on this measure and the facility standardized mortality ratio. While there is excellent evidence correlating phosphorus levels with mortality, there is no evidence that intervention to lower phosphorus levels affects clinical outcomes. Furthermore, there is no evidence that monthly monitoring of phosphorus leads to improved outcomes. Nonetheless, given the absence of such evidence, the preponderance of evidence suggests that very high phosphorus levels should be followed and treated. Several committee members commented that even if one concedes that it should be monitored, there probably is no need to do so on a monthly basis. Another committee member noted that there is no data one way or the other for frequency. Monthly measurement is primarily a function of usual practice because it is paid for on a monthly basis with other lab tests. 2. Scientific Acceptability of Measure Properties (based on decision logic): Workgroup: Yes Steering Committee: Y-19; N-3 2a. Reliability: Workgroup: H-5; M-3; L-1; I-0 2b. Validity: Workgroup: H-3; M-5; L-1; I-0 Rationale: 2a. Reliability The preliminary reliability ratings were mixed, but CMS did submit additional reliability testing that indicated the interunit reliability was 0.94. 2b. Validity – Validity testing demonstrated association between facility performance on this measure and the facility standardized mortality ratio. The lowest quintile of performance on this assessment measure had a 17% greater risk of mortality than the highest performing quintile; and the risk of mortality decreased as the quintile of performance increased. 3. Usability: Workgroup: H-6; M-1; L-2; I-0 Steering Committee: H-8; M-11; L-3; I-0 Rationale: Because of the limitations already noted under evidence, some Committee members did not think this measure would be that useful for evaluating quality.

0255 Measurement of Serum Phosphorus Concentration Specifications Submission
4. Feasibility: Workgroup: H-7; M-2; L-0; I-0 Steering Committee: H-16; M-5; L-1; I-0 Rationale: Phosphorus is measurable and should be relatively easy to get. Assessment of Criteria Met/Suitable for Endorsement: Workgroup: Y-6; N-2 Rationale: One member noted that while it is an important issue, it is going to be measured as a part of a patient's general care plan and should not necessarily be a performance measure. Some Committee members were concerned about misinterpretation of the importance if no measure related to serum phosphorus was recommended. For phosphorus, the correlative data to survival is so remarkably strong that it is important enough to be a performance measure.
Steering Committee Recommendation for Endorsement: Y-19; N-3 Rationale: Phosphorus has the greatest implications for mortality. However, the current state of science does not suggest a measure of intermediate outcome or intervention, so a measure of assessment frequency is the best that could be implemented.
Public and Member Comment Comments included: • endorse only for 2-3 years until replaced with intermediate outcome; • put in reserve status All NQF endorsed measures must undergo evaluation for continued endorsement every 3 years. Whether an intermediate outcome or intervention measure can be developed is dependent on the state of the science to support identifying specific levels that determine optimal care or effective interventions. The measure does not qualify for reserve status because it is not proximal to desired outcomes.

End Stage Renal Disease (ESRD)
Quality Measure Development and Maintenance
Mineral and Bone Disorder Clinical Technical Expert Panel Summary Report
Prepared by: Arbor Research Collaborative for Health and The University of Michigan Kidney Epidemiology and Cost Center
Conducted April 16-17, 2013 in Baltimore, MD
Sent to CMS on June 28, 2013

Contract No. 500-2008-00022I, Task Order No. HHSM-500-T0001

Table of Contents

Technical Expert Panel Summary

The Centers for Medicare & Medicaid Services (CMS) has contracted with Arbor Research Collaborative for Health (Arbor Research) and the University of Michigan Kidney Epidemiology and Cost Center (UM-KECC) to develop End-Stage Renal Disease (ESRD) Quality Measures (QMs) for the following four measure areas:

- Mineral and Bone Disorder
- Hemodialysis Adequacy
- Preventive Care (Pneumococcal, Hepatitis B, and Influenza Vaccinations)
- Dialysis Adequacy for Pediatric Patients (Peritoneal Dialysis Adequacy [PD])

The purpose of the project is to develop measurements that can be used to provide quality care to Medicare beneficiaries.

Technical Expert Panel Objectives

The objectives of these ESRD C-TEPs were described in the charter that was approved by the C-TEPs. The C-TEPs were charged with providing expertise and input to Arbor Research on the development and implementation of measures that will be used to assess and improve the quality of care for Americans with ESRD. The C-TEPs were to provide guidance and assist in the development and specification of new quality measures in specific clinical areas. In addition, the C-TEP members were to consider potential measures using the framework of CMS and the National Quality Forum (NQF). The four evaluation criteria are: *importance*, *scientific acceptability*, *feasibility, and usability*.

Technical Expert Panel Meeting

The Preventive Care, Mineral and Bone Disorder, and Hemodialysis Adequacy TEP met in Baltimore, MD on April 16-17, 2013. The Pediatric Peritoneal Dialysis Adequacy TEP met via conference call on April 11 and April 17, 2013.

The TEPs were comprised of individuals with the following areas of expertise and perspectives:

- Topic Knowledge: ESRD
- Performance Measurement
- Quality Improvement
- Consumer Perspective
- Purchaser Perspective
- Health Care Disparities

The following individuals participated in this TEP:

Name	Title	Organization	Measure Area	Conflict of Interest Disclosure
Patty Danielson, RN BC	Nurse Manager, Behavioral Health	Adventist Medical Center	Mineral and Bone Disorder	None
Kathy Schiro Harvey, MS RD CSR	Nutrition Manager	Puget Sound Kidney Centers	Mineral and Bone Disorder	None
Tamara Isakova, MD	Assistant Professor of Medicine	University of Miami Miller School of Medicine	Mineral and Bone Disorder	None
Mary Leonard, MD MSCE	Professor of Pediatrics and Epidemiology	The Children's Hospital of Philadelphia Research Institute	Mineral and Bone Disorder	Member of Amgen Prolia Scientific Methodological Advisory Committee
Julia Lewis, MD	Professor of Medicine	Vanderbilt University Medical Center	Mineral and Bone Disorder	Current PI for a Keryx-sponsored study; past grants/contracts with industry (self and spouse); investment with Nephrogenix through Vanderbilt
Hartmut Malluche, MD FACP	Professor and Chief, Division of Nephrology, Bone and Mineral Metabolism Department of Internal Medicine	University of Kentucky Medical Center	Mineral and Bone Disorder	Past or current research support from Celgen, Novartis, Vifor
Robin Mauer, FNP MSN	Nurse Practitioner	Washington University School of Medicine	Mineral and Bone Disorder	None
Klemens Meyer, MD	Professor of Medicine	Tufts University Medical Center	Mineral and Bone Disorder	Past grants from Covidien, Parexel ; salary support from DCI

Sharon Moe, MD FASN	Director, Division of Nephrology	Indiana University School of Medicine	Mineral and Bone Disorder	Research support from Genzyme; consultant to Genzyme, Sanofi, Amgen; Litholink, Kai pharma; past research support from Amgen

1. Background

1.1 Overview of measure areas to be discussed

This report summarizes discussions of the in-person CMS ESRD Technical Expert Panel (TEP) meeting held in Baltimore, MD, April 16-17. Prior to the TEP in-person meeting, TEP members reviewed the published clinical evidence (post-2011), clinical guidelines, and existing measures in the areas of mineral and bone disorder such as phosphorus, calcium, parathyroid hormone (PTH), vitamin D, fractures, and falls. In all of these areas, the TEP was charged with evaluating the relative importance of each of these topics in the ESRD population, assessing the evidence supporting potential quality measures in each area, and assessing the feasibility of implementing such measures. Specific topics discussed at the in-person meeting included PTH, dietary counseling, vitamin D, calcium and phosphorus, and areas of specific interest to CMS including serum vs. plasma measurements, corrected vs. uncorrected serum calcium measurements, and bone histology/bone fractures.

2. Corrected versus Uncorrected Serum Calcium in the Hypercalcemia Measure

2.1 Discussion

As requested by CMS, the TEP discussed whether or not they thought it would be acceptable for facilities to report corrected calcium specifically for the NQF endorsed hypercalcemia measure (NQF ID #1454). It was noted that there are 3 options: 1) leave the measure with uncorrected calcium, as it was developed by the 2010 TEP and endorsed by the NQF 2) change the measure to use corrected calcium, or 3) to allow either uncorrected or corrected calcium to be used for calculation of the hypercalcemia measure. In agreement with the 2010 TEP, it was felt that corrected calcium obtained using formulas typically applied in clinical practice may not reflect ionized calcium levels (which are biologically active) better than uncorrected. Some TEP members expressed concerns for the various formulas that are used, since evidence suggests that in ESRD patients they may be inaccurate in estimating ionized calcium concentrations, the gold standard. One TEP member had concerns that the use of various formulas may cause unintended consequences in terms of additional burden related to data collection.

CROWNWeb patient level data from May to November 2012 was used to construct boxplots showing the distribution of corrected vs. uncorrected calcium for May 2012 as well as the percentage of patients in different calcium ranges (<8.4, 8.4-9.5, 9.6-10.2, and >10.2) when using corrected and uncorrected calcium . It was noted by Arbor that the CROWNWeb data and analyses were limited to patients for whom calcium data had been submitted as it was not a required data field. The TEP was shown boxplots of corrected vs. uncorrected serum calcium values at the patient-level where there was roughly a 1% decrease in mean serum calcium values when using uncorrected values instead of corrected values. Based on this distribution, TEP estimated the percentage of patients who are classified as hypercalcemic (i.e. having serum calcium > 10.2 mg/dl) when using corrected and uncorrected calcium. While the difference in the % of patients classified as hypercalcemic using corrected and uncorrected calcium is numerically small, one TEP member questioned whether it could be clinically significant l. There was

some concern that using corrected calcium could over-estimate the number of patients with hypercalcemia in a small number of patients with low albumin. The TEP agreed that some of these issues are technical issues such as which albumin assay was used to obtain corrected calcium rather than issues related to the measure definition itself.

2.2 Recommendations

Despite some of the technical issues described above, all TEP members unanimously recommended to leave the measure unchanged and to retain the current specification for uncorrected calcium. This is the current version of the measure as endorsed by NQF (#1454).

3. Serum versus Plasma Measurements of Phosphorus and Calcium

3.1 Discussion

The TEP was asked by CMS to consider a request submitted to CMS and the NQF regarding the use of plasma versus serum in measurement of phosphorous and calcium, respectively. Specifically, CMS received a letter from the Kidney Care Partners, a renal stakeholder organization, requesting a review of the current NQF endorsed phosphorus measure (#0225) to accept measurement of either serum or plasma phosphorous. This would allow facilities that routinely measure phosphorus on plasma not to be penalized under the current phosphorus reporting measure under the Quality Incentive Program (QIP). CMS felt that it was a reasonable request and deferred a decision on this issue to the MBD TEP.

Published literature assessing the difference in phosphorus levels measures using plasma versus serum indicates that the difference may actually be not negligible (Carothers, 1976). As far as the TEP members could determine, unpublished data provided by Spectra laboratory were conducted on normal volunteers and therefore may not be applicable to the HD population. The TEP had a lot of discussion regarding the preclinical processing issues (e.g. sample handling at the dialysis facility; shipping conditions etc.) that may result in lower phosphorus levels if drawn on plasma and whether or not the tubes were spun or properly handled. A TEP member explained that one study showed that the shipping of tubes, letting them sit for long periods, and not following other pre laboratory technical processing can affect levels of many analyses, including phosphorus. One TEP member agreed and pointed out that the collection, shipping and processing of blood draws bears review because of issues with repeat labs, more blood loss and other factors compromising the integrity of the lab specimen. However, it was felt that this discussion is beyond the expertise and scope of this TEP.

TEP members agreed that they should focus on whether or not to accept measurement of plasma phosphorus. Are there reasons to believe that measurement on plasma would alter the care of the patient (assuming that everything was processed correctly)? It was pointed out that plasma phosphate concentrations are about 10% lower than serum phosphate levels. However considering the current measure does not specify a target range this would not be an immediate issue but may in the future if a specified range would need to be defined. There was caution to accept labs that use a different assay and getting a different laboratory measurement when there is already quite a bit of biological variability based on diet and the time of the day the sample was drawn. Thus, the addition of an additional

potential confounder to the physician's interpretation of the results is premature without additional studies comparing 'real world' collection of samples from a dialysis unit that would evaluate not only the analytic differences but also whether this does indeed improve sample handling and processing mistakes at the level of the dialysis unit.

The TEP discussed the need for another lab-specific TEP that would include CLIA (Clinical Laboratory Improvement Amendments) and/or CAP (College of American Pathology) experts that would understand the laboratory values and focus on the laboratory tests and procedures that are necessary. CMS stressed that it was important to move forward with clinical quality measures because these can affect care at the dialysis facility, while measures of laboratory processes do not directly affect care processes delivered at the dialysis facility. TEP members agreed on this point. It was noted by one TEP member that some labs will only take samples processed correctly, while others will process specimens, even if they could have been compromised at the source or in transit (cracked tubes, limited amount of blood, etc.). Such a panel could agree on minimum standards for the determination of an acceptable sample at a laboratory.

3.2 Recommendations

All TEP members voted and unanimously recommended to keep the measure unchanged – facilities would need to report phosphorus levels measured on serum. The TEP also recommended that this would apply to the measurement of calcium. Specifically, the NQF endorsed hypercalcemia measure (#1454) specifically state measurement of serum calcium, and a previous process measure for calcium (#0261) that was retired by NQF in 2011.

The TEP members also recommended that a lab-specific TEP be convened that included experts in CLIA and CAP for the purpose of defining measures for laboratory tests and procedures that may affect laboratory values.

4. Bone Biopsies

4.1 Discussion

CMS asked the TEP to consider whether there may be a role for bone histology as a quality indicator for the care of dialysis patients.

While most TEP members agreed that bone biopsies are the gold standard to analyze abnormalities in bone histology, they also pointed out that there is no evidence that routine adoption of bone biopsies would result in improved patient outcomes, including quality of life. There is no evidence base of trial data utilizing this technique to guide specific interventions. Moreover, bone biopsies are relatively invasive procedures, are not routinely performed across the country and while in general complication rates are low most patients would not welcome such procedures.

One TEP member felt strongly that bone histology provides important information, especially in selected patients (e.g. those with unexplained bone fractures) and it should be considered the gold standard to guide therapeutic strategies; furthermore, bone biopsies would be cost effective in the long run if

appropriate therapies are chosen based on biopsy results. All TEP members felt that bone biopsies should be considered on a case by case basis and performed as clinically indicated. It was felt that since the indication differs largely on clinical presentation in each patient, it would be nearly impossible to create a quality indicator that applies to the majority of dialysis patients.

The majority of the TEP agreed that based on the available evidence it would not be feasible to create an appropriate quality indicator related to bone histology. While bone biopsies may be needed for a specific patient that is not the case for the vast majority of dialysis patients. One TEP member, however, disagreed and believed it to be feasible; clinicians would need to be trained in performing and interpreting results of bone biopsies, and would have the most appropriate information to decide what the appropriate treatment should be.

4.2 Recommendations

The majority of the TEP members (eight out of nine) recommended that a quality measure for bone biopsies not be developed at this time due to insufficient evidence.

5. Review of Current Mineral and Bone Disorder Measures

5.1 Discussion of existing MBD measures

5.1.1Proportion of patients with hypercalcemia (NQF ID #1454)

None of the TEP members had any issues with the 10.2 threshold but one TEP member expressed concerns with the potential for missing values a calcium measurement – particularly if the patient was a new patient at the unit at the end of the month. The current exclusions to the measures are children, transient patients, and transplant patients with a functioning graft (the hypercalcemia measure also restricts to patients who have been on dialysis for >90 days). A TEP member pointed out that patients who were in the hospital may be an issue. Ideally if a patient is hospitalized labs should be re-drawn when they come back to the dialysis facility and re-start dialyzing; however, it is often the case that monthly labs that were not drawn at the time the patient was in the hospital are not drawn while the patient is discharged from the hospital and back on dialysis.

The TEP also discussed the current exclusions that exist for the hypercalcemia measure (pediatric patients as well as transient patients). One TEP member explained that while different thresholds may be appropriate for pediatric patients, the frequency of measurement should be the same noting that some things should be measured more often but at a minimum should be measured monthly.

There was some general discussion among TEP members about the possibility of restricting the measure to patients that dialyze at least 2 or 3 times in the month. It was pointed out that this is an exclusion that is generally applied to all QIP measures and is not exclusion for the NQF endorsed “Proportion of patients with hypercalcemia” measure. There was concern over some patients that never show up and transient patients (for example patients that go to Florida for several months). One TEP member explained that there is a difference between transient patients that would not be included in the denominator (and would be included in another facility’s denominator after a month of treatment) and

patients that belong to a facility and never show up. Some more discussion regarding the number of months/treatments that a patient should have to be included in the denominator continued.

In reviewing the hypercalcemia measure there was a brief discussion of the age of a pediatric patient (the current hypercalcemia measure uses patients less than or equal to 18 years old). One TEP member mentioned that some consider patients less than the age of 21 to be pediatric. It was decided to leave the exclusion as is (< 18 years).

The TEP attempted to maintain consistency among the measures and wanted to change the hypercalcemia measure to say percentage of patients rather than proportion of patients. There was a discussion because this measure is a little different than the other measures that are process measures in which a three month rolling average is used, that the wording "proportion of patients" should remain as is. There was also a brief discussion of the consistency among the different measures in using the 90 day criterion for patients in a facility in the denominator. It was decided that because the hypercalcemia measure is a 3 month rolling average that it makes sense for a patient to have to be in the facility for 90 days and therefore the TEP decided to not change the language or specifications to the hypercalcemia measure.

5.1.2 Measurement of Serum Phosphorus Concentration (NQF ID #0255)

The TEP was informed that among facilities reporting mineral metabolism data for the months of May-November 2012, the current reporting frequency of phosphorus is high (>99% patients having at least one observation per month). CMS also explained that the measure should make it feasible for facilities to reach 100% but that this issue has surfaced in many other TEPs and that there is not usually a 100% standard for the QIP (the threshold is not set).

In an effort to maintain consistency among the measures, the TEP discussed adding an exclusion for patients who were not on dialysis for at least 90 days. The TEP agreed that unlike the "Proportion of patients with hypercalcemia" measure, there was not the same need to have this exclusion. The TEP did not want the patient to wait 90 days before having a measurement completed for serum phosphorus.

5.1.3 Measurement of Serum Calcium Concentration (NQF ID #0261, previously endorsed by NQF – retired in 2011)

The TEP felt that it was just as important to measure serum calcium in pediatric patients at least once a month and decided to remove the pediatric exclusion criteria from the process measure (measuring serum calcium at least once a month). In contrast, in regards to the hypercalcemia measure, it was felt by the TEP that pediatric patients wouldn't necessarily have 10.2 as an appropriate threshold for hypercalcemia and would suggest leaving the exclusion in the hypercalcemia measure (as noted above).

The TEP discussed the current exclusion for transplant patients with a functioning graft. It was explained that these are patients that have delayed graph viability so need dialysis until the transplant kicks in. A TEP member said that somebody should check the labs anyway for these patients and these patients shouldn't be excluded.

In an effort to maintain consistency among the measures, the TEP discussed adding an exclusion for patients who were not on dialysis for at least 90 days. The TEP agreed that unlike the "Proportion of patients with hypercalcemia" measure, there was not the same need to have this exclusion. The TEP did not want the patient to wait 90 days before having a measurement completed for serum calcium.

5.2 Recommendations

5.2.1 Proposed revisions to the serum calcium and serum phosphorus process measures

The TEP recommended two changes to the current process measures for serum calcium and serum phosphorus:

1. The TEP recommended that pediatric patients (< 18 years) be included in the denominator for the process measures of monthly measurement of serum calcium and serum phosphorus.

2. The TEP also recommended that transplant recipients with a non-functioning graft should be included in the denominator and that the current exclusion of these patients in the two process measures should be taken out.

5.2.1.1 Measurement of serum phosphorus

Percentage of all peritoneal dialysis and hemodialysis patients included in the sample for analysis with serum phosphorus measured at least once within a month.

Numerator: Number of dialysis patients included in denominator with serum phosphorus measured at least once within a month

Denominator: All peritoneal dialysis and hemodialysis patients included in the sample for analysis.

Exclusions: Transient dialysis patients (in unit < 30 days)

5.2.1.2 Measurement of uncorrected serum calcium

Note: The previously NQF endorsed measurement of serum calcium measure was retired in 2011. The TEP recommended this measure be submitted to NQF for endorsement and reinstated for uncorrected serum calcium.

Measure description: Percentage of all peritoneal dialysis and hemodialysis patients included in the sample for analysis with uncorrected serum calcium measured at least once within a month.

Numerator: Number of dialysis patients included in denominator with uncorrected serum calcium measured at least once within a month

Denominator: All peritoneal dialysis and hemodialysis patients included in the sample for analysis.

5.2.2 Proposed revisions to the hypercalcemia measure

As the existing hypercalcemia uses a 3-month rolling average, while the measurement of serum phosphorous and measurement of serum calcium measures specify measurement at least once within

the month. The TEP recommended to revise the current hypercalcemia measure so that monthly calcium measurements would be required as part of the measure specification.

6. Parathyroid Hormone

6.1 Literature Review and scientific importance

There was also a lengthy discussion of the strength of evidence regarding PTH as a risk factor for adverse outcomes, in light of recent randomized trials including EVOLVE (EVOLVE Trial Investigators, 2012) and the ADVANCE study (Raggi 2011) . The TEP was divided on the strength of the evidence but concluded that they are the current strongest bodies of evidence that exist since the 2010 TEP convened.

It was recognized that the previously cited problem with assay variability could be overcome if the unit utilizes the same assay each time. Furthermore, given the normal physiologic oscillations in PTH, measurement should be more often if variability is to be minimized. Other TEP members agreed that the combination of laboratory values (PTH with calcium and phosphorus) may be more predictive of mortality, but since each lab value changes individually, it would be very difficult to make a recommendation based on a combination. .

A few TEP members suggested that phosphorus would be measured at the same time as PTH and mentioned that phosphorus and calcium both contribute to PTH secretion regulation.

6.2 Review of existing measures

The TEP reviewed some of the existing measures (percent of patients on a phosphate binder with iPTH measured within the last 3 months; percent of patients with iPTH greater than 100 pg/mL (or greater than 1.5 times the upper limit of normal for each assay used) and/or phosphorus greater than 4.5 mg/dL and are prescribed a low phosphorus diet for 1 month; percent of patients with one measurement of iPTH) and considered the final decisions of the 2010 TEP which was to not recommend a PTH process measure.

6.3 Recommendations

6.3.1 Proposed measure

Measurement of plasma PTH concentration

Measure description: Percentage of all peritoneal dialysis and hemodialysis patients included in the sample for analysis with plasma PTH measured, together with documentation of the specific PTH assay utilized*, at least once within a 3 month period

Numerator: Peritoneal dialysis and hemodialysis patients included in the sample for analysis with plasma PTH measured, together with documentation of the specific PTH assay utilized, at least once within a 3 month period

Denominator: All peritoneal dialysis and hemodialysis patients included in the sample for analysis

Exclusions: Transient patients (in unit < 30 days)

**This includes type of assay (intact, whole), and assay kit manufacturer*

Measure Criteria:

Importance

- Elevations in PTH are associated with increased morbidity and mortality.
- Therapies in dialysis patients that affect PTH are commonly used and hence, PTH should be monitored.
- Evidence in the EVOLVE and ADVANCE studies (that were published in 2011-2012, and hence not available for the 2010 TEP) demonstrate efficacy of strategies to lower PTH in a priori planned secondary analyses adding to the strength of evidence.

Scientific acceptability

- Prior concerns about assay variability are attenuated by the requirement to report the specific type of assay (e.g., intact, whole, other) and kit manufacturer utilized

Feasibility

- PTH is routinely measured in dialysis facilities

Usability

- Nephrologists routinely measure and interpret PTH levels
- With measurement and assay reporting, clinicians can interpret trends in PTH as per KDIGO recommendations

6.3.1.1 Discussion of numerator and denominator

Initial discussions considered the time period in which PTH should be measured. It was noted that a study showed that with more frequent PTH measurement the value of PTH was lower but there was also a significant increase in calcimimetics and vitamin D analogues (Greenberg 2011). One TEP member noted that increasing the measurements to monthly may make it more of a burden on facilities. CROWNWeb data analyses were examined for PTH including the proportion of patients with various PTH measurements in a 1 month period, 2 month period, 3 month, as well as a 4 month period. It was noted however, that because PTH is not a required data element in CROWNWeb that the results should be interpreted with caution. The rest of the TEP agreed that measuring PTH every 3 months would be sufficient and that it would be more of a burden and expensive to measure PTH monthly. There were, however, concerns about the possibility of increasing drug prescriptions and medication usage and the side effects that may result from this if the frequency of measurement was increased given a single paper that noted this. Conversely, in the studies that demonstrate efficacy of various treatments (calcitriol and its analogs, calcimimetics), that PTH was measured more frequently. A vote was conducted, with one TEP member voting for monthly measurements, and the other TEP members voting for at least 1 measurement in a 3 month period (quarterly). The TEP also decided to remain consistent with the exclusions that were stated in the process measures for calcium and phosphorus in which only transient patients (in the facility for <30 days) would be excluded but pediatric patients (< 18 years) would not be.

There was additional discussion of the definition and distinction between the mandatory versus voluntary fields in CROWNWeb particularly in terms of the feasibility of making PTH a mandatory field, and the feasibility of knowing what assay was used. One TEP member was concerned about the possibility that some facilities may not know what assay is being used. Another TEP member pointed out that it would be preferable for the unit to have an pre-populated field for a facility so that the facility would have one assay and only the patient's PTH level would have to be inserted but that if the facility changed assays (changed laboratories) then that pre-populated field could be changed. TEP members agreed that this would be the most ideal situation in asking a facility to report the assay used for PTH measurements.

6.3.1.2 Importance

A first draft of the importance text was drafted with language input from all TEP members.

Importance: *Elevations in PTH are associated with increased morbidity and mortality. Commonly used therapies in dialysis patients are prescribed to affect PTH and hence PTH should be monitored. Recent evidence in the EVOLVE and ADVANCE studies demonstrate efficacy of strategies to lower PTH in a priori secondary analyses.*

There was a long discussion among TEP members regarding the EVOLVE specifications and the ADVANCE study specifications since these two are the main studies that provide the evidence that treatments aimed at lowering PTH have outcomes of importance (coronary artery calcification and mortality/cardiovascular morbidity). It was concluded that they are the best RTCs available to support the importance of this parameter and although the primary end points of both studies were negative, the importance of the pre-specified secondary end points provides the strongest evidence to back up the importance of the measure. The TEP felt that the evidence provided by these studies was substantial enough to proceed with a measure.

6.3.1.3 Scientific acceptability

A first draft of the scientific acceptability text was drafted with language input from all TEP members.

Scientific Acceptability: *Prior concerns about assay variability are attenuated by required reporting of assay utilized such that trends in PTH can be monitored.*

There was a discussion of assay reporting for PTH given the earlier conversation of the variability that can affect the PTH measurement. There was some concern whether facilities should stick with the same assay or be allowed to change. PTH measures previously proposed by other measure developers were rejected because the scientific acceptability could not be determined. The EVOLVE and ADVANCE studies as well as several other studies were discussed briefly to ensure that scientific acceptability was more robust. One TEP member noted that most physicians should know the type of assay used based on the normative range for PTH given. Not all TEP members thought this would be the case. One TEP member was not sure if there was a linear relationship between higher or lower PTH values between different PTH assays. A TEP member hypothesized that NQF did not endorse the last a PTH measure proposed in 2010 by an organization other than CMS because of the concern over the assay variability which can be accounted for if the dialysis facilities report changes in the assay used and otherwise use

the same assay. The KDIGO guidelines suggested a range based on the normal value for the assay used. Furthermore, the EVOLVE and ADVANCE studies did not exist for the last TEP and therefore there is now more scientific acceptability.

6.3.1.4 Feasibility and usability

A first draft of the feasibility and usability text was drafted with language input from all TEP members.

Feasibility: *Routinely done in facilities.*

Usability: *Nephrologist routinely measure and interpret PTH levels. With measurement and assay reporting clinicians can interpret trends in PTH as per KDIGO recommendations.*

There was overall discussion of the language to use for these two points but there were also no objections by any TEP members regarding the content.

7. Dietary Counseling

7.1 Literature Review and Scientific Importance

One TEP member noted a paper (Shi 2013) that showed if the dietician spends more time with a patient, then the phosphorus levels tended to be more in a normal range but that the effect tapered off after 6 months. Another TEP member noted that dialysis facilities needed to have a dietician on staff but it wasn't clear if there was a requirement to see all patients. There was a discussion among all TEP members of the target patient population, which is to only counsel patients in need (high phosphorus levels versus all patients). A TEP member expressed concerns that this could turn into just E&M (Evaluation and Management) coding as an unintended consequence. There were also concerns that it is possible that a dietician may spend a lot of time with a patient and still may not be able to get the phosphorus levels down to an acceptable range. TEP members agreed at this point that the proposed measure would be a process measure, and therefore would not specify the duration of the interaction.

One TEP member thought that the measure would be more likely to be endorsed if it was specific to patients with high phosphorus. A TEP member responded by saying that there is not strong enough evidence to determine a threshold for high phosphorus; it was also mentioned that a 3 month rolling average should be considered if that was the path that was to be taken. Other TEP members agreed that the measure would become too complicated and that instead the measure should apply to every patient, that is, every patient would receive counseling. There were differing opinions about the frequency for providing counseling to patients. These varied from monthly to once a year. However there was consensus about educating patients to be able to recognize and avoid foods containing phosphorus as an additive as well as foods that are lower in phosphorus for the purpose of getting phosphorus lowered and maintaining it through diet. One TEP member thought that counseling should be once every month, while another thought it could be every 3 months. A TEP member pointed out two studies provide strong evidence that counseling a patient can be effective in lowering phosphorus – one intervention showed a significant difference of 0.6 (Sullivan 2009) lower phosphorus checked 3 months later while another intervention yielded a threshold of 5.5 (Mayne 2012). It was thought that

these studies were strong enough to possibly get the measure NQF endorsed. One TEP member noted that behavioral intervention of dietary counseling is very time intensive and requires consistency. The TEP discussed a similar sodium measure, "Dietary Sodium Reduction", that was not NQF endorsed during the 2010 cycle because of the failure to meet importance criteria.

One TEP member expressed concern about making additional rules in a facility that distracts people from doing all of the other patient-care activities that they were already doing so it is important that this measure doesn't distract from things like addressing other issues like albumin levels.

A TEP member stated that there is no mandate to counsel patients but the standard of care is to counsel patients and reevaluate every 3 months or if the patient has changed modalities. Overall, the TEP decided that if everyone agreed on the importance of the measure, such measure should be submitted. One TEP member mentioned that the language of the measure needed to be specific in that the dietary counseling was done and documented instead of just coming up with a plan. There was a belief that many patients may not get the actual counseling but instead the provider would document there is a plan. As discussions moved toward drafting measure specifications one TEP member stressed that the concept was that there are all kinds of dieticians but renal dieticians are specifically trained to deal with renal appropriate diets including phosphate control. The dietician would be able to counsel the patient on foods that are naturally high in phosphorus as well as foods that have phosphate as an additive.

7.2 Review of existing measures

The TEP discussed a comparable dietary counseling measure, "Dietary Sodium Reduction", specific to sodium intake. This measure was not NQF endorsed because of the importance criteria in that there was lack of evidence that dietary advice has an effect on sodium and that it is not clear what the consequences of high sodium are.

7.3 Recommendations

7.3.1 Proposed measure

Measure description: Percentage of all hemodialysis and peritoneal dialysis patients included in the sample for analysis with dietary counseling of the patient and/or caregiver on appropriate phosphorus sources and content as part of an overall healthy nutrition plan at least once within each six-month period

Numerator: Number of hemodialysis and peritoneal dialysis patients included in the denominator with dietary counseling, as described above, conducted at least once within six-months

Denominator: All hemodialysis and peritoneal dialysis patients included in the sample for analysis

Exclusions: Transient patients (in unit < 30 days), patients exclusively on non-oral food sources (i.e., total parenteral or enteral nutrition)

Measure Criteria:

Importance

- Hyperphosphatemia and severe hypophosphatemia are associated with increased morbidity and mortality.
- Dietary counseling as an intervention has been shown to reduce phosphorus levels in hyperphosphatemic patients.
- Recent studies have shown that reducing foods high in phosphate additives may result in reduction of serum phosphorus.
- Expert counseling, typically from a renal dietician, will limit potential adverse effects of unmonitored diets.

Scientific acceptability

- Renal dieticians are trained to individualize plans of care for improving phosphorus control.

Feasibility

- Dieticians are readily available in dialysis units and are adequately trained to implement the proposed measure.

Usability

- The above dietary counseling intervention is readily documentable.

7.3.1.1 Discussion of numerator and denominator

The discussion of the frequency of counseling patients was based on the current standard of care in which a patient is counseled at least once a year. It was decided that monthly was too much burden on the facility and if the TEP felt that every 3 months was necessary then they would need to restrict to just patients with hyperphosphatemia which means a threshold for high phosphorus would need to be established. The TEP did not feel comfortable establishing a threshold for high phosphorous and therefore every 6 months seemed appropriate. The TEP also decided that it should be for all patients including pediatric patients in which an appropriate caregiver should also be included. The only exclusion would be transient patients in the facility <30 days. The TEP also discussed the issue of patients with low serum phosphorus and whether increasing dietary phosphorus intake may impact patient outcomes.

The TEP was informed that based on the CMS Conditions for Coverage, patients are required to have a plan of care done within the first 30 days in a facility and then at 3 months and every year after that. They were asked to determine whether the dietary intervention proposed measure was specific enough to modify current practices that are already required. One TEP member countered with the fact that the same argument holds with phosphorus in which it is measured once a month anyway but there is still a requirement to measure it. It was stressed that it was a matter of seeing evidence that it is happening. Another TEP member expressed that this proposed measure is different because they are proposing every 6 months, as opposed to other dietary requirements outlined possibly in NKF in which the minimum may be every 12 months after the first 3 months of dialysis

One TEP member brought up a point about possibly excluding patients who are hearing or visually impaired as these would be barriers to receive and use verbal or written forms of counseling. Others agreed that these patients would still need dietary counseling but that it would have to be through another caregiver of the patient or through a translator. A TEP member brought up the point that patients with total parenteral nutrition or enteral nutrition could be excluded because these patients are already under extensive dietary intervention and the physician would change the prescription for the nutrition if there was an issue. The rest of the TEP agreed with this exclusion.

7.3.1.2 Importance

A first draft of the importance text was drafted by one TEP member. This text was then reviewed, edited, and discussed by the group.

Importance: *Hyperphosphatemia and severe low phosphorus levels are associated with increased morbidity and mortality. Dietary counseling has been shown to lower elevated levels.*

Discussions regarding importance are included in the summary in Section 7.1 Literature Review and Scientific Importance.

7.3.1.3 Scientific acceptability

A first draft of the scientific acceptability text was drafted by one TEP member. This text was then reviewed, edited, and discussed by the group.

Scientific Acceptability: *Dieticians offer the ability to individualize plan of care for improving MBD management*

Discussions regarding scientific acceptability are included in the summary in Section 7.1 Literature Review and Scientific Importance.

7.3.1.4 Feasibility and usability

A first draft of the feasibility and usability text was drafted by one TEP member. This text was then reviewed, edited, and discussed by the group.

Feasibility: *Dieticians are readily available in dialysis units and are adequately trained to deliver this intervention.*

Usability: *Results will limit adverse effects of patient's self-implemented diets (ex: low protein)*

Discussions regarding feasibility and usability are included in the summary in Section 7.1 Literature Review and Scientific Importance.

8. Vitamin D

8.1 Discussion

One TEP member brought up that the TEP had not discussed the subject of Vitamin D at this point. Another TEP member responded by saying the subject of Vitamin D in the general population is controversial and it would be more controversial in dialysis patients, given that the role of vitamin D in mineral metabolism in patients with ESRD may be even more complex . One TEP member added that the levels are all low and that they don't know the appropriate levels nor do we know the morbidity or mortality associated with levels. It was quickly agreed that there is not strong evidence to support a measure for Vitamin D at this time in ESRD patients.

8.2 Recommendation

The majority of the TEP agreed that there is not enough evidence to develop a Vitamin D measure.

9. Recommendations for Studies to improve the level of evidence in MBD, Funding, Policy, and CROWNWeb

9.1 Discussion

Since the TEP agreed that it is not possible to propose additional measures due to lack of evidence, CMS requested the TEP to come up with a list of studies that would provide the evidence needed. The TEP repeatedly raised the issue of the overall lack of evidence that was available due to the lack of randomized clinical trials that exist in order to inform recommendations for proposed measures, and meet the criterion of scientific acceptability. One TEP member said that if there had been a large randomized clinical trial that compared placebo and binders then the TEP would probably have felt that they would have sufficient evidence for a measure specifying either a phosphorus level or a treatment regimen; however, she acknowledged that it would not be funded due to concerns over the ethical concerns of a true placebo in an ESRD patients. However, a study that compared the achievement of two different levels of phosphorus would be ethical (for example randomized to phosphorus of 7 versus 4 mg/dl). However, a phosphate binder versus placebo study in CKD patients with hard end points was needed. The TEP discussed a pilot 9 month study in pre-dialysis CKD patients (Block2012) that showed no benefit on coronary calcification or bone density. . The TEP also determined that it would be useful to have studies that had hard outcomes such as cardiovascular events, fractures, mortality, hospitalizations etc. Randomized controlled trials that target different goals of phosphorus or PTH levels with hard outcomes would be the most ideal. Due to the lack of funding for studies in the field of MBD within the dialysis population, the TEP came up with a list of recommendations to CMS regarding ways to fund these studies. A few TEP members thought that CMS can direct and wholly or partially fund LDOs to try to conduct some of these studies for the sake of information.

Based on information provided by representatives from the CROWNWeb project team, the TEP members had a short discussion about the proposed measures. Some TEP members thought that the data entry for each patient would be too much of a burden to require lab values for all three of the

values (calcium, phosphorus, and PTH). There was some discussion that it might be easier for the facility to just input the percentage of patients instead of values for each patient. It was explained to the TEP that from a data standpoint it is much more accurate to input the patient's individual laboratory values.

One TEP member recommended that there is a difference in what we need from CROWNWeb and what we want from CROWNWeb. This TEP member said that the TEP is really only asking for a yes/no that phosphorus, calcium and PTH were done in the specified time frames for each measure whereas this could serve as rich resource for determining future measures. Another TEP member questioned why they would ask for the name of the assay if the TEP only wanted CW to report a yes/no for each of these measures instead of a value. A TEP member thought that it was important to know the assay so that it remained consistent. Other TEP members believed that it was easier to submit a value than to check a Yes/No box in CW and that most LDOs may have this process automated. The TEP discussed the burden of the data collection on the facility.

Due to the lack of definitive consensus on the required data elements for CROWNWeb, the TEP ultimately decided that it was best to leave the proposed measures as currently defined but suggested that the TEP have the opportunity to look at the data elements as they appear in CW and provide feedback prior to a change.

9.2 Recommendations for Studies that would Improve the Level of Evidence

The TEP's final recommendations for additional studies that they would like to see:

- Randomized controlled clinical trials in patients undergoing dialysis that target different goals of phosphorus or PTH levels with hard outcomes (mortality, fractures, hospitalizations, cardiovascular events) in ESRD patients
- Non-dialysis CKD patients - phosphorus binders vs. placebo for hard outcomes.
- Develop better dietary instructions appropriate for various levels of health literacy (i.e., reduced math and reading literacy)
- Incidence and prevalence of fractures, falls and frailty in dialysis patients needs further study
- Bone density, bone biomarkers, bone biopsies and other novel measures of bone health in dialysis patients

9.3 Recommendations for Funding

The TEP recognized that the key obstacle for any of the proposed studies is the lack of dedicated funding. Therefore, they specified the following funding recommendations for CMS consideration:

- CMS could partially or wholly fund studies that may improve care and result in health care cost savings in the field of MBD
- CMS could encourage the NIH/AHRQ to fund research in the field of MBD
- CMS could encourage LDOs to conduct national pragmatic trials in the field of MBD

9.4 Recommendations for Policy

The TEP also came up with a policy recommendation slide for limitations that arose within several discussions earlier and would be ideal in order to make future recommendations

- We encourage CMS to collaborate with the FDA as a major limitation to dietary phosphate control is a lack of quantitative food labeling for phosphate
- Furthermore phosphate additive are on the GRAS (generally regarded as safe) list and should be reevaluated.

9.5 Recommendations for CROWNWeb

CROWNWeb Related Request:

- The TEP should be consulted during the development of the implementation plan for collection of data by CMS for adopted quality measures

10. Conclusion and Summary Recommendations

The TEP addressed topic areas of interest to CMS, namely serum vs. plasma measurements for phosphorous and calcium, corrected vs. uncorrected serum calcium measurements, and bone histology/bone fractures. The TEP also recommended two new measures during their discussions at the in-person meeting.

- Measurement of Plasma PTH Concentration
- Percentage of Patients with Dietary Counseling

The TEP also suggested revisions to the existing measures:

- Measurement of serum phosphorus
- Measurement of uncorrected serum calcium (to be revised and then resubmitted to NQF, as this measure is currently retired; NQF #0216)

The TEP discussed the current hypercalcemia measure and recommended to not make any changes.

10.1 Recommendations for a future TEP

The TEP members recommended that a lab-specific TEP be convened that included experts in CLIA and CAP for the purpose of defining measures for laboratory tests and procedures that may affect laboratory values.

Due to the low level of evidence in the published literature on mineral and bone disorder, the TEP provided several recommendations for areas where further study and funding are needed as well as policy recommendations that would to aid in making future measure recommendations:

10.2 Recommendations for studies

- Randomized controlled clinical trials ESRD patients undergoing dialysis examining various levels of phosphorus or PTH levels and outcomes such as mortality, fractures, hospitalizations, and cardiovascular events
- Examining hard outcomes in Non-dialysis CKD patients on phosphorus binders vs. placebo
- Health literacy levels and dietary instructions
- Further study into the incidence and prevalence of fractures, falls
- Novel measures of bone health in dialysis patients

10.3 Recommendations for funding

- CMS could partially or wholly fund studies that may improve care and result in health care cost savings in the field of MBD
- CMS could encourage the NIH/AHRQ to fund research in the field of MBD
- CMS could encourage LDOs to conduct national pragmatic trials in the field of MBD

10.4 Recommendations for policy

- Collaboration between CMS and the FDA on quantitative food labeling
- Reevaluation of phosphate additives on the GRAS (generally regarded as safe) list

11. Summary of Measure Recommendations

Measure Name	**Measure Description**	**Type of Measure**
Measurement of Uncorrected Serum Calcium	Percentage of all peritoneal dialysis and hemodialysis patients included in the sample for analysis with uncorrected serum calcium measured at least once within a month	Process
Measurement of Serum Phosphorus	Percentage of all peritoneal dialysis and hemodialysis patients included in the sample for analysis with serum phosphorus measured at least once within a month	Process
Measurement of Plasma PTH Concentration	Percentage of all peritoneal dialysis and hemodialysis patients included in the sample for analysis with plasma PTH measured, together with documentation of the specific PTH assay utilized*, at least once within a 3 month period	Process

Percentage of Patients with Dietary Counseling	Percentage of all hemodialysis and peritoneal dialysis patients included in the sample for analysis with dietary counseling of the patient and/or caregiver on appropriate phosphorus sources and content as part of an overall healthy nutrition plan at least once within each six-month periods	Process

12. References

Block GA, Wheeler DC, Persky MS, et al. "Effects of phosphate binders in moderate CKD." *Journal of the American Society of Nephrology: JASN* (2012) 23(8):1407-15. PMID: 2282207Carothers J, Kurtz N, Lemann J. "Error introduced by specimen handling before determination of inorganic phosphate concentrations in plasma and serum." *Clinical Chemistry* (1976) 11:1909-12. PMID 975554

EVOLVE Trial Investigators, Chertow GM, Block GA, et al. "Effect of cinacalcet on cardiovascular disease in patients undergoing dialysis." *The New England journal of medicine* (2012) 367:2482-94. PMID: 23121374

Greenberg S, Gadde S, Pagala M, et al. "Optimal frequency of parathyroid hormone monitoring in chronic hemodialysis patients." *Clinical nephrology* (2011) 76:348-53. PMID: 22000553

Mayne TJ, Benner D, Ricketts K, et al. " Results of a Pilot Program to Improve Phosphorus Outcomes in Hemodialysis Patients." *Journal of Renal Nutrition : official journal of the Council on Renal Nutrition of the National Kidney Foundation* (2012) 22(5):472-9. PMID: 22056148

Raggi P, Chertow GM, Torres PU, et al. *"*The ADVANCE study: a randomized study to evaluate the effects of cinacalcet plus low-dose vitamin D on vascular calcification in patients on hemodialysis*." Nephrology, dialysis, transplantation: official publication of the European Dialysis and Transplant Association - European Renal Association* (2011) 26; 1327-39. PMID 21148030

Shi YX, Fan XY, Han HJ, et al. *"*Effectiveness of a nurse-led intensive educational programme on chronic kidney failure patients with hyperphosphataemia: randomised controlled trial.*" Journal of Clinical Nursing* (2013) 22: 7789-97.PMID 23480506

Sullivan C, Sayre SS, Leon JB et al. "Effect of Food Additives on Hyperphosphatemia Among Patients with End-stage Renal Disease". *JAMA: The Journal of the American Medical Association* (2009) 301; 629-35. PMID: 19211470

ESRD Quality Measure Development and Maintenance

Clinical Technical Expert Panel Members

The C-TEPs for Preventive Care, Pediatric Peritoneal Dialysis Adequacy, Mineral & Bone Disorder, and Hemodialysis Adequacy will meet face-to-face in Baltimore, Maryland on April 16-17, 2013. In addition, the TEP members will review materials provided in advance of the TEP meetings, participate in 2 to 3 pre-meeting conference calls, be available for follow-up communications to clarify their expert input and additional feedback for the final TEP summary report, and be available if needed to discuss any NQF recommendations.

The following individuals have been selected to participate in these **C-TEPs**:

	Name	Title	Organization	Measure Area
1.	Deepa Chand, MD MHSA	Medical Director, Kidney Dialysis	Akron Children's Hospital Medical Center	Pediatric Peritoneal Dialysis Adequacy
2.	Annabelle Chua, MD	Assistant Professor, Pediatrics	Baylor College of Medicine/Texas Children's Hospital	Pediatric Peritoneal Dialysis Adequacy
3.	Barbara Fivush, MD MHSA	Chief of Pediatric Nephrology	Johns Hopkins Children's Center	Pediatric Peritoneal Dialysis Adequacy
4.	Joseph Flynn, MD MS	Professor of Pediatrics	University of Washington School of Medicine	Pediatric Peritoneal Dialysis Adequacy
5.	Patti Spina, RN BSN CCRN	Pediatric Dialysis Unit Nurse Manager	Levine Children's Hospital	Pediatric Peritoneal Dialysis Adequacy
6.	Bradley Warady, MD	Chief, Pediatric Nephrology; Director, Dialysis and Transplantation; Professor of Pediatrics	University of Missouri, Kansas City School of Medicine	Pediatric Peritoneal Dialysis Adequacy
7.	Suhail Ahmad, MD	Senior Medical Director	Northwest Kidney Centers	Dialysis Adequacy
8.	William Dant	Chair of Dialysis Quality Initiative	Renal Support Network	Dialysis Adequacy
9.	John Daugirdas, MD	Professor of Medicine	University of Illinois at Chicago	Dialysis Adequacy
10.	Thomas Depner, MD	Professor of Medicine	University of California	Dialysis Adequacy
11.	Peter DeOreo, MD	Chief Medical Officer	Centers for Dialysis Care	Dialysis Adequacy
12.	Elizabeth Evans, DNP CWCN RN	Nurse Practitioner	Renal Medicine Associates	Dialysis Adequacy

	Name	Title	Organization	Measure Area
13.	Stuart Goldstein, MD	Professor of Pediatrics	University of Cincinnati College of Medicine	Dialysis Adequacy
14.	Eduardo Lacson, MD, MPH	Vice President, Clinical Science, Epidemiology and Research	Fresenius Medical Care NA	Dialysis Adequacy
15.	Michael Rocco, MD MSCE	Professor of Internal Medicine/Nephrology	Wake Forest University School of Medicine	Dialysis Adequacy
16.	Constance Anderson, BSN MBA	Vice President of Clinical Operations	Northwest Kidney Centers	Preventive Care
17.	Kevin Chan, MD, MSC	Senior Director of Clinical Outcomes Research and Medical Analytics	Fresenius Medical Care North America	Preventive Care
18.	Alfred Cheung, MD	Professor of Medicine, Division of Nephrology	University of Utah	Preventive Care
19.	David Gilbertson	Executive Director of Epidemiology and Biostatistics	United States Renal Data System (USRDS)	Preventive Care
20.	Raymond Hakim, MD, PhD	Attending Physician/Nephrologist	Vanderbilt University	Preventive Care
21.	Celeste Castillo Lee	Senior Project Manager, Office of the Provost	University of Michigan	Preventive Care
22.	Paul Martin, MD, FRCP, FRCPI	Chief, Division of Hepatology	University of Miami Miller School of Medicine	Preventive Care
23.	Alicia Neu, MD	Professor, Pediatric Nephrology	John Hopkins Medicine	Preventive Care
24.	David Van Wyck, MD	Vice President, Clinical Services	DaVita	Preventive Care
25.	Patty Danielson, RN BC	Nurse Manager, Behavioral Health	Adventist Medical Center	Mineral & Bone Disorder
26.	Kathy Schiro Harvey, MS RD CSR	Nutrition Manager	Puget Sound Kidney Centers	Mineral & Bone Disorder
27.	Tamara Isakova, MD	Assistant Professor of Medicine	University of Miami Miller School of Medicine	Mineral & Bone Disorder
28.	Mary Leonard, MD MSCE	Professor of Pediatrics and Epidemiology	The Children's Hospital of Philadelphia Research Institute	Mineral & Bone Disorder
29.	Julia Lewis, MD	Professor of Medicine	Vanderbilt University Medical Center	Mineral & Bone Disorder
30.	Hartmut Malluche, MD FACP	Professor and Chief, Division of Nephrology, Bone and Mineral Metabolism Department of Internal Medicine	University of Kentucky Medical Center	Mineral & Bone Disorder

	Name	Title	Organization	Measure Area
31.	Robin Mauer, FNP MSN	Nurse Practitioner	Washington University School of Medicine	Mineral & Bone Disorder
32.	Klemens Meyer, MD	Professor of Medicine	Tufts University Medical Center	Mineral & Bone Disorder
33.	Sharon Moe, MD FASN	Director, Division of Nephrology	Indiana University School of Medicine	Mineral & Bone Disorder