Biomarkers for Assessing and Managing Iron Deficiency Anemia in Late-Stage Chronic Kidney Disease
Biomarkers for Assessing and Managing Iron Deficiency Anemia in Late-Stage Chronic Kidney Disease

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Contract No. 290-2007-10055-I

Prepared by:
Tufts Evidence-based Practice Center, Tufts Medical Center
Boston, MA

Investigators:
Mei Chung, Ph.D., M.P.H.
Denish Moorthy, M.B.B.S., M.S.
Nira Hadar, M.S.
 Priyanka Salvi, M.S.
Ramon C. Iovin, Ph.D.
Joseph Lau, M.D.
Preface

The Agency for Healthcare Research and Quality (AHRQ) conducts the Effective Health Care Program as part of its mission to organize knowledge and make it available to inform decisions about health care. As part of the Medicare Prescription Drug, Improvement, and Modernization Act of 2003, Congress directed AHRQ to conduct and support research on the comparative outcomes, clinical effectiveness, and appropriateness of pharmaceuticals, devices, and health care services to meet the needs of Medicare, Medicaid, and the Children’s Health Insurance Program (CHIP).

AHRQ has an established network of Evidence-based Practice Centers (EPCs) that produce Evidence Reports/Technology Assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care. The EPCs now lend their expertise to the Effective Health Care Program by conducting comparative effectiveness reviews (CERs) of medications, devices, and other relevant interventions, including strategies for how these items and services can best be organized, managed, and delivered.

Systematic reviews are the building blocks underlying evidence-based practice; they focus attention on the strength and limits of evidence from research studies about the effectiveness and safety of a clinical intervention. In the context of developing recommendations for practice, systematic reviews are useful because they define the strengths and limits of the evidence, clarifying whether assertions about the value of the intervention are based on strong evidence from clinical studies. For more information about systematic reviews, see www.effectivehealthcare.ahrq.gov/reference/purpose.cfm.

AHRQ expects that CERs will be helpful to health plans, providers, purchasers, government programs, and the health care system as a whole. In addition, AHRQ is committed to presenting information in different formats so that consumers who make decisions about their own and their family’s health can benefit from the evidence.

Transparency and stakeholder input are essential to the Effective Health Care Program. Please visit the Web site (www.effectivehealthcare.ahrq.gov) to see draft research questions and reports or to join an email list to learn about new program products and opportunities for input. Comparative Effectiveness Reviews will be updated regularly.

We welcome comments on this CER. They may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850, or by email to epc@ahrq.hhs.gov.

Carolyn M. Clancy, M.D.
Director, Agency for Healthcare Research and Quality

Stephanie Chang, M.D., M.P.H.
Director, EPC Program
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Jean Slutsky, P.A., M.S.P.H.
Director, Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Elisabeth U. Kato, M.D., M.R.P.
Task Order Officer
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality
Acknowledgments
The authors would like to thank the clinical content expertise provided by Madhumathi Rao, M.D., Ph.D.

Key Informants

James Adamson, M.D., FACP, FCCP
Senior Vice President and Medical Officer
for National Accounts
Arkansas Blue Cross Blue Shield (BCBS)
Little Rock, AR

James C. Boyd, M.D.
Associate Professor of Pathology
University of Virginia School of Medicine
Director of Clinical Pathology
Charlottesville, VA

Haewook Han, P. h.D., R.D., CSR, LDN
Renal Nutrition Specialist
Department of Nephrology,
Harvard Vanguard Medical Associate
Boston, MA

Csaba P Kovesdy, M.D.
Associate Professor of Clinical Internal Medicine
University of Virginia School of Medicine
Chief of Nephrology
Salem Veterans Affairs Medical Center
Charlottesville, VA

Kathe LeBeau
Patient Advocacy Project Manager
Renal Support Network
Albany, NY

Maureen M. Okam, M.D., M.P.H.
Associate Director, Brigham and Women’s Hospital Ambulatory Infusion Center
Brigham and Women's Hospital
Boston, MA

James Rollins, M.D., M.S.H.A., Ph.D.
Director of Division of Items and Devices Coverage & Analysis Group
Centers for Medicare and Medicaid (CMS)
Baltimore, MD

Deborah Van Allen R.N., M.S.
Nurse Manager
Dialysis Clinic, Inc. (DCI Boston)
Boston, MA

Jay B. Wish, M.D.
Professor of Medicine
Case Western Reserve University School of Medicine
Medical Director, Dialysis Program
University Hospitals Case Medical Center
Cleveland, OH
Technical Expert Panel

Ali K. Abu-Alfa M.D., FASN
Professor of Medicine
Head, Division of Nephrology and Hypertension
American University of Beirut
Beirut-Lebanon
and
Adjunct Faculty
Section of Nephrology
Yale School of Medicine
New Haven, CT

Meredith Atkinson, M.D., M.H.S.
Assistant Professor of Pediatrics
Division of Pediatric Nephrology
John Hopkins University School of Medicine
Baltimore, MD

Jeffrey S. Berns, M.D.
Associate Dean for Graduate Medical Education
Professor of Medicine and Pediatrics
Renal-Electrolyte and Hypertension Division
Hospital of the University of Pennsylvania
Philadelphia, PA

Jay B. Wish, M.D.
Professor of Medicine
Case Western Reserve University School of Medicine
Medical Director, Dialysis Program
University Hospitals Case Medical Center
Cleveland, OH
Peer Reviewers

John W. Adamson, M.D.
Clinical Professor of Medicine
Hematology/Oncology
University of California, San Diego

Jeffrey S. Berns, M.D.
Associate Dean for Graduate Medical Education
Professor of Medicine and Pediatrics
Renal-Electrolyte and Hypertension Division
Hospital of the University of Pennsylvania
Philadelphia, PA

Alex C. Chin, Ph.D., DABCC, FACB
Clinical Chemist
Calgary Laboratory Services
Clinical Assistant Professor
Department of Pathology and Laboratory Medicine Faculty of Medicine University of Calgary
Calgary, AB Canada

Susan Furth, M.D., Ph.D.
Division Chief, Nephrology
Professor, Department of Pediatrics and Epidemiology, Perelman School of Medicine at the University of Pennsylvania
Senior Scholar, Clinical Epidemiology Unit, Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine at the University of Pennsylvania

Ameet R. Kini, M.D., Ph.D.
Associate Professor of Pathology
Director, Division of Hematopathology
Director, Hematology and Flow Cytometry Laboratories
Associate Director, Molecular Diagnostics
Loyola University Chicago Stritch School of Medicine
Maywood, IL

Mitchell Scott, Ph.D.
Professor of Pathology and Immunology
Co-Medical Director, Clinical Chemistry Division of Laboratory and Genomic Medicine
St. Louis, MO

Jay B. Wish, M.D.
Professor of Medicine
Case Western Reserve University School of Medicine
Medical Director, Dialysis Program
University Hospitals Case Medical Center
Cleveland, OH
Biomarkers for Assessing and Managing Iron Deficiency Anemia in Late-Stage Chronic Kidney Disease

Structured Abstract

Background. Anemia is a common complication of chronic kidney disease (CKD) that develops early in the course of CKD, and becomes increasingly severe as the disease progresses. The management of anemia in CKD patients requires an appropriate balance between stimulating the generation of erythroblasts (erythropoiesis) and maintaining sufficient iron levels for optimum hemoglobin (Hb) production. Thus, assessing iron status is integral to both iron and anemia management in CKD patients, as iron is essential for Hb formation (as is erythropoietin). However, classical laboratory biomarkers of iron deficiency exhibit a wide biological variability in CKD. In response, newer, less-variable markers have been proposed.

Purpose. To summarize the literature on the use of newer versus classical laboratory biomarkers of iron status as part of the management strategies for iron deficiency in stages 3–5 CKD patients (nondialysis and dialysis).

Data sources. All published articles identified through MEDLINE®, and the Cochrane Central Register of Controlled Trials, from inception to May 2012.

Study selection. Two reviewers independently selected studies on the basis of predetermined eligibility criteria. We considered studies of pediatric and adult nondialysis patients with stage 3, 4, or 5 CKD, patients with CKD undergoing dialysis (hemo- or peritoneal dialysis), and patients with a kidney transplant. Studies that compared newer laboratory biomarkers of interest such as hemoglobin content in reticulocytes (CHr), percentage of hypochromic red blood cells (%HYPO), erythrocyte zinc protoporphyrin (ZPP), soluble transferrin receptor (sTfR), hepcidin, and superconducting quantum interference devices (SQUID), with classical laboratory biomarkers, such as bone marrow iron stores, serum iron, transferrin saturation (TSAT), iron-binding capacity, and serum ferritin were included.

Data extraction. One reviewer abstracted article information into predesigned extraction forms; a second reviewer checked information for accuracy. A standardized protocol was used to extract details on designs, diagnoses, interventions, outcomes, and methodological issues.

Data synthesis. A total of 30 articles were accepted, including one Polish- and one Japanese-language publication. We did not identify any study that provided data directly addressing our overarching question (Key Question 1) regarding the impact of using newer laboratory biomarkers on patient-centered outcomes (mortality, morbidity, quality of life, and adverse effects). We identified 27 studies to answer Key Question 2, which addresses the performance of newer markers of iron status as a replacement for or in addition to classical markers.

The synthesis of data for Key Question 2 was complicated by the lack of generally accepted reference standard tests for determining iron deficiency in the context of CKD. Of the 27
included studies, 15 used classical markers of iron status to define “iron deficiency” as the reference standard in calculating the test accuracy (sensitivity and specificity) of newer markers of iron status. For the purpose of our review, this approach was analogous to assessing the concordance between classical and newer biomarkers of iron status; thus, these studies were only included for subquestion 2a (What reference standards are used for the diagnosis of iron deficiency in studies evaluating test performance?). The remaining 12 studies investigated the test performance of newer or classical markers of iron status, using a response to IV iron treatment as the reference standard for the diagnosis of iron deficiency. We therefore synthesized these 12 studies for Key Question 2. Of these 12 studies, most studies enrolled only adult CKD patients on hemodialysis (HD CKD patients), though a few examined adult peritoneal dialysis (PD) and nondialysis (ND) CKD patients. Only one study enrolled pediatric CKD patients. Although the reviewed studies evaluated many newer markers, such as CRh, %HYPO, RetHe, sTfR, hepcidin, and ZPP, the majority assessed CRh or %HYPO among adult HD CKD patients.

Based on our analysis, we concluded that there is a low level of evidence that both CRh and %HYPO have a similar or better overall test accuracy compared with classical markers (TSAT or ferritin) to predict a response to IV iron treatment (as the reference standard for iron deficiency). In addition, data from a few studies suggest that CRh (with cutoff values of <27 or <28 pg) and %HYPO (with cutoff values of >6% or >10%) have better sensitivities and specificities to predict iron deficiency than classical markers (TSAT <20 or ferritin <100 ng/mL). There is also a low level of evidence that sTfR has a similar test performance compared with classical markers (TSAT or ferritin) to predict a response to IV iron treatment, but the strength of evidence was insufficient to come to a conclusion regarding the test performance of newer markers of iron status as an add-on to older markers, and that of ZPP and hepcidin. It should be noted that, across studies, there exists a high degree of heterogeneity in the test comparisons, definitions for the reference standard (a response to IV iron treatment), iron status of the study populations (assessed by TSAT or ferritin), and background treatment. This heterogeneity may limit the comparability of findings across studies.

For Key Question 3 (impact on intermediate outcomes of newer markers compared with older markers), we identified only two short-term RCTs (4 and 6 months), enrolling a total of 354 adult HD CKD patients. We concluded that there is a low level of evidence for a reduction in the number of iron status tests and resulting intravenous iron treatments (a post hoc intermediate outcome) administered to patients whose iron management was guided by CRh compared with those guided by TSAT or ferritin. Both RCTs reported that Hct remained in the targeted ranges (an indication for the adequacy of anemia management) throughout the study period in all randomized arms, though the Hct target was higher in the U.S. trial than the Japanese trial.

For Key Question 4 (factors affecting the test performance and clinical utility of newer markers), we included 3 studies (1 RCT and 2 prospective cohorts) as well as relevant data from all 27 studies included in Key Questions 2; however, we found insufficient evidence to draw any conclusions, as only single studies or indirect comparisons across studies provided relevant data.

Limitations. The available data are very limited due to a high degree of heterogeneity. There exist many definitions of a response to IV iron treatment as the reference standard for iron deficiency. Moreover, there is a lack of a uniform regimen of intravenous iron treatment in terms of dosage, iron formulation, treatment frequency, and followup duration for the iron challenge
test (to define a response) across studies. Many studies included in our review were also rated as being at a high risk of bias, limiting their utility in informing clinical practice.

**Conclusions.** Combining the evidence addressing Key Questions 2, 3, and 4, we can conclude that all currently available laboratory biomarkers of iron status (either newer or classical markers) do not have an ideal predictive ability when used singly to determine iron deficiency as defined by a response to iron challenge test. Furthermore, we can conclude that there is insufficient evidence to determine the test performance of the combinations of newer biomarkers, or combinations of newer and classical biomarkers, for diagnosing iron deficiency. However, it may be that CHr and %HYPO have better predictive ability for a response to IV iron treatment than classical markers (TSAT <20 or ferritin <100 ng/mL) in HD CKD patients. In addition, results from two RCTs showed a reduction in the number of iron status tests and resulting IV iron treatments administered to patients whose iron management was guided by CHr, compared with those guided by TSAT or ferritin. These results suggest that CHr may reduce potential harms from IV iron treatment by lowering the frequency of iron testing, although the evidence for the potential harms associated with testing or test-associated treatment is insufficient.

Nevertheless, the strength of evidence supporting these conclusions is low, and there remains considerable clinical uncertainty regarding the use of newer markers in the assessment of iron status and management of iron deficiency in stages 3–5 CKD patients (both nondialysis and dialysis). In addition, factors that may affect the test performance and clinical utility of newer laboratory markers of iron status remain largely unexamined.
# Contents

**Executive Summary** .................................................................................................................. ES-1

**Introduction** ................................................................................................................................ 1
  - Background ............................................................................................................................... 1
    - Chronic Kidney Disease and Iron Management ................................................................. 1
    - Laboratory Biomarkers of Iron Status ................................................................................ 2
  - Scope and Key Questions ......................................................................................................... 5
    - Scope of the Review ........................................................................................................... 5
    - Key Questions and Analytic Framework ............................................................................ 5
  - Organization of This Report ..................................................................................................... 9

**Methods** ........................................................................................................................................ 10
  - AHRQ Task Order Officer ...................................................................................................... 10
  - Topic Refinement and Review Protocol ................................................................................ 10
  - Literature Search Strategy ....................................................................................................... 10
    - Search Strategy ................................................................................................................. 10
    - Inclusion and Exclusion Criteria ....................................................................................... 11
    - Study Selection ................................................................................................................... 12
    - Data Extraction ................................................................................................................. 12
  - Risk of Bias—Assessment of Individual Studies ..................................................................... 13
  - Data Synthesis ......................................................................................................................... 13
    - Summary Tables ............................................................................................................... 14
    - Graphical Presentation of Study Results ........................................................................... 14
    - Test Performance Terms and Definitions ......................................................................... 14
  - Strength of the Body of Evidence ........................................................................................... 15
  - Applicability ........................................................................................................................... 16
  - Peer Review and Public Commentary .................................................................................... 16

**Results** .......................................................................................................................................... 17
  - Introduction .............................................................................................................................. 17
  - Literature Searches .................................................................................................................. 17
  - Description of Included Studies .............................................................................................. 17
  - Key Question 1. Comparative Effectiveness of Newer Versus Older Markers of Iron Status for the Diagnosis and Management of Iron Deficiency Anemia ................................................. 18
  - Key Question 2. Test Performance of Newer Markers Compared With the Older Markers of Iron Status .................................................................................................................... 19
    - 2b. Adverse Effects or Harms Associated With Testing .................................................. 49
  - Key Question 3. Intermediate Outcomes Comparing Iron Management Guided by Newer Laboratory Markers With Those of Iron Management Guided by Older Laboratory Markers .............................................................................................................................. 49
  - Key Points (Table 15) ....................................................................................................... 49
  - Detailed Synthesis (Tables 19 and 20) .............................................................................. 56
  - Key Question 4: Factors Affecting Test Performance and Clinical Utility ............................ 56
  - Key Points ......................................................................................................................... 56
  - Detailed Synthesis (Tables 19 and 20) .............................................................................. 56
Table 14. Adverse effects or harms reported in the 27 studies included in
  Key Question 2 ....................................................................................................................... 49
Table 15. Overall strength of evidence for intermediate outcomes comparing
  iron management guided by newer laboratory markers with those of iron
  management guided by older laboratory markers ........................................................................ 50
Table 16. Characteristics of randomized controlled trials comparing intermediate
  outcomes of iron management guided by classical laboratory markers with
  those of iron management guided by newer laboratory markers of iron status
  in CKD patients undergoing hemodialysis ............................................................................. 53
Table 17. Dose of epoetin required to maintain hematocrit targets ........................................................................ 54
Table 18. Changes in iron status markers ..................................................................................... 55
Table 19. Characteristics of studies evaluating factors affecting test performance
  and clinical utility ................................................................................................................... 58
Table 20. Test accuracy of TSAT, ferritin, CHr, and sTfR for predicting change
  in hemoglobin in subgroups of IV iron and no IV iron treatment .............................................. 59
Table 21. Summary of the strength of evidence addressing Key Questions ................................ 60
Table 22. Research gaps and suggestions for future research ...................................................... 66

Figures
Figure A. Analytic framework .................................................................................................. ES-5
Figure 1. Roles of current and newly proposed markers of iron status ........................................ 4
Figure 2. Analytic framework ......................................................................................................... 7
Figure 3. Literature flow ............................................................................................................... 18
Figure 4. Indirect comparisons of the overall test accuracy of newer versus
  classical markers of iron status (at baseline) to predict response to IV iron
  among adult HD CKD patients—CHr, %HYPO, sTfR .......................................................... 29
Figure 5. Indirect comparisons of the overall test accuracy of newer versus
  classical markers of iron status (at baseline) to predict response to IV iron
  among adult HD CKD patients—ZPP, Ferritin, TSAT .......................................................... 30

Appendixes
Appendix A. Literature Search Strategy
Appendix B. Excluded Studies
Appendix C. Blank Extraction Form
Appendix D. Test Performance of Newer Markers for Assessing Iron Status as Defined by
  Classical Laboratory Markers
Appendix E. Quality Criteria and Individual Study Grades
Executive Summary

Background

Chronic kidney disease (CKD) is the gradual, progressive deterioration of kidney function, and a condition that affects an estimated 26 million American adults. A common complication of CKD is anemia, which results from inadequate erythropoietin or from iron deficiency as a result of inadequate absorption or mobilization. The management of anemia in CKD patients must strike an appropriate balance between stimulating generation of erythroblasts (erythropoiesis) and maintaining sufficient iron levels for optimum hemoglobin (Hb) production.\(^1\) Erythropoietic stimulating agents (ESAs) mobilize iron stores in promoting erythropoiesis; however, decreased iron stores or iron availability are the most common reasons for resistance to the effect of ESAs. Thus, most patients who receive ESA treatment will require supplemental (oral or intravenous) iron to ensure an adequate response with erythropoietic agents. Iron management (iron status assessment and iron treatment), therefore, is an essential part of the treatment of anemia associated with CKD,\(^1\) as there remain outstanding concerns regarding the adverse effects associated with elevated doses of ESAs\(^2\) and supplemental iron\(^3\).

Assessing iron status is integral to both iron and anemia managements in CKD patients. Bone marrow iron stores are often regarded as the best indicator of iron status (although this is not universally accepted);\(^1\) however, taking a bone marrow sample is invasive and involves risks of infection or bleeding at the biopsy site.\(^4\) Other classical iron status tests, of which ferritin and transferrin saturation (TSAT) are the most widely used, reflect either the level of iron in tissue stores or the adequacy of iron for erythropoiesis. Serum ferritin reflects storage iron—iron that is stored in liver, spleen, and bone marrow reticuloendothelial cells. The TSAT percentage value reflects iron that is readily available for erythropoiesis. Guidelines on monitoring iron status stipulate that hemodialysis (HD) patients receiving erythropoietin should have their iron status monitored every 3 months, and maintain a transferrin saturation (TSAT) >20 percent and a serum ferritin level >100 ng/mL (>200 ng/mL for CKD patients on HD).\(^5,6\) The National Kidney Foundation guidelines have been widely adopted in dialysis centers across the United States.

Though widely used, classical laboratory biomarkers of iron status are not without drawbacks when used in CKD patients: CKD is a pro-inflammatory state, and the biological variability of serum iron, transferrin saturation, and ferritin is known to be large in the context of underlying inflammation.\(^7,9\) In an attempt to find alternative methods to assess iron status in the setting of CKD, several novel biomarkers of iron status have been proposed:

- The hemoglobin (Hb) content of reticulocytes (CHr)/Reticulocyte hemoglobin equivalent (RetHe): CHr and RetHe measurements are functionally equivalent,\(^10\) but the two measurements are performed by different analyzers. CHr/RetHe, which examines both the precursors and mature red cells, provides an opportunity to detect and monitor acute and chronic changes in cellular hemoglobin status. CHr/RetHe measurement is a function of the amount of iron in the bone marrow that is available for incorporation into reticulocytes (immature red blood cells);\(^11\) decreased levels of CHr/RetHe indicate iron deficiency.
- The percentage of hypochromic erythrocytes (%HYPO): %HYPO is a measurement of Hb in red blood cell (RBC), which factors in the absolute Hb content as well as the size of the RBC.\(^12\) This can be used to measure functional iron deficiency. If iron supply is
low in the face of ESA therapy, then there is lesser amount of Hb being incorporated into each RBC, and as a result, %HYPO levels are high.

- Erythrocyte zinc protoporphyrin (ZPP): ZPP is a measure of iron incorporation in heme. When iron levels are low, zinc is used instead of iron in the formation of heme, a protein component of Hb. As a result, ZPP levels increase, indicating iron deficiency.13

- Soluble transferrin receptor (sTfR): sTfR measures the availability of iron in the bone marrow. When the bone marrow is stimulated by erythropoiesis stimulating agents (ESAs), it results in increased expression of transferrin receptors on the surface of erythroblasts, the precursors of RBC. If iron supply is low, then levels of transferrin containing iron are low, and there is a mismatch between the numbers of transferrin receptors and the transferrin-iron complexes to bind with them. Some of the transferrin receptors that are not bound by iron-containing transferrin then get detached and can be detected in the blood. Increased concentration of sTfRs in the blood is an indicator of iron deficiency.

- Hepcidin: Hepcidin is a peptide produced by the liver that regulates both iron absorption in the intestine as well as release of iron from macrophages. Increased levels of hepcidin have indeed been associated with a decrease in available iron.14

- Superconducting QUantum Interference Device (SQUID) is a noninvasive method for the detection and quantification of liver iron content,15 because of the paramagnetic properties of iron, magnetic resonance signal diminishes in liver as iron concentration increases.

Although a number of international guidelines have examined the use of both classical and new serum iron biomarkers, their recommendations differ. Across guidelines, it is agreed that the optimal management of anemia in HD patients depends on diagnosis and management of iron deficiency. However, a number of questions remain without consensus, including: Which combination of iron biomarkers is required? Should the newer biomarkers be used as a replacement for or in addition to classical markers?

In view of the considerable clinical uncertainty, the high biological variability associated with laboratory biomarkers, and the need for frequent assessment of iron status to guide treatment for anemia, a systematic review of the relevant literature is a priority.

**Objectives**

The purpose of this review is to evaluate the impact on patient-centered outcomes of the use of newer versus classical laboratory biomarkers of iron status as part of the management strategy for anemia in patients with CKD stages 3–5, that is, nondialysis or dialysis patients with CKD or kidney-transplant patients.

**Key Questions and Analytic Framework (Figure A)**

As test results have little direct impact on patient-relevant outcomes, the utility of a medical test is usually determined by its indirect effect on outcomes, that is, through its influence on therapeutic decision making and subsequently on patient outcomes. Although studies that assess the overall impact of tests on the clinical management process would provide the most direct evidence for this CER, they are often challenging or infeasible to conduct. Because we expected to find little of such evidence, the question of overall impact (Key Question 1, see below for full descriptions of all Key Questions) was broken out into three component Key Questions (Key
Questions 2 to 4). Combining evidence gathered to address these three component Key Questions can thus inform the conclusions for this review’s primary, overarching question.

**Key Question 1**

What is the impact on patient-centered outcomes of using newer laboratory biomarkers\(^a\) as a replacement for or an add-on to the older laboratory biomarkers of iron status\(^b\) for the assessing of iron status and management of iron deficiency in stages 3–5 CKD patients (nondialysis and dialysis), and in patients with a kidney transplant?

**Key Question 2**

What is the test performance of newer markers of iron status\(^a\) as a replacement for or an add-on to the older markers\(^b\) in stages 3–5 nondialysis and dialysis patients with CKD, and in patients with a kidney transplant?

- a. What reference standards are used for the diagnosis of iron deficiency in studies evaluating test performance?
- b. What are the adverse effects or harms associated with testing using newer and/or older markers of iron status?

**Key Question 3**

In stages 3–5 nondialysis and dialysis CKD patients with iron deficiency, what is the impact of managing iron status based on newer laboratory biomarkers either alone or in addition to older laboratory biomarkers on intermediate outcomes (e.g., improvement in Hb levels, dose of erythropoiesis-stimulating agents, time in target Hb range), compared with managing iron status based on older laboratory biomarkers alone?

- a. What are the adverse effects or harms associated with the treatments guided by tests of iron status?

**Key Question 4**

What factors affect the test performance and clinical utility of newer markers of iron status, either alone or in addition to older laboratory biomarkers, in stages 3–5 (nondialysis and dialysis CKD patients with iron deficiency)? For example:

- Biological variation in diagnostic indices
- Use of different diagnostic reference standards
- Type of dialysis (i.e., peritoneal or hemodialysis)
- Patient subgroups (i.e., age, sex, comorbid conditions, erythropoiesis-stimulating agent resistance, protein energy malnutrition secondary to an inflammatory state, hemoglobinopathies [e.g., thalassemia and sickle cell anemia])
- Route of iron administration (i.e., oral or intravenous)
- Treatment regimen (i.e., repletion or continuous treatment)

\(^a\)Content of hemoglobin [Hb] in reticulocytes, percentage of hypochromic red blood cells, erythrocyte zinc protoporphyrin, soluble transferrin receptor, hepcidin, and superconducting quantum interference devices.

\(^b\)Bone marrow iron stores, serum iron, transferrin saturation, iron-binding capacity, and ferritin.
• Interactions between treatments (i.e., patients treated with or without erythropoiesis-stimulating agents, patients treated with or without iron-replacement therapy)
• Other factors (based on additional information in the reviewed papers).
Figure A. Analytic framework

CKD=chronic kidney disease; ESA=erythropoiesis-stimulating agents; Hb=hemoglobin level; KQ=Key Question
Methods

Data Sources and Selection

We conducted literature searches of studies in MEDLINE® (from inception to May 2012) and the Cochrane Central Register of Controlled Trials (through the first quarter of 2012). Studies published in any language with adult human subjects were screened to identify articles relevant to each Key Question. We also consulted a Technical Expert Panel, and screened the reference lists of related guidelines and selected narrative reviews and primary articles for additional articles. For all Key Questions, we excluded studies with fewer than 10 patients with CKD. The eligibility criteria for study populations for all Key Questions included pediatric and adult nondialysis patients with stage 3, 4, or 5 CKD; patients with CKD undergoing dialysis (HD or PD); and patients with a kidney transplant. For interventions, eligible studies were those involving the newer laboratory biomarkers (see the list in the Key Questions section above) to diagnose and manage iron deficiency either as a replacement for classical markers or in addition to classical biomarkers. For comparators, eligible studies were those involving classical laboratory biomarkers (see the list in the Key Questions section) to diagnose and manage iron deficiency.

Key Question 1 outcomes included mortality, morbidity, quality of life measured using standardized scales, or adverse effects or harms associated with testing and associated treatments. Key Questions 2 and 4 outcomes included measures of test performance comparing newer markers with classical markers of iron status. We accepted any “reference standard” used in the original studies for the analyses of sensitivity and specificity, including functional iron deficiency as defined by response or nonresponse to treatment. For Key Questions 3 and 4, the intermediate outcomes included increase in Hb or hematocrit, more consistent maintenance of Hb or hematocrit, use of ESAs for maintenance of Hb, or adverse effects or harms associated with different management strategies.

For Key Question 2, we included any study design. For Key Question 3, we included only randomized controlled trials (RCTs), as well as non-RCTs and observational studies with concurrent comparison groups. Studies could have any length of followup or any setting. Data were extracted into standard forms. We extracted bibliographic data, eligibility criteria, and enrollment years for all studies. We also extracted population characteristics such as basic demographic data—age, sex, and race or ethnic group—as well as sample size, study design, descriptions of the test and reference standard, analytic details, and outcomes.

Quality (Risk of Bias) Assessment of Individual Studies

We assessed the risk of bias (methodological quality) for each study using the Agency for Healthcare Research and Quality’s Methods Guide for Effectiveness and Comparative Effectiveness Review (from here on referred to as the Methods Guide). Briefly, we rated each study as being at a high, medium, or low risk of bias on the basis of adherence (Yes, No, or Unclear/Not reported) to well-accepted standard methodologies (Quality Assessment of Diagnostic Accuracy Studies [QUADAS] tool for studies of diagnostic performance, and the Cochrane risk of bias tool for intervention studies) and assessed and reported each methodological quality item for all qualifying studies. We also considered the clarity and consistency in reporting as part of the overall judgment of risk of bias. Grading was outcome-specific, such that a given study that reported its primary outcome well but conducted an
incomplete analysis of a secondary outcome would be graded as having a different quality rating for each of the two outcomes. Studies of different study designs were graded within the context of their study design; RCTs and observational studies were graded separately to be at a high, medium, or low risk of bias. Only RCTs and prospective cohort studies could be rated as being at a low risk of bias.

**Data Synthesis**

We summarized all included studies in narrative form as well as in summary tables that condense the important features of the study populations, design, anemia and iron status indices, laboratory tests, reference standards, background treatment, intervention, outcomes, and results. We used summary tables to succinctly report measures of the main outcomes evaluated and additional information to assist their interpretation.

The synthesis of data for Key Question 2 was complicated by the lack of generally accepted reference standard tests for determining iron deficiency in the context of CKD.\(^1\) Thus, we accepted any “reference standard” used by the authors of the included primary studies for the analyses of test performance of newer or classical laboratory biomarkers of iron status. Based on our post hoc observation of this body of literature, we separated studies into two distinct groups. Specifically, current studies used two distinct methods to operationalize a reference standard for assessing test performance: (1) a response to intravenous (IV) iron treatment, often referred to as “functional iron deficiency”; and (2) classical laboratory biomarkers, alone or in combination with each other, often referred to as “absolute iron deficiency.”

When a study used a response to IV iron treatment as the reference standard for iron deficiency, it allowed us to directly compare the test performances of classical versus newer biomarkers in predicting a response. To facilitate the interpretation of study results, the reported sensitivity and specificity of both newer and classical laboratory biomarkers were visually depicted in receiver operating characteristic (ROC) space. We did not conduct meta-analyses, because there was a high degree of heterogeneity across studies in the definitions of reference standard (a response to IV iron treatment), baseline iron status of the study populations, and background treatment.

When a study used classical laboratory biomarkers (alone or in combination with each other) as the reference standard for iron deficiency, we were prevented from comparing the test performance of classical biomarkers with newer biomarkers. For the purpose of our review, this approach is analogous to assessing the concordance between classical and newer biomarkers of iron status. Since concordance cannot tell us which test is better and which is worse—both may be equally bad or equally good for defining “iron deficiency”—and cannot answer Key Question 2, these studies were only included for subquestion 2a (What reference standards are used for the diagnosis of iron deficiency in studies evaluating test performance?).

**Test Performance Terms and Definitions**

- **Receiver operating characteristic curve:** ROC curves compare sensitivity with specificity across a range of values for the ability to predict a dichotomous outcome (in this case, defined as the reference standard). The ROC curve graphically displays the trade-off between sensitivity and specificity, and is useful in assigning the best cutoffs for clinical use.
- **Overall test accuracy**: Overall accuracy of a test is expressed as area under the ROC curve (AUC). The AUC provides another useful parameter for comparing test performance between, for example, classical and newer laboratory biomarkers of iron status. The AUC summarizes the ROC curve in a single number but loses information about the trade-offs between sensitivity and specificity.

- **Test accuracy**: Test accuracy refers to sensitivity (true positive rate) and specificity (true negative rate) of a test. For any test, there is usually a trade-off between sensitivity and specificity. For example, a test may exhibit a high sensitivity and a low specificity, or vice versa.

- **Diagnostic odds ratio (DOR)**: The DOR is a single indicator of test performance that combines the strengths of sensitivity and specificity. The DOR offers advantages when logistic regression is used with diagnostic problems, because the DOR is equivalent to the regression coefficient, after exponentiation. DORs are conditional: They depend on the other variables that have been used in the model. Consequently, the conditional DOR of each test variable, adjusted for the other variable (e.g., inflammation markers), can be estimated.

**Grading the Body of Evidence**

We followed the Methods Guide in evaluating the strength of the body of evidence for each Key Question with respect to four domains: risk of bias, consistency, directness, and precision. The body of evidence was rated on a four-level scale—high, moderate, low, and insufficient—on the basis of our degree of confidence that the evidence reflected the true effect for the major comparisons of interest. Briefly, a high level of evidence indicates a high confidence that the evidence reflects the true effect, and that further research is unlikely to change our confidence in the estimate of effect. A moderate level of evidence indicates a moderate confidence that the evidence reflects the true effect, and that further research may change our confidence in the estimate, or may change the estimate. A low level of evidence indicates a low confidence that the evidence reflects the true effect, and that further research is likely to change our confidence in the estimate of effect and to change the estimate.

The rating of the strength of the body of evidence was based on the consensus of all team investigators. We evaluated the applicability of included studies to each patient population of interest, that is, nondialysis patients with stage 3, 4, or 5 CKD, patients with CKD undergoing HD or PD, and patients with a kidney transplant. We evaluated and summarized studies of pediatric, adult, and elderly adults separately.

**Results**

The results of our literature searches are presented first, followed by the results of our syntheses in order by Key Questions. The majority of the included studies were related to test performance (Key Question 2), and they addressed many different laboratory markers and reference standard pairs. Thus, we organized studies included in Key Question 2 by types of test performance outcomes (predictive ability or test agreement).

**Literature Search**

Our literature search yielded 6,407 citations. From these, 694 articles were retrieved for full-text screening on the basis of abstracts and titles. Full-text articles were screened on the basis of
study eligibility criteria. A total of 664 articles were rejected on double, independent full-text screening because they did not meet one or more of the population, intervention, comparator, outcome (PICO) criteria for a particular Key Question. At the conclusion, a total of 30 articles were accepted, including 1 Polish- and 1 Japanese-language publication. Twenty-seven articles reported data on the test performance of newer markers of iron status compared with classical markers (Key Question 2),\textsuperscript{10,18-43} two reported intermediate outcomes comparing iron management guided by newer laboratory markers with iron management guided by classical markers (Key Question 3),\textsuperscript{42,44} and three (in two articles) reported data on factors affecting test performance comparing newer with classical laboratory markers of iron status (Key Question 4).\textsuperscript{45,46} Most studies enrolled only adult CKD patients undergoing HD. The main findings of this comparative effectiveness review are presented below.

**Key Question 1. Comparative Effectiveness of Newer Versus Older Markers of Iron Status for the Diagnosis and Management of Iron Deficiency Anemia**

No study reported on patient-centered outcomes (mortality, morbidity, quality of life, and adverse effects) when using newer laboratory markers as a replacement for or an add-on to the classical laboratory markers for assessing iron status and management of iron deficiency in stages 3–5 CKD nondialysis and dialysis patients, or in patients with a kidney transplant.

**Key Question 2. Test Performance of Newer Markers Compared With the Older Markers of Iron Status**

2a. Reference Standards for the Diagnosis of Iron Deficiency

A total of 27 studies were included for Key Question 2. Reviewed studies used two distinct methods to operationalize a reference standard for assessing test performance: (1) a response to intravenous (IV) iron treatment; and (2) classical laboratory biomarkers, alone or in combination. However, there were large variations across studies in the definitions of these reference standards.

Of the 27 included studies, 15 used classical markers of iron status to define “iron deficiency” as the reference standard in calculating the test accuracy (sensitivity and specificity) of newer markers of iron status;\textsuperscript{10,18-20,24,25,27,29-33,36,39,42} These studies used the following definitions for iron deficiency: (1) TSAT $\leq 15$ percent;\textsuperscript{24} (2) TSAT $\leq 20$ percent;\textsuperscript{18-20,29,33,39,42} (3) ferritin $\leq 100$ ng/mL;\textsuperscript{20} (4) TSAT $\leq 20$ percent and ferritin $\leq 100$ ng/mL;\textsuperscript{25,27,29-31,39} (5) TSAT $\leq 20$ percent or ferritin $\leq 100$ ng/mL;\textsuperscript{27,32,36,42} (6) serum iron $< 40$ µg/dL, TSAT $< 20$ percent, ferritin $< 100$ ng/mL, and Hb $< 11$ g/dL;\textsuperscript{10} (7) TSAT $< 20$ percent, ferritin 100–800 ng/mL, and Hb $< 11$ g/dL;\textsuperscript{10} and (8) TSAT $< 16$ percent and ferritin $< 12$ ng/mL.\textsuperscript{30} The remaining 12 studies investigated the test performance of newer or classical markers of iron status, using a response to IV iron treatment as the reference standard for diagnosis of iron deficiency.\textsuperscript{21-23,26,28,34,35,37,38,40,41,43} (As described in Methods, these 12 studies, which used a response to IV iron treatment as the reference standard for iron deficiency, allowed us to directly compare the test performance of classical versus newer biomarkers in predicting a response. Thus, the results from these studies were synthesized to answer Key Question 2.)

However, there existed a large heterogeneity in the reference standards used in these studies as well. The most commonly used definition for a response to IV iron treatment was an increase in Hb concentration $\geq 1$ g/dL after a (variable) period of IV iron treatment.\textsuperscript{21,22,38,40,43} Other
reference standards included a ≥ 15 percent increase in Hb, an increase in Hct of ≥ 3 percent and/or a ≥ 30 percent reduction in erythropoietin (EPO) dose, >1 point increase in corrected reticulocyte index, and 5 percent increase in Hct or a decrease in EPO dose of >2,000 units per treatment. It should be noted that there was no uniform regimen of IV iron in terms of dosage or iron formulation across these studies. IV iron treatment duration also varied widely. The potential impact of IV iron treatment regimen on the test performance of newer or classical laboratory markers of iron status is not known.

Comparisons of Test Performance of Newer Versus Classical Markers of Iron Status to Predict a Response to Intravenous Iron Treatment

Twelve studies (10 prospective cohorts, 1 retrospective cohort, and 1 cohort study of unknown directionality) investigated the test performance of newer or classical markers of iron status, using a response to IV iron treatment as the reference standard for diagnosis of iron deficiency. Of these, eight reported comparative data between five of the newer markers (no studies addressed SQUID) and the classical markers (although not all studies performed formal statistical testing for the comparisons). Seven of the eight enrolled adult hemodialysis (HD CKD) patients, and one study enrolled adult nondialysis (ND CKD) patients. The remaining four studies investigated the test performance of newer laboratory markers alone. Of these four, three enrolled adult HD CKD patients, and one enrolled adult peritoneal dialysis (PD CKD) patients. None of the reviewed studies enrolled pediatric CKD patients, and we did not include studies evaluating the test performance of classical markers alone.

Content of Hemoglobin in Reticulocytes (CHr)/Reticulocyte Hemoglobin Equivalent

Eight cohort studies, enrolling 533 adult HD CKD patients, 1 cohort study enrolling 23 PD CKD patients, and 1 cohort study enrolling 95 ND CKD patients evaluated the test performance of CHr to predict a response to IV iron treatment. Of the eight studies in HD CKD patients, six compared the test performance of CHr with that of classical markers of iron status (TSAT or ferritin, alone or in combination with each other), and two studies reported the test performance of CHr alone. Of these studies, one was rated as being at low risk of bias, four at a medium risk of bias, and three at a high risk of bias. Studies enrolled primarily older patients who received maintenance ESA treatment; however, maintenance ESA doses varied across studies. Baseline iron status (based on mean serum ferritin and TSAT concentrations) also varied across studies.

Overall, there is a low level of evidence that CHr has similar or better overall test performance compared with classical markers (TSAT or ferritin) to predict a response to IV iron treatment among HD CKD patients. Four different definitions of a response to IV iron treatment were used among these eight studies. Studies examined the sensitivities and specificities at different cutoff values of CHr, ranging from <26 to <32 pg, to predict iron deficiency, but the available data did not allow us to assess threshold effect, due to the heterogeneity in the definitions of reference standards. Additional heterogeneity, such as the variable iron status of the study populations and background treatment across studies, further limited our ability to make comparisons across studies.

Only two studies reported the sensitivities and specificities of classical markers (TSAT <20 or ferritin <100 ng/mL) to predict iron deficiency, and data suggest that CHr (with cutoff values of <27 or <28 pg) provides a better sensitivity and specificity in predicting iron deficiency than classical markers (TSAT <20 or ferritin <100 ng/mL). Only one study performed
multivariate analyses to predict a response to IV iron treatment (defined as an increase in Hct of ≥3 percent and/or a ≥ 30 percent reduction in EPO dose), and reported that CHr (with cutoff of <28 pg) had much higher diagnostic odds ratio than serum ferritin (with cutoff of <300 ng/mL). The strength of evidence is insufficient to draw conclusions regarding the test performance of CHr compared with that of classical markers of iron status among PD or ND CKD patients. We did not identify any study that evaluated the test performance of CHr to predict a response to IV iron treatment among pediatric CKD patients.

Percent Hypochromic Red Blood Cells

Six cohort studies, enrolling a total of 365 adult HD CKD patients, evaluated the test performance of %HYPO to predict a response to IV iron treatment. One study was rated as being at a low risk of bias, two at a medium risk, and three at a high risk of bias. Studies enrolled primarily older patients who received maintenance ESA treatment; however, maintenance ESA doses varied across studies. Baseline iron status (based on mean serum ferritin and TSTA concentrations) also varied across studies.

Overall, there is a low level of evidence that %HYPO has similar or better overall test accuracy compared with classical markers (TSAT or ferritin) to predict a response to IV iron treatment among HD CKD patients. Three different definitions of a response to IV iron treatment were used among these six studies. Studies examined the sensitivities and specificities of %HYPO, with a cutoff value of either >6 percent or >10 percent, to predict iron deficiency. Data suggest that %HYPO (with cutoff values of >6 percent or >10 percent) has a better sensitivity and specificity in predicting iron deficiency than classical markers (TSAT <20 or ferritin <100 ng/mL). In addition, two studies (from the same group of investigators) performed a multivariate regression analysis and showed that %HYPO was the only significant predictor of a response to IV iron treatment among all other markers included in the model.

We did not identify any study evaluated the test performance of %HYPO to predict a response to IV iron treatment among adult PD or ND CKD patients, or among pediatric CKD patients.

Soluble Transferrin Receptor

Two cohort studies, enrolling a total of 157 adult HD CKD patients, evaluated the test performance of sTfR to predict a response to IV iron treatment. Both studies also compared the test performance of sTfR with that of classical laboratory markers (TSAT or ferritin). One study was rated as being at a high risk of bias, and one at a medium risk of bias. The response to IV iron treatment was defined differently in the two studies, either as an increase in Hb concentration ≥1g/dL after intravenous iron treatment, or as an increase in Hb >15 percent from baseline.

Overall, there is a low level of evidence that sTfR has similar overall test accuracy compared with classical markers (TSAT or ferritin) to predict a response to IV iron treatment (although defined differently in the two studies) among HD CKD patients. We did not identify any study that evaluated the test performance of sTfR to predict a response to IV iron treatment among adult PD or ND CKD patients, or among pediatric CKD patients.

Erythrocyte Zinc Protoporphyrin

Two cohort studies, enrolling a total of 187 adult HD CKD patients, evaluated the test performance of ZPP in predicting a response to IV iron treatment. Both studies also compared the test performance of ZPP with that of classical laboratory markers (TSAT or
ferritin). However, because the reference standards (Hb versus Hct/decrease in EPO dose) were not comparable, the two studies were evaluated separately. Therefore, the strength of evidence is insufficient to draw conclusions regarding the overall test performance or test accuracy of ZPP compared with that of classical laboratory markers (TSAT or ferritin).

We did not identify any study that evaluated the test performance of ZPP to predict a response to IV iron treatment among adult PD or ND CKD patients, or among pediatric CKD patients.

**Hepcidin**

One prospective cohort study evaluated the test performance of both isoforms of hepcidin (hepcidin-20 and hepcidin-25) to predict iron deficiency among 56 older adult HD CKD patients who were on maintenance ESA treatment. The study was rated as being at a low risk of bias. The strength of evidence is insufficient to draw conclusions regarding the test performance of hepcidin-20 or hepcidin-25 comparing with that of classical markers of iron status among adult HD CKD patients.

We identified no study evaluating the test performance of hepcidin to predict a response to IV iron treatment among adult PD or ND CKD patients, or among pediatric CKD patients.

2b. Adverse Effects or Harms Associated With Testing

Only 7 of the 27 identified studies reported information on harms.23,26,35,40-43 Specifically, three studies reported no adverse events associated with iron therapy during the study periods. A total of five deaths were reported across two studies. Studies did not attribute these deaths to either testing or treatment. However, iron testing itself is unlikely to cause deaths, and most of the reported harms were attributed to iron therapy (if reported).

**Key Question 3. Intermediate Outcomes Comparing the Iron Management Guided by the Newer Laboratory Markers With That Guided by the Older Laboratory Markers**

Two short-term RCTs (4 and 6 months), enrolling a total of 354 adult CKD patients (mean age 60 years old) undergoing HD, compared the intermediate outcomes of iron management guided by classical markers of iron status (TSAT and/or ferritin) with those of iron management guided by a newer marker of iron status (CHr). It should be noted that the two trials (one in the United States and one in Japan) employed different protocols for initiating intravenous iron therapy and anemia management, which affect the applicability of the trial findings.

The two trials showed different findings in terms of the doses of epoetin required to maintain hematocrit (Hct) targets. Specifically, the U.S. trial showed that guiding iron management via CHr resulted in similar epoetin dosing compared with iron management guided by ferritin or TSAT. In contrast, the Japanese trial found the doses of epoetin were significantly decreased (lower by 36 percent) in the group guided by TSAT, but did not change significantly in the group guided by CHr. However, it should be noted that the Hct target was higher in the U.S. trial, which may explain why the U.S. trial used much higher doses of epoetin than the Japanese trial during the trial period. Despite the differences in the protocols for initiating intravenous iron therapy, both trials reported a significant decrease in the intravenous iron doses administered to patients whose iron management was guided by CHr compared with those guided by TSAT or ferritin. Only the Japanese trial specifically monitored the adverse events associated with study
medication; no differences in the hospitalization or infection rates between the two iron management groups were reported.

There is a low level of evidence for a reduction in the number of iron status tests and resulting intravenous iron treatments needed to maintain target hematocrit in patients whose iron management was guided by CHr compared with those guided by TSAT or ferritin, with similar or lower ESA use. Both RCTs reported that Hct remained in the targeted ranges (an indication for the adequacy of anemia management) throughout the study period in all randomized arms, although the Hct target was higher in the U.S. trial than the Japanese trial. We identified no study comparing iron management guided by classical markers with that guided by newer markers (%HYPO, sTfR, Ret-He, ZPP, or hepcidin).

Key Question 4. Factors Affecting Test Performance and Clinical Utility

Only a single study or indirect comparisons across studies provided data on the potential impacts of some factors (e.g., interactions between iron and ESA treatment, route of iron administration, and treatment regimen) on the test performance of newer or classical laboratory markers of iron status. Therefore, the strength of evidence is insufficient to draw conclusions regarding factors that may affect the test performance or clinical utility of laboratory markers of iron status.

Interactions Between Iron and ESA Treatment

One trial randomized 134 HD CKD patients to either no IV iron or IV iron (1 gram of ferric gluconate).45 This trial was rated as being at a medium risk of bias and enrolled a special population of HD CKD patients with high ferritin (500-1200 ng/mL) and low TSAT levels (≤ 25 percent), possibly due to functional iron deficiency. Baseline epoetin doses were raised by 25 percent in both groups, starting with the first hemodialysis session of week 1 and then maintained for the entire study until the first hemodialysis session of week 6.

Within the no-intravenous-iron group (25 percent epoetin dose increase alone), the sensitivity and specificity pairs for a TSAT cutoff of ≥19 percent and a ferritin cutoff of ≥726 ng/mL were 29 and 70 percent, and 27 and 69 percent, respectively. The sensitivity and specificity pairs for a CHr cutoff of ≥31.2 pg and a sTfR cutoff of ≥5.9 mg/L were 27 and 69 percent, and 35 and 77 percent, respectively.

In contrast, in the intravenous iron group, a cutoff of CHr of ≥31.2 pg had a higher sensitivity (64 percent) and specificity (75 percent) in predicting treatment response. However, the test accuracies were lower for sTfR, TSAT, and ferritin.

Use of Different Diagnostic Reference Standards

Included in Key Question 2a, one study examined the test performance of RetHe using two different reference standards, and showed that the test performance of RetHe was less favorable for assessing “functional iron deficiency” (TSAT<20 percent, ferritin 100-800 ng/mL, and Hb <11 g/dL) than for assessing “traditional parameters for iron deficiency” (serum iron < 40 µg/dL, TSAT<20 percent, ferritin <100 ng/mL, and Hb <11 g/dL) in HD CKD patients.10 The heterogeneity in the definitions for the reference standard (a response to IV iron treatment) may explain the differences in study findings.
Discussion

Key Findings and Strength of Evidence

We did not identify any study that provided data directly addressing our overarching question regarding the impact on patient-centered outcomes (mortality, morbidity, quality of life, and adverse effects) of using newer laboratory biomarkers. In the absence of direct evidence, the overarching question could be answered by the component questions (Key Questions 2, 3, and 4). A number of studies addressing these component questions were identified. A summary of the strength of evidence addressing each Key Question is provided in Table A.
<table>
<thead>
<tr>
<th>Key Question</th>
<th>Strength of Evidence</th>
<th>Summary, Comments, and Conclusions</th>
</tr>
</thead>
</table>
| **Key Question 2. What is the diagnostic test accuracy of newer markers of iron status as a replacement for or an add-on to classical laboratory markers?** | Low / Insufficient (depending on the test comparisons, study populations, or test performance outcomes) | • Among adult HD CKD patients, there is a low level of evidence that:  
  o CHr has similar or better overall test accuracy compared with TSAT or ferritin to predict a response to IV iron treatment. Data from two studies suggest that CHr (with cutoff values of <27 or <28 pg) has a better sensitivity and specificity in predicting iron deficiency than classical markers (TSAT <20 or ferritin <100 ng/mL).  
  o %HYPO has similar or better overall test accuracy compared with TSAT, and better overall test accuracy compared with ferritin to predict a response to IV iron treatment. Data suggest that %HYPO (with cutoff values of >6% or >10%) has a better sensitivity and specificity to predict iron deficiency (as defined by a response to IV iron treatment) than classical markers (TSAT <20% or ferritin <100 ng/mL).  
  o sTfR has a similar test performance compared with classical markers (TSAT or ferritin) to predict a response to IV iron treatment.  
• There is insufficient evidence regarding:  
  o Test performance of newer markers of iron status as an add-on to older markers.  
  o Test performance comparing ZPP and hepcidin to predict a response to IV iron treatment in adult HD CKD patients.  
  o Test performance comparing newer with classical laboratory markers to predict a response to IV iron treatment, in adult PD CKD and ND CKD patients, and in pediatric CKD patients. |
| **Key Question 2a. What reference standards are used for the diagnosis of iron status in studies evaluating test accuracy?** | Not rated (descriptive data) | • There is a lack of generally accepted reference standard tests for determining iron deficiency in the context of CKD. This is reflected by the fact that current studies use two distinct methods to operationalize a reference standard for assessing test performance: (1) a response to intravenous (IV) iron treatment, often referred as “functional iron deficiency”; and (2) classical laboratory biomarkers, alone or in combination with each other, often referred as “absolute iron deficiency.” However, across studies, the definitions of these reference standards vary widely. |
| **Key Question 2b. What are the adverse effects or harms associated with testing using newer and/or older markers of iron status?** | Insufficient | • Only 7 of the 27 studies reported information:  
  o 3 studies reported no adverse events associated with iron therapy during the study periods.  
  o A total of 5 deaths reported. Studies did not attribute these deaths to either testing or any treatment.  
  o Most of the reported harms were attributed to iron therapy. |
| **Key Question 3. What is the impact of managing iron status based on newer laboratory biomarkers either alone or in addition to older laboratory biomarkers on intermediate outcomes?** | Low | • Two short-term RCTs (4 and 6 months) showed a reduction in the number of iron status tests and resulting intravenous iron treatments (a post-hoc intermediate outcome) administered to patients whose iron management was guided by CHr compared with those guided by TSAT or ferritin.  
• Both RCTs reported that Hct remained in the targeted ranges (an indication for the adequacy of anemia management) throughout the study period in all randomized arms, although the Hct target differed between the two trials.  
• One trial showed that guiding iron management via CHr resulted in similar epoetin dosing compared with iron management guided by ferritin or TSAT. In contrast, the other trial found doses of epoetin were significantly decreased (lower by 36 percent) in the group guided by TSAT, but did not change significantly in the group guided by CHr.  
• No study compared iron management guided by classical markers with that of newer markers (%HYPO, sTfR, Ret-He, ZPP, or hepcidin). |
Table A. Summary of the strength of evidence addressing Key Questions (continued)

<table>
<thead>
<tr>
<th>Key Question</th>
<th>Strength of Evidence</th>
<th>Summary, Comments, and Conclusions</th>
</tr>
</thead>
</table>
| Key Question 3a. What are the adverse effects or harms associated with the treatments guided by tests of iron status? | Insufficient | • Only 1 RCT explicitly monitored the adverse events:  
  o There were a total of three deaths (2 patients in the CHr group; 1 patient in the TSAT group) due to bacterial pneumonia (at week 4 in the CHr group), sudden death by unknown cause (at week 16 in the CHr group), and liver tumor (at week 7 in the TSAT group).  
  o One patient in the TSAT group dropped out because of massive bleeding due to a femoral bone fracture and need for blood transfusion.  
  o There were no significant differences in the hospitalization or infection rates of the two iron management groups. |
| Key Question 4. What factors affect the test performance and clinical utility of newer markers of iron status? | Insufficient | • Only single study or indirect comparisons across studies provided data on the potential impacts of some factors on the test performance of newer or classical laboratory markers of iron status:  
  o One RCT found an interaction between iron and ESA treatment on test accuracy of CHr. A higher baseline CHr predicted greater likelihood of a response to anemia and iron treatment only in the IV iron (plus epoetin) treatment group, but not in the no IV iron (epoetin only) treatment group.  
  o One study showed that the test accuracy of RetHe was lower for assessing “functional iron deficiency” (TSAT<20%, ferritin 100-800 ng/mL, and Hb <11 g/dL) than for assessing “traditional parameters for iron deficiency” (serum iron < 40 µg/dL, TSAT<20%, ferritin <100 ng/mL, and Hb <11 g/dL) in HD CKD patients.  
  o Indirect comparisons across studies suggested potential impacts of route of iron administration and treatment regimen on the test accuracy of newer and classical laboratory markers of iron status.  
  • No study performed analyses by patient subgroups.  
  • No study examined the impacts of biological variation or type of dialysis in diagnostic indices on the test performance or clinical utility of laboratory markers of iron status. |

%HYPO=percent hypochromic red blood cells; CHr=content of hemoglobin in reticulocytes; CKD=chronic kidney disease; ESA=erythropoiesis-stimulating agents; Hb=hemoglobin; HD=hemodialysis; IV=intravenous; ND=nondialysis; PD=peritoneal dialysis; RCT=randomized controlled trial; sTfR=soluble transferring receptor RetHe=reticulocyte hemoglobin equivalent; TSAT=transferrin saturation; ZPP=erythrocyte zinc protoporphyrin

Findings in Relationship to What Is Already Known

Our findings are consistent with the recommendations in the Kidney Disease Outcome Quality Initiative (KDOQI) and the National Institute for Health and Clinical Excellence (NICE) guidelines for anemia management in CKD.1,6 These guidelines recommend that the initial assessment of iron deficiency anemia include ferritin to assess iron stores, and serum TSAT or CHr (KDOQI) or %HYPO (NICE) to assess adequacy of iron for erythropoiesis. We found that there is a low level of evidence that both CHr and %HYPO have a similar or better overall test accuracy compared with classical markers (TSAT or ferritin) to predict a response to IV iron treatment among HD CKD patients. Our confidence in the totality of evidence was limited by the heterogeneity and potential risk of bias in the body of literature (see “Limitation of the Evidence Base” for more details). In addition, many important questions remain unanswered, such as the test performance of newer markers of iron status as an add-on to older markers and factors that may affect the test performance or clinical utility of laboratory markers of iron status.

We identified one study showing an improvement in the test performance by using a combination of laboratory biomarkers, such as the combination of %HYPO >6 with TSAT ≤20 percent, the combination of %HYPO >6 percent with CHr ≤29 pg, and the combination of
%HYPO >6 with ZPP >52 µmol/mol.\(^{37}\) However, there are potentially a large number of test combinations to be evaluated, and without a widely accepted reference standard for the diagnosis of iron deficiency in the context of CKD, new studies are unlikely to significantly contribute to what is already known or change existing clinical practice.

**Applicability and Implications for Clinical and Policy Decisionmaking**

We assessed the applicability of the included studies by organizing them according to each patient population of interest, that is, nondialysis patients with stage 3, 4, or 5 CKD, patients with CKD undergoing HD or PD, or patients with a kidney transplant. A majority of this review’s findings are applicable to only adult HD CKD patients. Whether test performance and clinical utility of newer or classical markers of iron status vary by different CKD populations are not known.

We identified two RCTs that compared intermediate outcomes of iron management guided by CHr with those of iron management guided by classical markers of iron status (TSAT and/or ferritin).\(^{42,44}\) These two trials (one conducted in the United States and one in Japan) employed different protocols for initiating IV iron therapy and anemia management. These differences may reflect differences in the healthcare systems of their respective countries, and should be considered as part of clinical decisionmaking.

**Limitations of the Evidence Base**

The available data are very limited due to a high degree of heterogeneity. There exist many definitions of a response to IV iron treatment as the reference standard for iron deficiency. Moreover, there is a lack of a uniform regimen for intravenous iron treatment across studies in terms of dosage, iron formulation, treatment frequency, and followup duration for the iron challenge test (to define a response).

In addition to heterogeneity of the evidence base, many studies included in our review were rated as being at a high risk of bias, limiting their utility in informing clinical practice.

**Research Gaps**

The most directly applicable study designs for clinical decisionmaking would be studies that compare two or more iron and anemia management strategies, follow the patients through decisions and treatments, and then report on patient outcomes. However, it is unlikely such studies can be conducted, due to the large number of patients and resource requirements. Typically, the assessment of diagnostic tests follows the Fryback approach,\(^{47}\) progressing from the establishment of technical and clinical validity, to the assessment of test impact on clinicians’ diagnostic thinking and therapeutic decisionmaking, as well as clinical outcomes. Finally, a global assessment of the test from a societal perspective can be performed. Thus, we suggest that future research address the gaps that we identified for each of the component questions in this review. We also identified several cross-cutting methodological issues that affect all of the Key Questions and should be addressed. Ultimately, when a reference standard of iron deficiency is finally established, and test performance data are sufficient and reliable, decision analysis could be used to assess how employing combinations of different markers to guide iron management strategies might influence clinical outcomes.
A summary of the research gaps we identified, as well as our suggestions for future research, are provided in Table B.

Table B. Research gaps and suggestions for future research

<table>
<thead>
<tr>
<th>Key Question</th>
<th>Research Gaps</th>
<th>Suggestions for Future Research</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Key Question 2. What is the diagnostic test accuracy of newer markers of iron status as a replacement for or an add-on to classical laboratory markers?</strong></td>
<td>Insufficient evidence for the test performance of newer markers of iron status as an add-on to older markers</td>
<td>• It is important to use an independent reference standard when assessing the test performance. See “Cross-cutting issues” for the research gaps for establishing a reference standard for iron deficiency.</td>
</tr>
<tr>
<td></td>
<td>Many existing studies are at a high risk of bias, limiting their utility in informing clinical practice</td>
<td>• General principles for the design of studies of diagnostic tests include the use of an appropriate reference standard, adequate description of the index and reference tests, blinded interpretation of test results, and independence of the index and reference standard tests.48</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Studies assessing diagnostic accuracy should instead aim to enroll patients representative of the spectrum of disease typically seen in clinical practice.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Future studies should provide details about the study base and sampling methods.</td>
</tr>
<tr>
<td><strong>Key Question 3. What is the impact of managing iron status based on newer laboratory biomarkers either alone or in addition to older laboratory biomarkers on intermediate outcomes?</strong></td>
<td>There is no uniform iron management algorithms across studies</td>
<td>• Future observational studies should assess the outcomes of different iron management algorithms or test-and-treat protocols, considering differences in CKD populations, clinical settings, and potential harms or burden to the patients.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Assessing impact of the most promising iron management algorithms on both intermediate and patient outcomes through prospective observational studies or RCTs.</td>
</tr>
<tr>
<td><strong>Key Question 4. What factors affect the test performance and clinical utility of newer markers of iron status?</strong></td>
<td>Insufficient evidence to draw conclusions regarding factors that may affect the test performance or clinical utility of laboratory markers of iron status</td>
<td>• Future studies are need to evaluated the following factors, suggested by the experts:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>o Biological variation in diagnostic indices</td>
</tr>
<tr>
<td></td>
<td></td>
<td>o Use of different diagnostic reference standards</td>
</tr>
<tr>
<td></td>
<td></td>
<td>o Type of dialysis (i.e., peritoneal or hemodialysis)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>o Patient subgroups (i.e., age, sex, comorbid conditions, erythropoiesis-stimulating agent resistance, protein energy malnutrition secondary to an inflammatory state, hemoglobinopathies [e.g., thalassemia and sickle cell anemia])</td>
</tr>
<tr>
<td></td>
<td></td>
<td>o Route of iron administration (i.e., oral or intravenous)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>o Treatment regimen (i.e., repletion or continuous treatment)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>o Interactions between treatments (i.e., patients treated with versus without ESA, patients treated with vs. without iron-replacement therapy)</td>
</tr>
<tr>
<td></td>
<td>Whether test performance and clinical utility of newer or classical markers of iron status vary by different CKD populations are not known</td>
<td>• Almost all existing studies enrolled only single CKD population (ND, HD, or PD CKD patients). Future studies should include wider CKD populations, and plan for subgroup analyses.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Power calculations should be performed to take into account for the planed subgroup analyses.</td>
</tr>
</tbody>
</table>
Table B. Research gaps and suggestions for future research (continued)

<table>
<thead>
<tr>
<th>Key Question</th>
<th>Research Gaps</th>
<th>Suggestions for Future Research</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross-cutting issues (for Key Question 2, 3, and 4)</td>
<td>There is no reference standard for determining iron deficiency in CKD patients</td>
<td>• A response to IV iron treatment is considered by many clinicians as the reference method for diagnosing iron deficiency but future research is needed to establish a standardized definition for appropriate CKD populations, and a standardized testing protocol specifying the regimen of IV iron challenge in terms of dosage and iron formulation and proper duration of iron challenge testing.</td>
</tr>
<tr>
<td></td>
<td>Existing studies were underpowered leading to imprecise estimates</td>
<td>• Future studies should be larger, ideally designed based on power calculations, to be able to reliably detect plausible effect sizes and provide precise estimates of diagnostic accuracy.49</td>
</tr>
</tbody>
</table>
| | There is no decision analysis to assess how using combinations of different markers to guide iron management strategies might influence clinical outcomes | • Patient outcomes of interest are  
  ○ Mortality  
  ○ Morbidity (e.g., cardiac or liver toxicity and infection)  
  ○ Quality of life, measured using standardized scales, including: Kidney Disease Quality of Life (KDQOL), Health Related Quality of Life (HRQOL), Medical Outcomes Study Short Form-36 (SF-36), and Pediatric Quality of Life Inventory (PQLI)  
  ○ Adverse effects or harms associated with testing and associated treatments (e.g., test-related anxiety, adverse events secondary to venipuncture, effects of iron overload with iron treatments, and cardiovascular complications from use of erythropoietin at higher Hb levels)  
  • For studies assessing clinical outcomes, blinding to test results to the outcome assessors is essential to avoid bias.48,50 |

CKD=chronic kidney disease; HD=hemodialysis; IV=intravenous

Conclusions

Combining the evidence addressing Key Questions 2, 3, and 4, we can conclude that all currently available laboratory biomarkers of iron status (either newer or classical markers) do not demonstrate an ideal predictive ability when they were used singly to determine iron deficiency as defined by a response to iron challenge test. Furthermore, there is insufficient evidence to determine the test performance of the combinations of newer biomarkers, or combinations of newer and classical biomarkers, for diagnosing iron deficiency. However, it may be that CHr and %HYPO have better predictive ability for a response to IV iron treatment than classical markers (TSAT <20% or ferritin <100 ng/mL) in HD CKD patients. In addition, results from two RCTs showed a reduction in the number of iron status tests and resulting IV iron treatments administered to patients whose iron management was guided by CHr compared with those guided by TSAT or ferritin. These results suggest that CHr may reduce potential harms from IV iron treatment by lowering the frequency of iron testing, although the evidence for the potential harms associated with testing or test-associated treatment is insufficient.

Nevertheless, the strength of evidence supporting these conclusions is low and there remains considerable clinical uncertainty regarding the use of newer markers in the assessment of iron status and management of iron deficiency in stages 3–5 CKD patients (both nondialysis and dialysis). In addition, factors that may affect the test performance and clinical utility of newer laboratory markers of iron status remain largely unexamined.
References


Introduction

Chronic kidney disease (CKD) is the gradual, progressive deterioration of kidney function leading to a toxic accumulation of wastes inside the body, which in turn gives rise to complications such as high blood pressure, decreased bone health, nerve damage, and anemia. The most common causes of CKD are diabetes and hypertension, though others include glomerulonephritis, inherited diseases such as polycystic kidney disease, congenital malformations of the kidney, autoimmune disorders such as lupus, and mechanical obstructions and chronic infections of the urinary tract. CKD patients are classified as having progressed to one of five stages, depending on the severity of their condition (CKD stage 1-5). When CKD progresses to its end stage (stage 5), dialysis or kidney transplantation become necessary.

CKD currently affects an estimated 26 million American adults, with a far higher number considered at risk. In addition to the significant detriment to the physical, mental, and social health of patients and their families that it poses, CKD comprises a tremendous individual and global financial burden.

Background

Chronic Kidney Disease and Iron Management

Anemia is a common complication of CKD which develops early in the course of CKD and becomes increasingly severe as the disease progresses. Anemia remains common among patients presenting for renal transplantation, and persists in the post-transplant period. Anemia, with its associated fatigue, cognitive impairment, and diminished quality of life, is a significant problem for dialysis patients. According to the United States Renal Data System, 67 percent of patients initiating dialysis had hemoglobin (Hb) values below 11.0 g/dL. The most common cause of anemia in dialysis patients is inadequate erythropoietin production due to kidney damage. The second most common cause, iron deficiency, stems from inadequate diet and absorption, procedure-related iron losses from repeated laboratory testing, and blood retention in the dialyzer and tubing during dialysis.

Despite its prevalence, anemia is generally treatable, and antianemic therapy is associated with reductions in mortality, morbidity, hospitalization, and medical costs in dialysis patients. Before the development of erythropoietic stimulating agents (ESAs), blood transfusion was the primary treatment option for anemia associated with CKD. Now the management of anemia in CKD patients requires an appropriate balance between stimulating the generation of erythroblasts (erythropoiesis) and maintaining sufficient iron levels for optimum Hb production. ESAs are analogues of the natural hormone erythropoietin produced by the kidneys, the primary site of erythropoietin production in the adult. Erythropoietin enhances the growth and differentiation of erythroid progenitors. With increasing renal dysfunction, decreased levels of erythropoietin are observed, resulting in progressive anemia. With the advent of ESA therapy, the risk for transfusion-related complications (e.g., transfusion-transmitted infection, transfusion reactions, immunologic sensitization, and iron overload) has been substantially reduced. ESAs mobilize iron stores in promoting erythropoiesis; however, decreased iron stores or iron availability are the most common reasons for resistance to the effect of ESAs. Thus, most patients who receive ESA treatment will require supplemental (oral or intravenous) iron to ensure an adequate response with erythropoietic agents. For this reason, iron management is an
essential part of the treatment of anemia associated with CKD, as there are concerns regarding the adverse effects associated with elevated doses of ESAs and supplemental iron.

Guidelines regarding the monitoring of iron deficiency and subsequent regimen of iron supplementation in patients on maintenance hemodialysis were first published by the National Kidney Foundation as part of their Kidney Disease Outcome Quality Initiative (KDOQI) in 1997, and then updated in 2000 and 2006. These guidelines describe the protocol to be followed in the management of anemia in CKD patients, including monitoring of iron status. As per the guidelines, Hb testing should be carried out annually in all patients with CKD, and such patients should be treated with ESAs when anemia is detected. Additionally, the guidelines stipulate that hemodialysis patients receiving erythropoietin should be monitored for iron deficiency using percent saturation of transferrin (TSAT, calculated as iron/total iron-binding capacity × 100), and serum ferritin (referred to as “ferritin”) concentrations every 3 months. However, the KDOQI guideline noted that there are no studies that have addressed the clinical benefit, cost-effectiveness, or risk benefit comparison of using different TSAT and ferritin levels for the diagnosis of iron deficiency. Older markers like serum iron and stainable iron in bone marrow are no longer used for monitoring in CKD patients. Serum iron is currently only assessed to aid in the calculation of TSAT. When treatment is required, the guidelines recommend the administration of sufficient iron to maintain a TSAT >20 percent and ferritin >100 ng/mL (>200 ng/mL for CKD patients on hemodialysis). Use of iron status markers is integral to assessment of deficiency, and to setting treatment goals in the successful management of anemia and iron deficiency in CKD patients. The National Kidney Foundation guidelines have been widely adopted in dialysis centers across the United States.

**Laboratory Biomarkers of Iron Status**

Assessing iron status is integral to both iron and anemia managements in CKD patients, as iron is essential for Hb formation (as is erythropoietin). Bone marrow iron stores are often regarded as the best indicator of iron status (although this is not universally accepted); however, taking a bone marrow sample is invasive and carries the risks of infection or bleeding at the biopsy site. Other classical iron status tests, of which ferritin and TSAT are the most widely used, reflect either the level of iron in tissue stores or the adequacy of iron for erythropoiesis. Serum ferritin reflects storage iron—iron that is stored in liver, spleen, and bone marrow reticuloendothelial cells. The percent TSAT (serum iron multiplied by 100 and divided by total iron binding capacity [TIBC]) reflects iron that is readily available for erythropoiesis. The TIBC essentially measures circulating transferrin. The transferrin molecule contains two binding sites for transporting iron from iron storage sites to erythroid progenitor cells. A TSAT of 50 percent indicates that half of the binding sites are occupied by iron. TSAT and ferritin level are individually most accurate as predictors of iron deficiency or iron overload when it is either extremely low (TSAT) or extremely high (ferritin).

Though widely used, current laboratory biomarkers of iron status are not without drawbacks when used in CKD patients: CKD is a pro-inflammatory state, and the biological variability of serum iron, transferrin saturation, and ferritin is known to be large in the context of underlying inflammation. This is because transferrin and ferritin are both acute-phase reactants, and in the presence of an inflammatory condition, transferrin concentration decreases and ferritin concentration increases. There is also considerable variability in comparisons of different assays used to measure serum iron.
Assessing the accuracy and reliability of laboratory biomarkers of iron status is likewise problematic, due to the lack of an established reference standard for these assays. This gap engenders an unavoidable component of measurement error in the reference standard used to assess diagnostic performance. Stainable iron from a bone marrow biopsy was previously used as a “gold standard,” but this is seldom performed, as bone marrow biopsy involves risks of infection or bleeding at the biopsy site. Further complicating the matter, patients with CKD may suffer from different manifestations of iron deficiency, including absolute iron deficiency (inadequate supply of iron in the body), functional iron deficiency (adequate supply but inefficient assimilation from body stores), and an extreme case of functional iron deficiency known as reticuloendothelial blockade (inadequate release of stored iron from macrophage cells of the body). These are typically identified by interpreting combinations of changes in the levels of ferritin and TSAT. The particular type of iron deficiency may affect the validity and reliability of laboratory test results for iron status and thus result in a dilemma regarding treatment decisions.

In an attempt to find a more accurate and reliable test, several novel biomarkers of iron status have been proposed. These may address the disadvantages of using ferritin and TSAT in a pro-inflammatory state in CKD patients. Figure 1 provides an overview of iron metabolism in the body, and the role of classical as well as newer laboratory biomarkers in assessing the status of iron status. The figure indicates that these newer markers assess aspects of iron metabolism that are not assessed by those in current use, with the exception of the paramagnetic assessment of iron in the liver using Superconducting QUantum Interference Device (SQUID). These newer markers, highlighted in yellow, are thought to be less influenced by the underlying state of inflammation in CKD, and their measurement more accurately reflects the state of iron supply and demand, as compared with older markers.

As illustrated in Figure 1, three markers assess the impact of iron deficiency on formation and composition of red blood cells (RBC), usually in the context of increased demand brought on by ESA use (functional iron deficiency). The Hb content of reticulocytes (CHr) is a function of the amount of iron in the bone marrow that is available for incorporation into reticulocytes (immature RBCs)—decreased levels of CHr indicate iron deficiency. Another is the percentage of hypochromic erythrocytes (%HYPO). This is a measurement of Hb in RBC, which factors in the absolute Hb content as well as the size of the RBC. This can be used to measure functional iron deficiency. (If iron supply is low in the face of ESA therapy, then there is lesser amount of Hb being incorporated into each RBC, and as a result, %HYPO levels are high.) However, this test cannot be used on stored blood, as storing blood samples causes an increase in RBC size, leading to invalid %HYPO results. The third, erythrocyte zinc protoporphyrin (ZPP) is a measure of iron incorporation in heme. When iron levels are low, zinc is used instead of iron in the formation of heme, a protein component of Hb. As a result, ZPP levels increase, indicating iron deficiency.

A fourth marker, soluble transferrin receptor (sTfR), measures the availability of iron in the bone marrow. When the bone marrow is stimulated by ESAs, it results in increased expression of transferrin receptors on the surface of erythroblasts, the precursors of RBC. If iron supply is low, then levels of transferrin containing iron are low, and there is a mismatch between the numbers of transferrin receptors and the transferrin-iron complexes to bind with them. Some of the transferrin receptors which are not bound by iron-containing transferrin then get detached and can be detected in the blood. Increased concentration of sTfRs in the blood is an indicator of iron deficiency.
Another lesser known marker, hepcidin, a peptide produced by the liver that regulates both iron absorption in the intestine as well as release of iron from macrophages, has also been suggested as a marker of iron deficiency in CKD patients. Increased levels of hepcidin have indeed been associated with a decrease in available iron.30

It has also been hypothesized that paramagnetic assessment of iron in the liver could indicate deficiency in iron stores, but this test has only been used in the context of iron overload.31

Figure 1. Roles of current and newly proposed markers of iron status

Although a number of international guidelines have examined the use of both classical and new serum iron biomarkers, their recommendations differ. Across guidelines, it is agreed that the optimal management of anemia in hemodialysis patients depends on accurate assessment of iron status. However, a number of questions remain, including: Which combination of iron biomarkers is required? Should the newer biomarkers be used as a replacement for or in addition to classical markers?

Accurate assessment and careful management of iron status is expected to garner increased attention following the Centers for Medicare and Medicaid Services’ recent adoption of a bundled reimbursement system for dialysis, where payments are made for groups of services rather than for individual treatments.32 In view of this development and considerable clinical
uncertainty, the high biological variability associated with laboratory biomarkers, and the need for frequent assessment to guide treatment for anemia, a systematic review of the relevant literature is of priority. The focus of the current review is to evaluate the strength of evidence for using these newly suggested markers, either as replacements for or additions to currently used markers, in managing iron-replacement therapy in patients with CKD.

Scope and Key Questions

Scope of the Review

The purpose of this review is to evaluate the impact on patient-centered outcomes of the use of newer versus classical laboratory biomarkers of iron status as part of the management strategies for anemia in patients with stages 3-5 CKD patients, that is, nondialysis or dialysis, or kidney-transplant patients. The newer laboratory biomarkers of interest include CHr, %HYPO, ZPP, sTfR, hepcidin, and SQUID. The classical laboratory biomarkers of interest include bone marrow iron stores, serum iron, TSAT, iron-binding capacity, and ferritin. These parameters were defined a priori with input from a panel of Key Informants and clinical experts (see Topic Refinement and Review Protocol for more details on the process).

As test results have little direct impact on patient-relevant outcomes, the utility of a medical test is usually determined by its indirect effect on outcomes, that is, through its influence on therapeutic decisionmaking and subsequently on patient outcomes. Although studies that assess the overall impact of tests on the clinical management process would provide the most direct evidence for this CER, they are often challenging or infeasible to conduct. Because we expected to find little of such evidence, the question of overall impact (Key Question 1, see below for full descriptions of all Key Questions) was broken out into three component Key Questions (Key Questions 2 to 4). Combining evidence gather to address these three component Key Questions can thus inform the conclusions for this reviews primary, overarching question.

Key Questions and Analytic Framework

Figure 2 depicts the analytic framework used in structuring this report. Broadly, it shows how the individual Key Questions are addressed within the context of the Populations, Interventions, Comparators, and Outcomes of interest.

Key Question 1 subsumes Key Questions 2, 3 and 4, which collectively address the impact on patient centered outcomes of using the newer laboratory biomarkers as a replacement for or in addition to classical laboratory biomarkers of iron status for assessing and management of iron deficiency. Specifically, Key Question 2 addresses the performance of newer markers of iron status as a replacement for or in addition to classical markers, and Key Question 3 focuses on comparative studies of management strategies where treatment decisions are guided by test results. Since these tests are also used for monitoring purposes (e.g., predict a response to intravenous iron treatment or setting treatment targets), treatment decisions may be altered by results of the subsequent tests at every time point of their measurement. In this way, the impact of testing on outcomes is mediated through a series of treatment decisions. We aim to capture “test effectiveness” by incorporating management strategies. Additionally, we aim to evaluate whether newer laboratory markers represent iron status, and better define (with respect to older markers) targets for iron therapy.
Tests of iron status as well as the treatments guided by these tests may be associated with adverse effects or harms. These can be related to testing directly, such as test-related anxiety, adverse events secondary to venipuncture, or indirectly, through downstream treatment decisions that were influenced by testing, such as iron overload with iron treatments. Sub-Key Question 2b and 3a address these potential harms.

Key Question 4 addresses the factors that may affect test performance and clinical utility of newer markers of iron status, such as biological variation in diagnostic indices, use of different diagnostic reference standards, and patient subgroups.

The full text of the Key Questions addressed in this report appears below.
Figure 2. Analytic framework

CKD=chronic kidney disease; ESA=erythropoiesis-stimulating agents; Hb=hemoglobin level
Key Question 1 (Overarching Question)
What is the impact on patient centered outcomes of using newer laboratory biomarkers\(^a\) as a replacement for or an add-on to the older laboratory biomarkers of iron status\(^b\) for the assessing iron status and management of iron deficiency in stages 3-5 CKD patients (nondialysis and dialysis), and in patients with a kidney transplant?

Key Question 2
What is the test performance of newer markers of iron status\(^a\) as a replacement for or an add-on to the older markers\(^b\) in stages 3-5 CKD patients nondialysis and dialysis, and in patients with a kidney transplant?
   a. What reference standards are used for the diagnosis of iron deficiency in studies evaluating test performance?
   b. What are the adverse effects or harms associated with testing using newer and/or older markers of iron status?

Key Question 3
In stages 3–5 CKD patients, nondialysis and dialysis, with iron deficiency, what is the impact of managing iron status based on newer laboratory biomarkers either alone or in addition to older laboratory biomarkers on intermediate outcomes (e.g., improvement in Hb levels, dose of ESA, time in target Hb range), compared with managing iron status based on older laboratory biomarkers alone?
   a. What are the adverse effects or harms associated with the treatments guided by tests of iron status?

Key Question 4
What factors affect the test performance and clinical utility of newer markers of iron status, either alone or in addition to older laboratory biomarkers, in stages 3–5 CKD patients (nondialysis and dialysis) with iron deficiency? For example:
   - Biological variation in diagnostic indices
   - Use of different diagnostic reference standards
   - Type of dialysis (i.e., peritoneal or hemodialysis)
   - Patient subgroups (i.e., age, sex, comorbid conditions, erythropoiesis-stimulating agent resistance, protein energy malnutrition secondary to an inflammatory state, hemoglobinopathies [e.g., thalassemia and sickle cell anemia])
   - Route of iron administration (i.e., oral or intravenous)
   - Treatment regimen (i.e., repletion or continuous treatment)
   - Interactions between treatments (i.e., patients treated with versus without ESA, patients treated with versus without iron-replacement therapy)
   - Other factors (based on additional information in the reviewed papers)

---
\(^a\)Content of hemoglobin [Hb] in reticulocytes, percentage of hypochromic red blood cells, erythrocyte zinc protoporphyrin, soluble transferrin receptor, hepcidin, and superconducting quantum interference devices.
\(^b\)Bone marrow iron stores, serum iron, transferrin saturation, iron-binding capacity, and ferritin.
Organization of This Report

The results chapter of this report is organized in the order of the Key Questions. The majority of the included studies were related to test performance (Key Question 2), and they addressed many different laboratory markers and reference standard pairs. Thus, we organized studies included in Key Question 2 alphabetically by newer laboratory markers of iron status.

A list of abbreviations and acronyms can be found at the end of the report, following the references.
Methods

The methods for this comparative effectiveness review (CER) adhere to those suggested by the Agency for Healthcare Research and Quality in its Methods Guide for Effectiveness and Comparative Effectiveness Reviews, hereafter referred to as the Methods Guide (available at www.effectivehealthcare.ahrq.gov/methodsguide.cfm). The main sections in this chapter reflect the elements of the protocol established for the CER; certain methods map to the PRISMA checklist. All methods were determined a priori. Any deviations from or modifications to the original protocol are described in this chapter.

AHRQ Task Order Officer

The AHRQ Task Order Officer (TOO) was responsible for overseeing all aspects of this project. The TOO facilitated a common understanding among all parties involved in the project, resolved ambiguities, and fielded all EPC queries regarding the scope and processes of the project. The TOO and other staff at AHRQ reviewed the report for consistency, clarity, and to ensure that it conforms to AHRQ standards.

Topic Refinement and Review Protocol

During a topic refinement phase, the initial questions that had previously been nominated for this report were refined with input from a panel of Key Informants. Key Informants included two representatives from the original nominating organization (American Association of Clinical Chemistry), two nephrologists, one hematologist, one renal dietician, one nurse manager, one public payer representative, and one private payer representative. After a public review of the proposed Key Questions, the clinical experts were reconvened to form the Technical Expert Panel (TEP), which served in an advisory capacity to help refine Key Questions, identify important issues, and define parameters for the review of evidence. Discussions among the EPC, TOO, Key Informants, and, subsequently, the TEP occurred during a series of teleconferences and via email. In addition, input from the TEP was sought during compilation of the report when questions arose concerning the scope of the review.

Literature Search Strategy

Search Strategy

We conducted literature searches of studies in MEDLINE® (from inception to May, 2012) and the Cochrane Central Register of Controlled Trials (through the first quarter of 2012). All studies published in any language with adult human subjects were screened to identify articles relevant to each Key Question. Our search strategy employed the National Library of Medicine’s Medical Subject Headings (MeSH) keyword nomenclature developed for MEDLINE. The full search strategy is described in Appendix A. The search strategy included MeSH or search terms for both newer and older laboratory biomarkers of interest, and MeSH or search terms for iron or erythropoietin treatment drugs and formulations. We combined these two groups of search strategies with MeSH or search terms for population and study designs of interest. We checked our search strategy against those used in relevant guidelines and systematic reviews. We also make sure our search covered key articles identified from the reference lists of key papers.
We did not search for unpublished studies, as such works and their data are not peer reviewed. However, we did search the Food and Drug Administration 510(k) database (www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmm/pmm.cfm) for all listed automated hematology analyzer with Product Code GKZ in July, 2012. We limited the search to products that received approval since 2008, and did not find relevant data in CKD patients. We also screened the reference lists of related guidelines and selected narrative reviews and primary articles for additional articles. Finally, we searched ClinicalTrials.gov for ongoing or completed studies using the search string “iron deficiency AND (dialysis OR kidney disease)”. When potential relevant studies were found, we conducted Internet and Pubmed® search for associated peer-reviewed publications.

Inclusion and Exclusion Criteria

The eligibility criteria for populations, interventions, comparators, outcomes, and study designs or settings (PICOS) are enumerated in Table 1. For all Key Questions, we excluded studies with fewer than 10 patients with CKD.

Table 1. Study eligibility criteria

<table>
<thead>
<tr>
<th>Key Question/PICO</th>
<th>Inclusion Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Key Question 1 (Overarching Question)</strong></td>
<td></td>
</tr>
</tbody>
</table>
| Populations | • Pediatric and adult nondialysis patients with stage 3, 4, or 5 CKD  
• Patients with CKD undergoing dialysis (hemo- or peritoneal dialysis)  
• Patients with a kidney transplant |
| Interventions | • Newer laboratory biomarkers to assess iron status and manage iron deficiency either as a replacement for or in addition to older laboratory biomarkers |
| Comparators | • Older laboratory biomarkers to assess iron status and manage iron deficiency |
| Outcomes | • Mortality  
• Morbidity (e.g., cardiac or liver toxicity and infection)  
• Quality of life, measured using standardized scales, including: Kidney Disease Quality of Life (KDQOL), Health Related Quality of Life (HRQOL), Medical Outcomes Study Short Form-36 (SF-36), and Pediatric Quality of Life Inventory (PQLI)  
• Adverse effects or harms associated with testing and associated treatments (e.g., test-related anxiety, adverse events secondary to venipuncture, effects of iron overload with iron treatments, and cardiovascular complications from use of erythropoietin at higher Hb levels) |
| Study designs | • Randomized controlled trials  
• Nonrandomized controlled trials  
• Observational studies with concurrent comparison groups |
| **Key Question 2, 3 and 4** |
| Populations | • Pediatric and adult nondialysis patients with stage 3, 4, or 5 CKD  
• Patients with CKD undergoing dialysis (hemo- or peritoneal dialysis)  
• Patients with a kidney transplant |
| Interventions | • Newer laboratory biomarker alone or in combination with older laboratory biomarkers of iron status |
| Comparators | • Older laboratory biomarkers of iron status, which include bone marrow iron stores, serum iron, transferrin saturation, iron-binding capacity, and ferritin |
Table 1. Study eligibility criteria (continued)

<table>
<thead>
<tr>
<th>Key Question/PICO</th>
<th>Inclusion Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Key Question 2, 3 and 4 (continued)</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Outcomes</strong></td>
<td>Key Question 2 and 4:</td>
</tr>
<tr>
<td></td>
<td>• Measures of test performance (e.g., concordance, sensitivity, specificity, predictive values, AUC) comparing newer with older markers of iron status. We accepted any “reference standard” used by the study authors for the analyses of sensitivity and specificity in the original study, including functional iron deficiency as defined by response or nonresponse to treatment</td>
</tr>
<tr>
<td></td>
<td>• Adverse effects or harms associated with laboratory testing</td>
</tr>
<tr>
<td></td>
<td>Key question 3 and 4:</td>
</tr>
<tr>
<td></td>
<td>• Intermediate outcomes</td>
</tr>
<tr>
<td></td>
<td>• Increase in Hb or hematocrit, or more consistent maintenance of Hb or hematocrit within the desired range</td>
</tr>
<tr>
<td></td>
<td>• Use of ESA for maintenance of Hb within the desired range (stable dose in contrast to escalating dose resulting in net decreased ESA dose in hypo-responsive patients or actual decreased ESA dose in relatively responsive patients)</td>
</tr>
<tr>
<td></td>
<td>• Adverse effects or harms associated with different management strategies</td>
</tr>
<tr>
<td><strong>Study designs</strong></td>
<td>Key Question 2:</td>
</tr>
<tr>
<td></td>
<td>• Any design</td>
</tr>
<tr>
<td></td>
<td>Key Question 3 and 4:</td>
</tr>
<tr>
<td></td>
<td>• Randomized controlled trials</td>
</tr>
<tr>
<td></td>
<td>• Nonrandomized controlled trials</td>
</tr>
<tr>
<td></td>
<td>• Observational studies with concurrent comparison groups</td>
</tr>
<tr>
<td><strong>Study settings</strong></td>
<td>• Any setting: primary or specialty care, in-facility or home, and inpatient or outpatient</td>
</tr>
</tbody>
</table>

AUC=area under the curve; CKD=chronic kidney disease; ESA=erythropoiesis stimulating agents; Hb=hemoglobin; HD=hemodialysis; IV=intravenous; RetHe=reticulocyte hemoglobin equivalent; sTfR=soluble transferrin receptor; TSAT=transferrin saturation; ZPP=erythrocyte zinc protoporphyrin.

*a* Hemoglobin (Hb) content in reticulocytes, percentage of hypochromic red blood cells, erythrocyte zinc protoporphyrin, soluble transferrin receptor, hepcidin, and superconducting quantum interference devices.

*b* Bone marrow iron stores, serum iron, transferrin saturation, iron-binding capacity, and ferritin.

**Study Selection**

We screened all abstracts available in English. Abstracts were screened based on eligibility criteria, with exclusions cross-checked by a second investigator. All studies that were accepted based on their abstracts were then reviewed in full. For those articles not available in English, we first employed Google Translate (translate.google.com) in attempting to determine their eligibility. If we had any question on the eligibility of non-English articles, we identified native language speakers to assist in full-text screening. It should be noted that most non-English articles in our literature search had English abstracts, and in many cases, non-English articles were excluded at the abstract screening level.

Full-text articles were evaluated independently by two investigators for eligibility. Disagreement on an article’s eligibility was resolved by consensus. A list of excluded articles and the reasons for excluding these articles are tabulated in Appendix B.

**Data Extraction**

Each study was extracted by one investigator, and reviewed and confirmed by at least one other investigator. Any disagreements were resolved by discussion amongst the team members.
Data were extracted into standard forms. The basic elements and design of these forms were similar to those we have used for other comparative effectiveness reviews, such as queries capturing population characteristics, sample size, study design, descriptions of the test and reference standard, analytic details, and outcomes. Prior to extraction, the form was customized to capture all elements relevant to the Key Questions. We used separate forms for questions related to test performance (Key Question 2) and the effectiveness of test-oriented treatments (Key Question 3). We tested the forms on several studies and revised as necessary prior to data extraction of all articles. A blank extraction form is provided in Appendix C.

**Risk of Bias—Assessment of Individual Studies**

We assessed the risk of biases (methodological quality) for each individual study using the assessment instrument described in the AHRQ Methods Guide. Briefly, we rated each study as being of high, medium, or low risk of bias on the basis of adherence (Yes, No, or Unclear/Not reported) to generally accepted standard methodologies (Quality Assessment of Diagnostic Accuracy Studies [QUADAS] tool for studies of diagnostic performance and the Cochrane risk of bias tool for intervention studies), and assessed and reported each methodological quality item for all qualifying studies. We also considered the clarity and consistency in reporting as part of the overall judgment of risk of bias. Grading was outcome-specific, such that a given study that reported its primary outcome well but conducted an incomplete analysis of a secondary outcome would be graded as having different quality for the two outcomes. Studies of different study designs were graded within the context of their study design; RCTs and observational studies were graded separately to be at a high, medium, or low risk of bias. Only RCTs and prospective cohort studies could be rated as having a low risk of bias.

**Data Synthesis**

We summarized all included studies in narrative form as well as in summary tables (see below) that condense the important features of the study populations, design, anemia and iron status indices, laboratory tests, reference standards, background treatment, intervention, outcomes, and results. Where appropriate we summarized the characteristics of eligible studies using summary statistics (means, medians, ranges and standard deviations).

The synthesis of data for Key Question 2 was complicated by the fact that there is a lack of generally accepted reference standard tests for determining iron deficiency in the context of CKD. Thus, we accepted any “reference standard” used by the authors of the included primary studies for the analyses of test performance of newer or classical laboratory biomarkers of iron status. Based on our post-hoc observation of this body of literature, we separated the included studies into two distinct groups. Specifically, current studies use two distinct methods to operationalize a reference standard for assessing test performance: (1) a response to intravenous (IV) iron treatment, often referred to as “functional iron deficiency”; and (2) classical laboratory biomarkers, alone or in combination with each other, often referred to as “absolute iron deficiency”.

When a study used a response to IV iron treatment as the reference standard for iron deficiency, it allowed us to directly compare the test performance of classical with newer biomarkers in predicting a response. To facilitate the interpretation of study results, the reported sensitivity and specificity of both newer and classical laboratory biomarkers were visually depicted in receiver operating characteristic (ROC) space. We did not conduct meta-analyses because there was a high degree of heterogeneity across studies in the definitions of reference
standard (a response to IV iron treatment), baseline iron status of the study populations, and background treatment.

When a study used classical laboratory biomarkers (alone or in combination with each other) as the reference standard for iron deficiency, we were prevented from comparing the test performance of classical with newer biomarkers. For the purpose of our review, this approach was analogous to assessing the concordance between classical and newer biomarkers of iron status. Since concordance cannot tell us which test is better and which is worse—both may be equally bad or equally good for defining “iron deficiency”—and cannot answer Key Question 2, these studies were only included for subquestion 2a (What reference standards are used for the diagnosis of iron deficiency in studies evaluating test performance?).

Summary Tables

Summary tables succinctly report measures of the main outcomes evaluated, and additional information to assist their interpretation. We used separate summary tables for questions related to test performance (Key Question 2) and the effectiveness of test-oriented treatments (Key Question 3). For Key Question 2, we included information regarding study population, laboratory analysis or assay, index test cutoff, reference standard, percentage of patients with iron deficiency, test performance outcomes (e.g., sensitivity, specificity, and area under the ROC curve [AUC]), and risk of bias. For Key Question 3, we included additional information regarding iron treatment regimen, anemia management protocol targets, followup duration, the mean outcome values, their 95 percent confidence intervals (CI), standard deviations (SD) or other measures of variability and when available, the mean difference (between groups) and its corresponding P value, or CI, as appropriate.

Graphical Presentation of Study Results

To facilitate the interpretation of study results, the reported sensitivity and specificity of both newer and classical laboratory biomarkers of iron status were visually depicted in ROC space. Each point on the ROC curve represents a sensitivity/specificity pair corresponding to a particular decision threshold. A test with perfect discrimination (no overlap in the two distributions) has a ROC curve that passes through the upper left corner (100% sensitivity, 100% specificity). Therefore, the closer the ROC curve is to the upper left corner, the higher the overall accuracy of the test.

When applicable, a published ROC curve that showed individual data points for multiple cutoffs on the curve was digitized using Engauge Digitizer, an open source digitizing software package (digitizer.sourceforge.net/). The digitization was accomplished by obtaining the image file of the published graph or plot, recording locations of data points and axes, and using the software to convert the data points on the graph into estimated data values. The digitized data were then exported into Stata® (a data analysis and statistical software suite) to recreate the ROC curve.

Test Performance Terms and Definitions

There are many quantitative indicators of test performance. Below, we list the test performance terms and definitions used in the current report:

- **Receiver operating characteristic curve**: ROC curves compare sensitivity with specificity across a range of values for the ability to predict a dichotomous outcome.
(defined as the reference standard). The ROC curve graphically displays the trade-off between sensitivity and specificity, and is useful in assigning the best cut-offs for clinical use.

- **Overall test accuracy**: Overall accuracy of a test is expressed as area under the ROC curve (AUC). The AUC provides another useful parameter for comparing test performance between, for example, classical and newer laboratory biomarkers of iron status. The AUC summarizes the ROC curve in a single number but loses information about the tradeoffs between sensitivity and specificity.

- **Test accuracy**: Test accuracy refers to sensitivity (true positive rate) and specificity (true negative rate) of a test. For any test, there is usually a trade-off between sensitivity and specificity. For example, a test may exhibit a high sensitivity and a low specificity, or vice versa.

- **Diagnostic odds ratio (DOR)**: The DOR is a single indicator of test performance that combines the strengths of sensitivity and specificity. The DOR offers advantages when logistic regression is used with diagnostic problems, because the DOR equals the regression coefficient, after exponentiation. DORs are conditional: They depend on the other variables that have been used in the model. Consequently, the conditional DOR of each test variable, adjusted for the other variable (e.g., inflammation markers), can be estimated.

**Strength of the Body of Evidence**

We followed the Methods Guide in evaluating the strength of the body of evidence for each Key Question with respect to four domains: risk of bias, consistency, directness, and precision. Briefly, we defined the risk of bias—low, medium, or high—on the basis of design and methodological quality of the underlying studies.

We rated the consistency of the data as: no inconsistency, inconsistency present, or not applicable if there was only one study available. We assessed the direction, magnitude, and statistical significance of all studies to make a determination. We described our logic where studies were not unanimous. For Key Question 2, we judged consistency based on the studies’ location in the ROC space as a measure of consistency.

We assessed the precision of the evidence (assessed as precise or imprecise) on the basis of the degree of certainty surrounding an effect estimate. A precise estimate was an estimate that would allow a clinically useful conclusion. An imprecise estimate was one for which the confidence interval was wide enough to include clinically distinct conclusions (e.g., both clinically important benefits and harms—a situation in which the direction of effect is unknown), a circumstance that would preclude a conclusion. For Key Question 2, we judged precision based on the distance of the study’s positive and negative LR scores from our pre-determined LR cutoffs.

We assess the directness based on the types of outcomes. We considered studies provided patient-center outcomes as the direct evidence to address our key questions. Finally, we rated the body of evidence based on a four-level scale—high, moderate, low, and insufficient—on the basis if our level of confidence that the evidence reflected the true effect for the major comparisons of interest. The rating of the strength of the body of evidence was based on the consensus of all team investigators.
Applicability

We followed the Methods Guide in evaluating the applicability of included studies to each patient population of interest,32 that is, nondialysis patients with stage 3, 4, or 5 CKD, patients with CKD undergoing hemo- or peritoneal dialysis, and patients with a kidney transplant. We evaluated and summarized studies of pediatric, adult, and elderly adults separately.

Peer Review and Public Commentary

The initial draft report was prereviewed by the TOO and an AHRQ Associate Editor (a senior member of a sister EPC). Following revisions, the draft report was sent to invited peer reviewers and was simultaneously uploaded to the AHRQ Web site where it was available for public comment for 30 days. All reviewer comments (both invited and from the public) were collated and individually addressed. The revised report and the EPC’s responses to invited and public reviewers’ comments were again reviewed by the TOO and Associate Editor prior to completion of the report. The authors of the report had final discretion as to how the report was revised based on the reviewer comments, with oversight by the TOO and Associate Editor.
Results

Introduction

In this Chapter, the results of literature searches come first, followed by the descriptions of all included studies and the overall strength of evidence table. The results of our syntheses were presented in the order of the Key Questions, from Key Question 1 to 4. Within each Key Question, we first summarize the key points of the findings and then present a more detailed synthesis of the literature. Please refer to Chapter 2. Methods for the methods used to synthesize the literature.

The majority of the included studies were related to test performance (Key Question 2), and they addressed many different laboratory markers and reference standard pairs. Thus, we organized studies included in Key Question 2 alphabetically by newer laboratory markers of iron status.

A list of abbreviations and acronyms can be found at the end of the report, following the references.

Literature Searches

The literature search yielded 6407 citations. From these, 694 articles were retrieved for full-text screening on the basis of abstracts and titles. Full-text articles were screened on the basis of study eligibility criteria; thirty articles were judged to have met the inclusion criteria. Figure 3 summarizes the study selection flow. A total of 664 articles were rejected on double, independent full-text screening because they did not meet one or more of the PICO criteria for a particular Key Question (see Appendix B for the list of rejected articles and the reasons for their rejection). The two most common reasons for rejection were: a) no diagnostic outcomes reported (studies reported only correlations between markers or the measurements of levels of markers before and after treatment); b) no comparative data for the outcomes of management strategies where treatment decisions were guided by test results (newer versus classical markers). Finally, a total of 30 articles were accepted,41-70 including one Polish and one Japanese language publication.

Description of Included Studies

Thirty articles were included. Twenty seven articles reported data on the test performance of newer markers of iron status compared with classical markers (Key Question 2),41-67 two reported the intermediate outcomes comparing the iron management guided by the newer laboratory markers with that guided by the classical markers (Key Question 3),66,70 and three (in two articles) reported data on the factors that affected the test performance comparing newer with classical laboratory markers of iron status (Key Question 4).68,69 Most studies enrolled only adult CKD patients undergoing hemodialysis. Eighteen studies did not reported information regarding their funding sources. Four studies were funded by the industry.41,64,68,70 Eight studies received funding from nonprofit sources, such as national kidney training fellowships,43,59 internal university hospital grant,52 academic foundation grant,56 or government funding.42,47,62,66 Detailed characteristics of included studies are presented later with results for each Key Question.
Key Question 1. Comparative Effectiveness of Newer Versus Older Markers of Iron Status for the Diagnosis and Management of Iron Deficiency Anemia

No study reported on patient centered outcomes (mortality, morbidity, quality of life, and adverse effects) when using newer laboratory markers as a replacement for or an add-on to the classical laboratory markers for assessing iron status and management of iron deficiency in stages 3–5 CKD nondialysis and dialysis patients, and in patients with a kidney transplant.

This question of overall impact on patient centered outcomes was broken out into three component Key Questions (Key Questions 2 to 4). Combining evidence gather to address these three component Key Questions can thus inform the conclusions for this reviews primary, overarching question.
Key Question 2. Test Performance of Newer Markers Compared With the Older Markers of Iron Status

2a. Reference Standards for the Diagnosis of Iron Deficiency in Studies Evaluating Test Performance

A total of 27 studies were included for Key Question 2. Current studies use two distinct methods to operationalize a reference standard for assessing test performance: (1) a response to intravenous (IV) iron treatment; and (2) classical laboratory biomarkers, alone or in combination with each other. However, across studies, there are large variations in the definitions of these reference standards.

Of the 27 included studies, 15 used classical markers of iron status to define “iron deficiency” as the reference standard in calculating the test accuracy (sensitivity and specificity) of newer markers of iron status.41-43,45,48,49,51-57,60,63,66 These studies used the following definitions: (1) TSAT ≤ 15%;48 (2) TSAT ≤ 20%;41-43,53,57,63,66 (3) ferritin ≤100 ng/mL;43 (4) TSAT ≤20 percent and ferritin ≤100 ng/mL;49,51,53-55,63 (5) TSAT ≤20% or ferritin ≤100 ng/mL;51,56,60,66 (6) serum iron < 40 µg/dL, TSAT<20%, ferritin <100 ng/mL, and Hb <11 g/dL;45 (7) TSAT<20 percent, ferritin 100-800 ng/mL, and Hb <11 g/dL;45 and (8) TSAT <16 percent and ferritin <12 ng/mL.54 Many of these studies evaluated more than one newer marker at different test cutoffs, including content of hemoglobin in reticulocytes (CHr), percent hypochromic red blood cells (%HYPO), reticulocyte hemoglobin content (RetHe), soluble transferrin receptor (sTfR), and erythrocyte zinc protoporphyrin (ZPP). As described in Methods, results from these 15 studies are analogous to assessing the concordance between classical and newer biomarkers of iron status. Since concordance between the tests cannot tell us which test is better and which is worse—both may be equally bad or equally good for defining “iron deficiency”—and cannot answer Key Question 2, the results of these 15 studies are only described in Appendix D.

Of the 27 included studies, 12 studies investigated the test accuracy of newer or classical markers of iron status, using a response to IV iron treatment as the reference standard for diagnosis of iron deficiency.44,46,47,50,52,58,59,61,62,64,65,67 However, there exists a high degree of heterogeneity in the reference standards used across studies as well (details are described later in Table 2). The most commonly used definition for a response to IV iron treatment was an increase in hemoglobin (Hb) concentration ≥1 g/dL after a (variable) period of IV iron treatment.44,46,47,50,52,58,59,61,62,64,65,67 Other reference standards include a ≥ 15 percent increase in Hb,61 an increase in Hct of ≥3 percent and/or a ≥ 30 percent reduction in erythropoietin (EPO) dose,47 >1 point increase in corrected reticulocyte index,52 and 5 percent increase in Hct or a decrease in EPO dose of >2000 units per treatment.65 It should be noted that there was no uniform regimen of IV iron in terms of dosage and iron formulation across these studies. There was also a wide range of durations of IV iron treatment across studies. The potential impact of IV iron treatment regimen on the test performance of newer or classical laboratory markers of iron status is not known.

As described in Methods, these 12 studies, which used a response to IV iron treatment as the reference standard for iron deficiency, allowed us to directly compare the test performance of classical with newer biomarkers in predicting a response. Thus, the results from these studies were synthesized to answer Key Question 2.
Comparisons of Test Performance of Newer Versus Classical Markers of Iron Status To Predict a Response to Intravenous Iron Treatment

In this section, we summarize the findings from 12 studies (10 prospective cohorts, one retrospective cohort, and one cohort study of unclear directionality) evaluating the test performance of newer or classical laboratory markers of iron status, using a response to IV iron treatment as the reference standard for diagnosis of iron deficiency. Of these 12 studies, eight reported comparative data between five of the newer markers (no studies addressed SQUIDD) and the classical markers (although not all studies performed formal statistical testing for the comparisons). Seven of these eight enrolled adult hemodialysis (HD CKD) patients, and one study enrolled adult nondialysis (ND CKD) patients. The remaining four studies investigated the test performance of newer laboratory markers alone. Of these four, three enrolled adult HD CKD patients, and one enrolled adult peritoneal dialysis (PD CKD) patients. None of the reviewed studies enrolled pediatric CKD patients, and we did not include studies evaluating the test performance of classical markers alone.

Table 2 tabulates the newer or classical markers of iron status that were investigated in each study. In summary, content of hemoglobin in reticulocytes (CHr) was investigated in 10 studies, percent hypochromic red blood cells (%HYPO) in six studies, soluble transferrin receptor (sTfR) and erythrocyte zinc protoporphyrin (ZPP) in two studies each, and hepcidin and reticulocyte hemoglobin content (RetHe) in one study each. Five studies investigated more than one newer marker. Both transferrin saturation (TSAT) and ferritin were investigated in the seven studies that reported comparative data between newer and classical markers. The most commonly used definition for a response to IV iron treatment was an increase in hemoglobin (Hb) concentration ≥1 g/dL after a period of IV iron treatment (Table 2). However, there was no uniform regimen of IV iron in terms of dosage and iron formulation. There was also a wide range of durations of IV iron treatment across studies. The potential impacts of IV iron treatment regimen on the test performance of newer or classical laboratory markers of iron status are not known. Additionally, there was a high degree of heterogeneity in definitions for the reference standard (a response to IV iron treatment) and background treatment across studies (Table 3). This heterogeneity prevented us from performing meta-analyses and limits our confidence in the validity of evaluating the consistency of findings across studies.

Interpretations of the summarized results for the overall test accuracy (measured by area under the ROC curve) or sensitivity and specificity (at specified cutoff values) comparing newer with classical markers of iron status to predict iron deficiency (as defined by a response to IV iron treatment) in adult HD CKD patients are described in Table 4. To facilitate indirect comparisons across studies through visual inspections, the test accuracy of the newer or classical markers of iron status for diagnosing iron deficiency among adult HD CKD patients were plotted in a receiver operating characteristics (ROC) space (Figures 4 and 5). Individual markers of iron status were plotted in a separate panel of Figure 4 and 5. Data in this figure were extracted from the seven studies that reported comparative data between newer and classical markers, and three additional studies that investigated the test performance of newer laboratory markers alone. The results from each of the single studies examining adult ND CKD patients and adult PD CKD patients were not plotted in the ROC space.
Summary of Key Points (Tables 2 to 4; Figures 4 and 5)

- Among adult HD CKD patients, there is a low level of evidence that:
  - CHr has a similar or better overall test accuracy compared with classical markers (TSAT or ferritin) to predict a response to IV iron treatment. Data suggest that CHr (with cutoff values of <27 or <28 pg) has a better sensitivity and specificity to predict iron deficiency than classical markers (TSAT <20 or ferritin <100 ng/mL).
  - %HYPO has similar or better overall test accuracy compared with TSAT, and better overall test accuracy compared with ferritin, to predict a response to IV iron treatment. Data suggest that %HYPO (with cutoff values of >6% or >10%) has a better sensitivity and specificity to predict iron deficiency (as defined by a response to IV iron treatment) than classical markers (TSAT <20% or ferritin <100 ng/mL).
  - sTfR has a similar test performance compared with classical markers (TSAT or ferritin) to predict a response to IV iron treatment.
  - There exists a high degree of heterogeneity across studies in the background treatment and the definitions of the reference standard (a response to IV iron treatment), limiting our ability in evaluating the consistency of findings.

- There is insufficient evidence regarding:
  - Test performance of newer markers of iron status as an add-on to older markers.
  - Test performance comparing erythrocyte ZPP, RetHe, and hepcidin to predict a response to IV iron treatment in adult HD CKD patients.
  - Test performance comparing newer (CHr, %HYPO, RetHe, sTfR, ZPP, and hepcidin) with classical laboratory markers to predict a response to IV iron treatment in adult PD and ND CKD patients, and in pediatric CKD patients.
<table>
<thead>
<tr>
<th>Study, Year [UI]</th>
<th>Population</th>
<th>Total N Enrolled</th>
<th>IV Iron Treatment</th>
<th>Reference Standard (Response to IV Iron Therapy)</th>
<th>Ferritin</th>
<th>TSAT</th>
<th>CHr</th>
<th>RetHe</th>
<th>%HYPO</th>
<th>ZPP</th>
<th>sTfR</th>
<th>Hepcidin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bovy, 2007&lt;sup&gt;44&lt;/sup&gt; [17237481]</td>
<td>HD CKD</td>
<td>32</td>
<td>IV iron sucrose (1200 mg total)—100 mg at the end of dialysis session over 4 wks</td>
<td>≥1 g/dL increase in Hb during the 4-week IV iron Tx</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Buttarello, 2010&lt;sup&gt;46&lt;/sup&gt; [20472854]</td>
<td>HD CKD</td>
<td>69</td>
<td>IV iron gluconate and α-darbepoetin to maintain Hb between 11.0 &amp; 12.0 g/dL</td>
<td>≥1 g/dL increase in Hb at any time after the third wk of IV iron Tx</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Fishbane, 1995&lt;sup&gt;65&lt;/sup&gt; [7872320]</td>
<td>HD CKD</td>
<td>62</td>
<td>1,000 mg IV iron dextran in 100 mg doses over 10 sequential HD Tx</td>
<td>5% increase in Hct or a decrease in EPO dose of &gt;2000 units/treatment over 3-6 mths</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Mitsuiki, 2003&lt;sup&gt;58&lt;/sup&gt; [14586744]</td>
<td>HD CKD</td>
<td>27</td>
<td>40 mg of chondroitin sulfate-iron colloid IV once a wk after the regular dialysis session</td>
<td>Change in Hct ≥3% (or change in Hb ≥1 g/dL) within 8 wks after IV iron Tx</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Mittman, 1997&lt;sup&gt;29&lt;/sup&gt; [9398141] U.S.</td>
<td>HD CKD</td>
<td>79</td>
<td>Single bolus of 500 mg IV iron dextran over 2 hours during a regular hemodialysis session</td>
<td>&gt;1 point increase in corrected reticulocyte index at any point during the 2 wks after IV iron Tx</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>
Table 2. An evidence map of studies of newer or classical markers of iron status in predicting a response to intravenous iron treatment in adult CKD patients (continued)

<table>
<thead>
<tr>
<th>Study, Year [UI]</th>
<th>Population</th>
<th>Total N_{enrolled}</th>
<th>IV Iron Treatment</th>
<th>Reference Standard (Response to IV Iron Therapy)</th>
<th>Ferritin</th>
<th>TSAT</th>
<th>CHr</th>
<th>RetHe</th>
<th>%HYPO</th>
<th>ZPP</th>
<th>sTfr</th>
<th>Hepcidin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tessitore, 2001 [11427634]</td>
<td>HD CKD</td>
<td>125</td>
<td>IV sodium ferric gluconate complex in sucrose as a slow (2 min) IV bolus at end of dialysis with 31 or 62 mg iron as per predialysis serum transferrin (&lt; or &gt; 170 mg/dL, respectively)</td>
<td>≥15% increase in Hb at any 2 consecutive measurements (evaluated every 2 wks)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Tessitore, 2010 [20538788]</td>
<td>HD CKD</td>
<td>56</td>
<td>1 g intravenous iron (62.5 mg ferric gluconate at 16 consecutive dialysis sessions)</td>
<td>≥1 g/dL increase in Hb after 6 wks IV iron treatment</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Van Wyck, 2005 [16316362]</td>
<td>ND CKD</td>
<td>95</td>
<td>IV iron sucrose 1,000 mg in divided doses over 14 days, as either 500 mg IV infusions on study days 0 and 14 or 200 mg injections on five different days from day 0 to day 14.</td>
<td>≥ 1 g/dL increase in Hb after 8 wks IV iron Tx</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
Table 2. An evidence map of studies of newer or classical markers of iron status in predicting a response to intravenous iron treatment in adult CKD patients (continued)

<table>
<thead>
<tr>
<th>Study, Year [UI]</th>
<th>Population</th>
<th>Total N enrolled</th>
<th>IV Iron Treatment</th>
<th>Reference Standard (Response to IV Iron Therapy)</th>
<th>Ferritin</th>
<th>TSAT</th>
<th>CHr</th>
<th>RetHe</th>
<th>%HYPO</th>
<th>ZPP</th>
<th>sTfR</th>
<th>Hepcidin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chuang, 200347 [12543894]</td>
<td>HD CKD</td>
<td>95</td>
<td>IV iron saccharate 100 mg at end of each dialysis session, three times a week for 4 wks, then 100 mg every 2 wks for 5 mths</td>
<td>Rise in Hct of ≥3% or a reduction in rHuEpo dose of ≥30% over the baseline values at the end of the study</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fishbane, 199752 [9211366]</td>
<td>HD CKD</td>
<td>50</td>
<td>1,000 mg of IV iron dextran infused over two hours as a single-dose infusion</td>
<td>1 point increase in the corrected reticulocyte index within two wks of IV iron Tx</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Silva, 199867 [9794562]</td>
<td>HD CKD</td>
<td>33</td>
<td>IV iron saccharate 20 mg diluted in 10 mL saline, and given in last 10 minutes of dialysis</td>
<td>≥ 1 g/dL increase in Hb during the 6 mths of IV iron Tx</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Domrongkitchaiporn, 199950 [10401012]</td>
<td>PD CKD</td>
<td>23</td>
<td>IV iron—1000 mg ferric saccharate-infused over 2 hours in two divided doses 1 wk apart</td>
<td>Sustained &gt;1 g/dL increase in Hb within 3 mths of IV iron Tx</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

%HYPO=percent of hypochronic red blood cell; CHr=content of hemoglobin in reticulocytes; CKD=chronic kidney disease; ESRD=end stage renal disease; Hb=hemoglobin; Hct=hematocrit; HD=hemodialysis; IV=intravenous; mths=months; ND=nondialysis; PD=peritoneal dialysis; RetHe=reticulocyte hemoglobin equivalent; rHuEpo=recombinant human erythropoietin; sTfR=soluble transferrin receptor; TSAT=transferrin saturation; Tx=treatment; UI=universal identifier/Pubmed ID; wk=week; ZPP=erythrocyte zinc protoporphyrin
√= marker was investigated
√*= best predictors of iron deficiency among all other markers
## Table 3. Characteristics of studies evaluating the ability of newer or classical markers of iron status to predict the response to IV iron treatment

<table>
<thead>
<tr>
<th>Study, Year [UI] Country</th>
<th>Study Design Recruitment Method</th>
<th>Sampling Population</th>
<th>Enrolled / Analyzed</th>
<th>Demographics</th>
<th>Anemia and Iron Status Indices</th>
<th>Background Treatment</th>
<th>Risk of Bias</th>
</tr>
</thead>
</table>
| Bovy, 2007[44] Belgium   | Prospective cohort Selected sample | HD CKD              | 32/32              | Male (%): 59  
Age (yr): 65  
Race (%): NR | Hb (g/dL): 12.3  
Hct (%): 38.8  
ferritin (ng/mL): 347  
TSAT (%): 21 | ESA dose: 153.5 IU/kg/wk  
Iron washout: 4 wks | Medium |
| Buttarello, 2010[46] Italy | Prospective cohort Selected sample | HD CKD              | 69/59              | Male (%):NR  
Age (yr): NR  
Race (%): NR | Hb (g/dL): 11.0  
Hct (%): NR  
ferritin (ng/mL): 238  
TSAT (%): 18 | ESA dose: NR  
Iron washout: 3 wks | Medium |
| Fishbane, 1995[56] U.S. | Prospective cohort | HD CKD              | 62/62              | Male (%): 47  
Age (yr): 52  
Race (%): NR | Hb (g/dL): NR  
Hct (%): NR  
ferritin (ng/mL): NR  
TSAT (%): NR | ESA dose: NR  
Iron washout: No washout, though subjects with transfusions within 3 months were excluded | High |
| Mitsuiki, 2003[58] Japan | Retrospective cohort Selected sample | HD CKD              | 27/27              | Male (%): 30  
Age (yr): 59  
Race (%): NR | Hb (g/dL): NR  
Hct (%): 26.8  
ferritin (ng/mL): 83.6  
TSAT (%): 27.7 | ESA dose: 4139 IU/wk  
Iron washout: 12 wks | Medium |
| Mittman, 1997[59] U.S.  | Prospective cohort | HD CKD              | 79/79              | Male (%): 50  
Age (yr): 63  
Race (%): Black-75 | Hb (g/dL): NR  
Hct (%): 34.1  
ferritin (ng/mL): 155.5  
TSAT (%): 24.5 | ESA dose: NR  
Iron washout: 4 wks | Medium |
| Tessitore, 2001[61] Italy | Cohort (prospective or retrospective NR) Selected sample | HD CKD              | 125/125            | Male (%): 80  
Age (yr): 31 to 84  
Race (%): NR | Hb (g/dL): 9.9  
Hct (%): NR  
ferritin (ng/mL): 201  
TSAT (%): 22 | ESA dose: 7216 IU/wk  
Iron washout: 3 wks | High |
Table 3. Characteristics of studies evaluating the ability of newer or classical markers of iron status to predict the response to IV iron treatment (continued)

<table>
<thead>
<tr>
<th>Study, Year [UI] Country</th>
<th>Study Design Recruitment Method</th>
<th>Sampling Population</th>
<th>N Enrolled / N Analyzed</th>
<th>Demographics</th>
<th>Anemia and Iron Status Indices</th>
<th>Background Treatment</th>
<th>Risk of Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tessitore, 2010[62] Italy</td>
<td>Prospective cohort Selected sample</td>
<td>HD CKD</td>
<td>56/56</td>
<td>Male (%): 57 Age (yr): 67 Race (%): NR</td>
<td>Hb (g/dL): 11.6 Hct (%): NR ferritin (ng/mL):146 TSAT (%): 20</td>
<td>ESA dose: 8000 IU/wk Iron washout: 10 wks</td>
<td>Low</td>
</tr>
<tr>
<td>Chuang, 2003[47] Taiwan</td>
<td>Prospective Cohort Selected sample</td>
<td>HD CKD</td>
<td>95/65</td>
<td>Male (%): 51 Age (yr): 60 Race (%): NR</td>
<td>Hb (g/dL): 9.8 Hct (%): 30.1 ferritin (ng/mL): 244 TSAT (%): 38.5</td>
<td>ESA dose: 90 IU/wk/kg Iron washout: 12 wks</td>
<td>High</td>
</tr>
<tr>
<td>Fishbane, 1997[52] U.S.</td>
<td>Prospective cohort Random sampling</td>
<td>HD CKD</td>
<td>50/32</td>
<td>Male (%): NR Age (yr): NR Race (%): NR</td>
<td>Hb (g/dL): NR Hct (%): 32.7 ferritin (ng/mL): 231 TSAT (%): NR</td>
<td>ESA dose: NR Iron washout: 4 wks</td>
<td>High</td>
</tr>
<tr>
<td>Silva, 1998[67] Portugal</td>
<td>Prospective cohort Selected sample</td>
<td>HD CKD</td>
<td>33/33</td>
<td>Male (%): 61 Age (yr): 58 Race (%): NR</td>
<td>Hb (g/dL):10.8 Hct (%): NR ferritin (ng/mL):137 TSAT (%): 27</td>
<td>ESA dose: 118.2 IU/kg/wk Iron washout: NR (61% patients received oral iron)</td>
<td>High</td>
</tr>
<tr>
<td>Van Wyck, 2005[54] U.S.</td>
<td>Prospective cohort</td>
<td>ND CKD (stage 3-5)</td>
<td>95/79</td>
<td>Male (%): 33 Age (yr): 62 Race (%): Caucasian-56 Black-38 Other-6</td>
<td>Hb (g/dL): 10.2 Hct (%): NR ferritin (ng/mL): 92.6 TSAT (%): 16.4</td>
<td>ESA dose: NR Iron washout: 24 wks</td>
<td>Medium</td>
</tr>
<tr>
<td>Domrongkitchaiporn, 1999[66] Thailand</td>
<td>Prospective Cohort Selected sample</td>
<td>PD CKD</td>
<td>23/21</td>
<td>Male (%): 67 Age (yr): 51 Race (%): NR</td>
<td>Hb (g/dL): 8.4 Hct (%): NR ferritin (ng/mL): 643 TSAT (%): 33.9</td>
<td>ESA dose: 71 IU/wk/kg Iron washout: 4 wks</td>
<td>Medium</td>
</tr>
</tbody>
</table>

CKD=chronic kidney disease; ESA=erythropoiesis stimulating agents; ESRD=end stage renal disease; Hb=hemoglobin; Hct=hematocrit; HD=hemodialysis; Hr=content of hemoglobin in reticulocytes; IV=intravenous; IU=international units; ND=nondialysis; NR=not reported; PD=peritoneal dialysis; TSAT=transferrin saturation; UI=universal identifier/Pubmed ID; wk=week; yr=year
Table 4. Interpretations of the summarized results for the direct comparisons of the overall test accuracy or sensitivity and specificity (at specified cutoff values) of newer versus classical markers of iron status (at baseline) to predict a response to intravenous iron treatment\(^a\) in seven cohort studies among adult HD CKD patients

<table>
<thead>
<tr>
<th>Iron Status Marker</th>
<th>Total Number of Studies (Total N) [Risk of Bias]</th>
<th>Overall Test Accuracy When Compared With TSAT</th>
<th>Sensitivity and Specificity When Compared With TSAT &lt;20%</th>
<th>Overall Test Accuracy When Compared With Ferritin &lt;100 ng/mL</th>
<th>Sensitivity and Specificity When Compared With Ferritin &lt;100 ng/mL</th>
<th>Sensitivity and Specificity When Compared With TSAT &lt;20% or Ferritin &lt;100 ng/mL</th>
<th>Other Comparative Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHr /RetHe</td>
<td>6 CHr studies(^{44,46,58,59,61,62}) [1 low,(^62) 4 medium,(^{44,46,58,59}) 1 high risk(^{61})]</td>
<td>NS difference (4 studies)(^{44,61,62}) CHr &lt;30 or &lt;29 pg worse (1 study)(^{44}) CHr &lt;27 or &lt;28 pg better (1 study)(^{59})</td>
<td>NS difference (2 studies)(^{44,62}) CHr better (3 studies)(^{46,60,61})</td>
<td>CHr &lt;29 pg worse (1 study)(^{44}) CHr &lt;30 pg better (1 study)(^{44}) CHr &lt;27 or &lt;28 pg better (1 study)(^{59})</td>
<td>CHr &lt;30 or &lt;29 pg worse (1 study)(^{44}) CHr &lt;27 or &lt;28 pg better (1 study)(^{59})</td>
<td>Combination of %HYPO &gt;6% with CHr ≤29 pg produced minor improvement in sensitivity and specificity (1 study)(^{61})</td>
<td></td>
</tr>
<tr>
<td>RetHe</td>
<td>1 RetHe study(^{46}) [1 medium risk(^{46})]</td>
<td>RetHe better (1 study)(^{46})</td>
<td>NS difference (1 study)(^46) %HYPO &gt;10% better (1 study)(^{44}) %HYPO &gt;6% better (1 study)(^{61})</td>
<td>NS difference (1 study)(^{44}) %HYPO &gt;10% better (1 study)(^{44}) %HYPO &gt;6% better (1 study)(^{61})</td>
<td>%HYPO &gt;10% better (1 study)(^{44}) %HYPO &gt;6% better (1 study)(^{61})</td>
<td>Combination of % HYPO &gt;6 with TSAT≤20% produced a substantial increase in sensitivity but reduce in specificity (1 study)(^{61})</td>
<td></td>
</tr>
<tr>
<td>%HYPO</td>
<td>4 studies(^{44,44,61,62}) [1 low,(^{62}) 2 medium,(^{44,46}) 1 high risk(^{61})]</td>
<td>NS difference (1 study)(^{44}) %HYPO &gt;10% better (1 study)(^{44}) %HYPO &gt;6% better (1 study)(^{61})</td>
<td>NS difference (1 study)(^{44}) %HYPO &gt;10% better (1 study)(^{44}) %HYPO &gt;6% better (1 study)(^{61})</td>
<td>NS difference (1 study)(^{44}) %HYPO &gt;10% better (1 study)(^{44}) %HYPO &gt;6% better (1 study)(^{61})</td>
<td>%HYPO &gt;10% better (1 study)(^{44}) %HYPO &gt;6% better (1 study)(^{61})</td>
<td>%HYPO was the only significant predictor of a response to IV iron treatment among all other markers(^b) (2 study)(^{61,62}) Combination of % HYPO &gt;6 with TSAT≤20% produced a substantial increase in sensitivity but reduce in specificity (1 study)(^{61})</td>
<td></td>
</tr>
<tr>
<td>ZPP</td>
<td>2 studies(^{61,65}) [187] [2 high risk(^{61,65})]</td>
<td>NS difference (1 study)(^{61}) ZPP &gt;90 µmol/mol better (1 study)(^{65}) ZPP &gt;52 µmol/mol better (1 study)(^{65})</td>
<td>NS difference (1 study)(^{61}) ZPP &gt;90 µmol/mol better (1 study)(^{65}) ZPP &gt;52 µmol/mol better (1 study)(^{65})</td>
<td>NS difference (1 study)(^{61}) ZPP &gt;90 µmol/mol better (1 study)(^{65}) ZPP &gt;52 µmol/mol better (1 study)(^{65})</td>
<td>ZPP &gt;90 µmol/mol better (1 study)(^{65}) ZPP &gt;52 µmol/mol better (1 study)(^{65})</td>
<td>Combination of % HYPO &gt;6 with ZPP &gt;52 µmol/mol produced a substantial increase in sensitivity but reduce in specificity (1 study)(^{61})</td>
<td></td>
</tr>
</tbody>
</table>
Table 4. Interpretations of the summarized results for the direct comparisons of the overall test accuracy or sensitivity and specificity (at specified cutoff values) of newer versus classical markers of iron status (at baseline) to predict a response to intravenous iron treatment in seven cohort studies among adult HD CKD patients (continued)

<table>
<thead>
<tr>
<th>Iron Status Marker</th>
<th>Total Number of Studies (Total N) [Risk of Bias]</th>
<th>Overall Test Accuracy When Compared With TSAT</th>
<th>Sensitivity and Specificity When Compared With TSAT &lt;20%</th>
<th>Overall Test Accuracy When Compared With Ferritin</th>
<th>Sensitivity and Specificity When Compared With Ferritin &lt;100 ng/mL</th>
<th>Sensitivity and Specificity When Compared With TSAT &lt;20% or Ferritin &lt;100 ng/mL</th>
<th>Other Comparative Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>sTfR</td>
<td>2 studies44,61 (157) [1 medium,44 1 high risk61]</td>
<td>NS difference (2 studies)44,61</td>
<td>sTfR &gt;1.5 pg better (1 study)61</td>
<td>NS difference (2 studies)44,61</td>
<td>sTfR &gt;1.5 pg better (1 study)61</td>
<td>sTfR &gt;1.5 pg better (1 study)61</td>
<td>NS difference (2 studies)44,61</td>
</tr>
<tr>
<td>Hepcidin</td>
<td>1 study62 (56) [1 low risk62]</td>
<td>NS difference (1 study)62</td>
<td>NS difference (1 study)62</td>
<td>NS difference (1 study)62</td>
<td>NS difference (1 study)62</td>
<td>NS difference (1 study)62</td>
<td>NS difference (1 study)62</td>
</tr>
</tbody>
</table>

%HYPO=percent hypochromic red blood cells; AUC=area under the curve; CHr=content of hemoglobin in reticulocytes; IV=intravenous; NS=not significant; RetHe=reticulocyte hemoglobin equivalent; sTfR=soluble transferrin receptor; TSAT=transferrin saturation; ZPP=erythrocyte zinc protoporphyrin

a Response to IV iron treatment (the reference standard) was defined variably across studies (see also Table 2).
b The multivariate logistic regression analysis included HFE genotype, ferritin, TSAT, %Hypo, CHr, Hep-25 and Hep-20 in the same model.
Figure 4. Indirect comparisons of the overall test accuracy of newer with classical markers of iron status (at baseline) to predict response to IV iron among adult HD CKD patients—CHr, %HYPO, sTfR

%HYPO=percent hypochromic red blood cells; CHr=content of hemoglobin in reticulocytes; CRI=corrected reticulocyte index; EPO=erythropoietin stimulating agents; Hb=hemoglobin; HD=hemodialysis; IV=intravenous; sTfR=soluble transferrin receptor; TSAT=transferrin saturation; ZPP=erythrocyte zinc protoporphyrin

Note: Each symbol represents one reference standard, and sensitivity/specificity pairs from the same study (using different cutoffs) are connected with lines. Each study was labeled by its first author’s last name (next to the corresponding symbol). Studies that fall in the shaded area to the left of the near vertical line have a positive likelihood ratio ≥ 10, and studies that fall in the shaded area above the near horizontal line have a negative likelihood ratio ≤ 0.1. Studies that reported LR+ ≥ 10 and LR- ≤ 0.1 were deemed to have adequate predictive ability of the marker’s test result for the response to IV iron.
Figure 5. Indirect comparisons of the overall test accuracy of newer versus classical markers of iron status (at baseline) to predict response to IV iron among adult HD CKD patients—ZPP, Ferritin, TSAT

%HYPO=percent hypochromic red blood cells; CHr=content of hemoglobin in reticulocytes; CRI=corrected reticulocyte index; EPO=erythropoietin stimulating agents; Hb=hemoglobin; HD=hemodialysis; IV=intravenous; sTfR=soluble transferrin receptor; TSAT=transferrin saturation; ZPP=erythrocyte zinc protoporphyrin

Note: Each symbol represents one reference standard, and sensitivity/specificity pairs from the same study (using different cutoffs) are connected with lines. Each study was labeled by its first author’s last name (next to the corresponding symbol). Studies that fall in the shaded area to the left of the near vertical line have a positive likelihood ratio $\geq 10$, and studies that fall in the shaded area above the near horizontal line have a negative likelihood ratio $\leq 0.1$. Studies that reported LR$^+$ $\geq 10$ and LR$^-$ $\leq 0.1$ were deemed to have adequate predictive ability of the marker’s test result for the response to IV iron.
Content of Hemoglobin in Reticulocytes (CHr)/Reticulocyte Hemoglobin Equivalent (RetHe)

Key Points (Table 5)

Eight cohort studies enrolling 533 adult HD CKD patients,\textsuperscript{44,46,47,52,58,59,61,62} one cohort study enrolling 23 PD CKD patients,\textsuperscript{50} and one cohort study enrolling 95 ND CKD patients\textsuperscript{64} evaluated the test accuracy of CHr to predict a response to IV iron treatment. Of the eight studies in HD CKD patients, six compared the test performance of CHr with that of classical markers of iron status (TSAT or ferritin, alone or in combination with each other), and two studies reported the test performance of CHr alone. There were four different test platforms used across studies, and considerable heterogeneity likely existing between testing platforms. Of these studies, one was rated as being at low risk of bias, four at a medium risk of bias, and three at a high risk of bias. Studies enrolled primarily older patients who received maintenance ESA treatment; however, maintenance ESA doses varied across studies. Baseline iron status (based on mean serum ferritin and TSTA concentrations) also varied across studies.

Overall, there is a low level of evidence that CHr has similar or better overall test accuracy compared with classical markers (TSAT or ferritin) to predict a response to IV iron treatment among HD CKD patients. Four different definitions of a response to IV iron treatment were used among these eight studies. Studies examined the sensitivities and specificities at different cutoff values of CHr, ranging from <26 to <32 pg, to predict iron deficiency, but data did not allow us to assess threshold effect, due to the heterogeneity in the definitions of reference standards. Other heterogeneity, such as the variable iron status of the study populations and background treatment across studies, further limited our ability in making comparisons across studies. Two studies also reported the sensitivities and specificities of classical markers (TSAT <20 or ferritin <100 ng/mL) to predict iron deficiency, and data suggest that CHr (with cutoff values of <27 or <28 pg) has a better sensitivity and specificity to predict iron deficiency than classical markers (TSAT <20 or ferritin <100 ng/mL).\textsuperscript{44,59} Only one study performed multivariate analyses to predict a response to IV iron treatment (defined as an increase in Hct of ≥3 percent and/or a ≥ 30 percent reduction in EPO dose), and reported that CHr (with cutoff of <28 pg) had a much higher diagnostic odds ratio than serum ferritin (with cutoff of <300 ng/mL).\textsuperscript{47}

The strength of evidence is insufficient to draw conclusions regarding the test performance of CHr compared with that of classical markers of iron status among PD or ND CKD patients. We identified no study that evaluated the test performance of CHr to predict a response to IV iron treatment among pediatric CKD patients.

### Table 5. Overall strength of evidence for the test performance of reticulocyte hemoglobin content (CHr) comparing with that of classical markers of iron status to predict a response to IV iron treatment

<table>
<thead>
<tr>
<th>Number of Studies (Total N Analyzed)</th>
<th>Risk of Bias</th>
<th>Consistency</th>
<th>Directness</th>
<th>Precision</th>
<th>Overall Strength of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>8 (533 HD CKD patients)</td>
<td>1 low risk</td>
<td>Consistent</td>
<td>Indirect</td>
<td>Imprecise</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>4 medium risk</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 high risk</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (23 PD CKD patients)</td>
<td>1 medium risk</td>
<td>NA (only one study)</td>
<td>Not applicable (no direct comparison)</td>
<td>Imprecise</td>
<td>Insufficient</td>
</tr>
<tr>
<td>1 (95 ND CKD patients)</td>
<td>1 medium risk</td>
<td>NA (only one study)</td>
<td>Indirect</td>
<td>Imprecise</td>
<td>Insufficient</td>
</tr>
</tbody>
</table>

CKD=chronic kidney disease; HD=hemodialysis; NA=not applicable; ND=nondialysis; PD=peritoneal dialysis
Detailed Synthesis (Tables 3 and 6)

HD CKD Patients

Eight studies evaluated the test performance of CHr to predict a IV response in 533 adult CKD patients. Of these, one (with a total of 69 adult HD CKD patients) also evaluated the ability of RetHe to predict the response to IV iron treatment, and showed that CHr and RetHe are similar in terms of test performance. Study sample sizes ranged from 27 to 125 patients. The mean age of patients, reported in five studies, ranged from 59 to 67 years old; one additional study reported subjects’ ages (31 to 84 years), while the remaining two did not report subjects’ age. The baseline mean Hb concentrations (reported in 5 studies) ranged from 9.9 to 12.3 g/dL, mean ferritin concentrations from 84 to 347 ng/mL (reported in 8 studies), and mean TSAT from 18 to 39 percent (reported in 7 studies). Most studies reported that patients were on maintenance ESA treatment during the trial of iron treatment; however, maintenance ESA doses varied across studies. The indices monitored for assessing a response were Hb, hematocrit, and the corrected reticulocyte index (which is calculated by multiplying the reticulocyte count by the hematocrit and dividing the result by 40). The iron formulations used were ferric gluconate, iron sucrose, chondroitin-sulfate iron colloid, iron dextran, and iron saccharate. The duration of iron treatment also varied across studies, ranging from 2 weeks to 6 months. Of the eight total studies, two evaluated the ability of change in CHr values from baseline to 2 or 4 weeks to predict response to IV iron treatment.

Studies examined the sensitivities and specificities at different cutoff values of CHr, ranging from <26 to <32 pg, to predict iron deficiency; however, the data did not allow us to assess threshold effect, due to the heterogeneity in the definitions of reference standards. Four different definitions of reference standards (a response to IV iron treatment) were used: 1) an increase in Hb of ≥1 g/dL; 2) a ≥ 15 percent increase in Hb; 3) an increase in Hct of ≥3 percent and/or a ≥ 30 percent reduction in EPO dose; and 4) >1 point increase in corrected reticulocyte index. There was no uniform regimen of intravenous iron treatment in terms of dosage and iron formulation. There was also a wide range of durations of intravenous iron treatment (2 weeks to 5 months) across studies. One study was rated as being at a low risk of bias, six at a medium risk of bias, and three at a high risk of bias. The common limitations among the studies rated as being at medium or high risk of bias included potential selection bias (due to inclusion of nonconsecutive patients), inadequate description of recruitment and the study population, and inadequate information on the blinding between the test readers of the index and reference tests.

Studies reported either a similar (not statistically different) or better overall test accuracy for CHr as compared with TSAT and ferritin based on the AUC values (Table 3). Only one out of the eight studies performed multivariate analyses to predict a response to IV iron treatment (defined as an increase in Hct of ≥3 percent and/or a ≥ 30 percent reduction in EPO dose). The logistic regression model included both newer and classical markers as independent variables, with the marker cutoffs being derived from ROC curves. The study reported that CHr <28 pg was associated with a 29-fold increased in the odds of a response (odds ratio=29; 95 percent CI 5 to 157), which was much higher than the odds ratio for serum ferritin (OR=8.71; 95 percent CI: 1.55, 48.96, with cutoff of <300 ng/mL). This study also reported the odds ratios for predicting a response based on a >1.2 pg change in CHr from baseline to 2 weeks (OR=29.04 [5.36,157.33]) and a >1.2 pg change in CHr from baseline to 4 weeks (OR=6.2 [1.94,19.8]). In the lone study where CHr was used in combination with a newer marker (%HYPO with a cutoff < 6 percent), the combination showed a higher sensitivity with no change in specificity.
Only two studies reported the sensitivities and specificities of CHr (at different cutoff values) in comparison with classical markers (TSAT <20 or ferritin <100 ng/mL) to predict iron deficiency (as defined by a response to IV iron treatment). Data from these two studies showed that CHr cutoff values of <27 or <28 pg had a better sensitivity and specificity to predict iron deficiency than classical markers (TSAT <20 or ferritin <100 ng/mL). However, the two studies used different definitions for a response to IV iron treatment, which limited the interpretation of findings across studies.

To aid the indirect comparisons across studies, the ability of CHr, ferritin, and TSAT to predict a response to IV iron treatment was plotted in ROC space (Panel A of Figure 4, and Panels E and F of Figure 5, respectively). Through visual inspection of the ROC curves for the three markers, it appears that the curves for CHr are closer to the upper left hand corner (denoting perfect ability to predict response) than the curves for ferritin and TSAT, indicating better overall test accuracy.

**PD CKD Patients**

In PD CKD patients with anemia, one cohort study with 23 patients evaluated the ability of CHr to predict a response to IV iron treatment, defined as an increase ≥ 1.0 g/dL of Hb within three months of starting treatment. The study was rated as being at a medium risk of bias due to potential selection bias. This study assessed multiple cutoffs, ranging from 28 to 31 pg of CHr to predict a response. The reported ranges of sensitivity and specificity were 20 to 53 percent and 83 to 67 percent, respectively, from lowest to highest CHr cutoffs. The study also assessed multiple cutoffs for serum ferritin (<100 to <800 ng/mL) and TSAT (<20 to <50 percent) to predict a response. Ferritin <100 ng/mL had sensitivity and specificity of 13 and 100 percent, respectively. Similarly, TSAT <20% had sensitivity and specificity of 20 and 100 percent, respectively. The authors reported that none of the sensitivity specificity pairs for various cutoffs for CHr, ferritin and TSAT provided reliable estimates to predict response to iron. This conclusion is consistent with our interpretations based on the calculated Likelihood Ratios falling below our prespecified limits (LR+ ≥10 and LR- ≤0.1) suggesting that none of these tests have adequate predictive ability for diagnosing iron deficiency in PD CKD patients.

**ND CKD Patients**

One cohort study, enrolling 95 ND CKD patients, evaluated the test accuracy of CHr to predict response to iron treatment, defined as a Hb increase ≥ 1.0 g/dL. This cohort study (at a medium risk of bias) analyzed data from the IV iron arm of an RCT comparing the efficacy of IV iron sucrose with oral ferrous sulfate over a period of 8 weeks.

The study publication reported ROC curves for CHr, ferritin, and TSAT with different cutoffs indicated in the text; however, the locations of the cutoffs were not indicated on the curve. Hence, the ROC curves were digitized to obtain sensitivity/specificity pairs. It was assumed that the cutoffs were presented in ascending order of sensitivity. The CHr cutoffs used to define response to IV iron ranged from <25 to <35 pg. The ranges of sensitivity and specificity to determine response to IV iron were 0 to 95 percent and 97 to 24 percent, respectively, from lowest to highest CHr cutoffs.

The study also assessed multiple cutoffs for ferritin (<50 to <300 ng/mL), TSAT (<5 to <25 percent), and the combination of ferritin and TSAT to predict a response. The authors reported that CHr, ferritin and TSAT had “poor clinical utility” at each cutoff value examined. Through visual inspection of the ROC curves for the all markers, it appears that CHr covered larger AUC
than ferritin and TSAT, indicating better overall test accuracy. However, none of these markers were close to the upper left hand corner (denoting perfect ability to predict response).
Table 6. Summary results of the ability of CHr to predict the response to IV iron treatment

<table>
<thead>
<tr>
<th>Study, Year [UI]</th>
<th>Testing Platforms or Methods</th>
<th>N&lt;sub&gt;analyzed&lt;/sub&gt; (% Responders)</th>
<th>CHr Cut-off (pg)</th>
<th>Sens, %</th>
<th>Spec, %</th>
<th>AUC (CI)</th>
<th>Ferritin AUC (CI)</th>
<th>Comparison CHr vs. Ferritin P value</th>
<th>TSAT AUC (CI)</th>
<th>Comparison CHr vs. TSAT P value</th>
<th>Other Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tessitore, 2010&lt;sup&gt;62&lt;/sup&gt; [20538788]</td>
<td>PBSCIlc mass spectrometer and copperloaded immobilized metal-affinity capture ProteinChip arrays (IMAC30-Cu2+)</td>
<td>56 (38)</td>
<td>&lt;32</td>
<td>57</td>
<td>75</td>
<td>0.697 (0.537,0.855)</td>
<td>0.552 (0.391, 0.713)</td>
<td>NS</td>
<td>0.593 (0.431, 0.754)</td>
<td>CHr AUC not significantly different from AUC of hepcidin isoforms (P &gt;0.12)</td>
<td></td>
</tr>
<tr>
<td>Bovy, 2007&lt;sup&gt;44&lt;/sup&gt; [17237481]</td>
<td>ADVIA 120 cell counter system, Bayer</td>
<td>32 (38)</td>
<td>&lt;29</td>
<td>25</td>
<td>100</td>
<td>0.752 (0.583,0.921)</td>
<td>0.834 (0.685,0.983)</td>
<td>NS</td>
<td>0.896 (0.778,1.0)</td>
<td>NS</td>
<td>NR</td>
</tr>
<tr>
<td>Buttarello, 2010&lt;sup&gt;45&lt;/sup&gt; [20472854]</td>
<td>ADVIA 120 hematology system, Bayer (CHr)</td>
<td>59 (NR)</td>
<td>&lt;31.2</td>
<td>47</td>
<td>83</td>
<td>0.74 (0.60, 0.89)</td>
<td>0.53 (0.38, 0.69)</td>
<td>NR</td>
<td>0.56 (0.40, 0.72)</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td>XE 5000 (RetHe)</td>
<td>59 (NR)</td>
<td>&lt;30.6</td>
<td>45</td>
<td>83</td>
<td>0.72(0.58,0.86) P&lt;0.003</td>
<td>0.53 (0.38, 0.69)</td>
<td>NR</td>
<td>0.56 (0.40, 0.72)</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Mitsuiki, 2003&lt;sup&gt;58&lt;/sup&gt; [14586744]</td>
<td>ADVIA 120 hematology system, Bayer</td>
<td>27 (63)</td>
<td>&lt;32</td>
<td>100</td>
<td>90</td>
<td>0.95 (0.89,1.00)</td>
<td>0.591 (0.415,0.767)</td>
<td>NR</td>
<td>0.676 (0.474, 0.878)</td>
<td>NR</td>
<td>NR</td>
</tr>
</tbody>
</table>
Table 6. Summary results of the ability of CHr to predict the response to IV iron treatment (continued)

<table>
<thead>
<tr>
<th>Study, Year [UI]</th>
<th>Testing Platforms or Methods</th>
<th>N analyzed (% Responders)</th>
<th>CHr Cut-off (pg)</th>
<th>Sens, %</th>
<th>Spec, %</th>
<th>AUC (CI)</th>
<th>Ferritin AUC (CI)</th>
<th>Comparison CHr vs. Ferritin P value</th>
<th>TSAT AUC (CI)</th>
<th>Comparison CHr vs. TSAT P value</th>
<th>Other Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mittman, 1997 [9398141]</td>
<td>Technicon H3RTC Hematology Analyzer, Bayer Diagnostic</td>
<td>79 (59)</td>
<td>&lt;26</td>
<td>44</td>
<td>88</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&lt;27</td>
<td>67</td>
<td>82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&lt;28</td>
<td>78</td>
<td>71</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Change in CHr from baseline to 2 wks &gt;2 pg</td>
<td>100</td>
<td>31</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Change in CHr from baseline to 2 wks &gt;2.5 pg</td>
<td>89</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Change in CHr from baseline to 2 wks &gt;3 pg</td>
<td>56</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chuang, 2003 [12543894]</td>
<td>Technicon H*3 automated cell counter, Bayer Laboratory</td>
<td>65 (65)</td>
<td>&lt;28</td>
<td>78</td>
<td>87</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Change in CHr from baseline to 2 wks &gt;1.2 pg</td>
<td>80</td>
<td>83</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>OR=27.85 (5.37,144.3) with the best cutoff &gt;1.2 pg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Change in CHr from baseline to 4 wks &gt;1.2</td>
<td>87</td>
<td>83</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>OR=6.2 (1.94,19.8, P=0.002) with a cut off &gt;1.2 pg</td>
</tr>
</tbody>
</table>

Adult HD CKD (continued)
<table>
<thead>
<tr>
<th>Study, Year [UI]</th>
<th>Testing Platforms or Methods</th>
<th>N&lt;sub&gt;analyzed&lt;/sub&gt; (% Responders)</th>
<th>CHr Cut-off (pg)</th>
<th>Sens, %</th>
<th>Spec, %</th>
<th>AUC (CI)</th>
<th>Ferritin AUC (CI)</th>
<th>Comparison CHr vs. Ferritin P value</th>
<th>Comparison CHr vs. TSAT P value</th>
<th>Other Results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Adult HD CKD (continued)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fishbane, 1997 [9211366]</td>
<td>Technicon H*3, Bayer Laboratory</td>
<td>32</td>
<td>&lt;26</td>
<td>100</td>
<td>80</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Tessitore, 2001 [11427634]</td>
<td>Advia 120 Hematology Analyser, Bayer Diagnostics</td>
<td>125 (41)</td>
<td>≤29</td>
<td>57</td>
<td>93</td>
<td>0.798 (0.714, 0.880)</td>
<td>0.633 (0.514, 0.752)</td>
<td>P&lt;0.05</td>
<td>0.753 (0.669, 0.837)</td>
<td>NS</td>
</tr>
<tr>
<td><strong>Adult PD CKD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Domrongkitchaiporn, 1999 [10401012]</td>
<td>Technicon H*3, Bayer Laboratory</td>
<td>21 (71)</td>
<td>&lt;28</td>
<td>20</td>
<td>83</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&lt;29</td>
<td>47</td>
<td>83</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&lt;31</td>
<td>53</td>
<td>66</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Adult ND CKD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Van Wyck, 2005 [16316362]</td>
<td>NR</td>
<td>35 (44)</td>
<td>&lt; 25</td>
<td>0</td>
<td>97</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&lt; 27</td>
<td>3</td>
<td>92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&lt; 29</td>
<td>12</td>
<td>86</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&lt; 31</td>
<td>33</td>
<td>84</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&lt; 33</td>
<td>83</td>
<td>57</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&lt; 35</td>
<td>95</td>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AUC=area under the curve; CHr=content of hemoglobin in reticulocytes; CI=95% confidence interval; CRI=corrected reticulocyte index; IV=intravenous; NR=not reported; NS=not significant; OR=odds ratio; rHuEpo=recombinant human erythropoietin; SE=standard error; Sens=sensitivity; Spec=specificity; TSAT=transferrin saturation, UI=universal identifier/Pubmed ID
Percent Hypochromic Red Blood Cells

Key Points (Table 7)

Six cohort studies, enrolling a total of 365 adult HD CKD patients, evaluated the test performance of %HYPO to predict a response to IV iron treatment.44,46,52,61,62,67 There were two different test platforms used across studies. One study was rated as being at a low risk of bias, two at a medium risk, and three at a high risk of bias. Studies enrolled primarily older patients who received maintenance ESA treatment; however, maintenance ESA doses varied across studies. Baseline iron status (based on mean serum ferritin and TSTA concentrations) also varied across studies.

Overall, there is a low level of evidence that %HYPO has similar or better overall test accuracy compared with classical markers (TSAT or ferritin) to predict a response to IV iron treatment among HD CKD patients. Three different definitions of a response to IV iron treatment were used among these six studies. Studies examined the sensitivities and specificities of %HYPO, with a cutoff value of either >6% or >10%, to predict iron deficiency. Data suggest that %HYPO (with cutoff values of >6% or >10%) has a better sensitivity and specificity to predict iron deficiency than classical markers (TSAT <20 or ferritin <100 ng/mL). Furthermore, two studies (from the same group of investigators) performed a multivariate regression analysis, and it showed that %HYPO was the only significant predictor of a response to IV iron treatment among all other markers included in the model.61,62

We did not identify any study evaluated the test performance of %HYPO to predict a response to IV iron treatment among adult PD or ND CKD patients, or among pediatric CKD patients.

Table 7. Overall strength of evidence for the test performance of Percent Hypochromic Red Blood Cells (%HYPO) comparing with that of classical markers of iron status to predict a response to IV iron treatment

<table>
<thead>
<tr>
<th>Number of Studies (Total N Analyzed)</th>
<th>Risk of Bias</th>
<th>Consistency</th>
<th>Directness</th>
<th>Precision</th>
<th>Overall Strength of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 (356 CKD patients on HD)</td>
<td>1 low risk</td>
<td>Consistent</td>
<td>indirect</td>
<td>Imprecise</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>2 medium risk</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 high risk studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CKD=chronic kidney disease; HD=hemodialysis

Detailed Synthesis (Tables 3 and 8)

Six cohort studies, enrolling a total of 356 HD CKD patients, evaluated the ability of %HYPO to predict a response to IV iron treatment.44,46,52,61,62,67 All studies also compared the predictive ability of %HYPO with that of classical laboratory markers (TSAT or ferritin). One study recruited anemia HD CKD patients,46 and two studies excluded patients with high normal serum ferritin values.52,67 Most studies reported that patients were on maintenance ESA treatment during the IV iron treatment; however maintenance ESA doses varied across studies. The mean age of patients ranged from 57 to 80 years old (reported in four studies). Baseline mean Hb concentrations ranged from 9.9 to 12.3 g/dL (reported in five studies), mean ferritin concentrations ranged from 137 to 347 ng/mL (reported in six studies), and mean TSAT ranged from 18 to 27 percent (reported in five studies). Four studies defined a response to IV iron treatment as an increase in Hb concentration ≥1 g/dL after treatment,44,46,62,67 one study defined
response as ≥15 percent increase in Hb at any two consecutive measurements, and one study defined response as >1 point increase in corrected reticulocyte index within 2 weeks. There was no uniform regimen of intravenous iron treatment in terms of dosage and iron formulation. There was also a wide range of durations of intravenous iron treatment (2 weeks to 6 months) across studies. One study was rated as being at a low risk of bias, two at a medium risk, and three at a high risk of bias. The studies rated as being at a medium or high risk of had issues related to potential selection bias and inadequate descriptions of the study population, patient recruitment, and tests.

Three of the four studies showed that %HYPO reported a significantly better overall test accuracy as compared with TSAT and ferritin, based on the AUC values. These studies defined a response to IV iron treatment as either an increase in Hb concentration ≥1 g/dL after treatment, or ≥15 percent increase in Hb at any two consecutive measurements. Two studies also reported the sensitivities and specificities of classical markers (TSAT <20 or ferritin <100 ng/mL) to predict iron deficiency, and data suggest that %HYPO (with cutoff values of >6% or >10%) has a better sensitivity and specificity to predict iron deficiency than classical markers (TSAT <20 or ferritin <100 ng/mL). Furthermore, two studies performed a multivariate regression analysis, which showed that %HYPO was the only significant predictor of a response to IV iron treatment among all other markers (HFE genotype, ferritin, TSAT, %Hypo, CHr, Hep-25 and Hep-20 in the same model). Combination of markers were assessed in two studies. In one study, when %HYPO was combined with newer or classical markers, the sensitivity of the test combination was higher than %HYPO alone but the reported specificity was lesser than that of %HYPO alone. In the other study, the combination of %HYPO and classical markers resulted in lower sensitivity and high specificity.

To aid the indirect comparisons across studies, the abilities of %HYPO, ferritin, and TSAT to predict a response to IV iron treatment were plotted in ROC space (Panel B of Figure 4, and Panels E and F of Figure 5, respectively). Through visually inspection of the ROC curves for the three markers, it appears that there is a better test performance for %HYPO as compared with TSAT and ferritin, with the ROC curves for %HYPO being closer to the upper left hand corner (denoting perfect ability to predict the response) than the ROC curves for ferritin and TSAT. This is also supported by the higher AUC values reported for %HYPO as compared with ferritin and TSAT in all studies.
<table>
<thead>
<tr>
<th>Study, Year [UI]</th>
<th>Population</th>
<th>Testing Platforms or Methods</th>
<th>%HYPO Cutoff (%)</th>
<th>N&lt;sub&gt;analyzed&lt;/sub&gt; (% responders)</th>
<th>Sens, %</th>
<th>Spec, %</th>
<th>AUC (CI)</th>
<th>Ferritin AUC (CI)</th>
<th>Comparison %HYPO vs. Ferritin P value</th>
<th>TSAT AUC (CI)</th>
<th>Comparison %HYPO vs. TSAT P value</th>
<th>Other Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tessitore, 2010 [20538788]</td>
<td>HD CKD</td>
<td>Advia 120 Hematology Analyser</td>
<td>&gt;6</td>
<td>56 (38)</td>
<td>76</td>
<td>89</td>
<td>0.844 (0.737, 0.950)</td>
<td>0.552 (0.391, 0.713)</td>
<td>NS</td>
<td>0.593 (0.431, 0.754)</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>Bovy, 2007 [17237481]</td>
<td>HD CKD</td>
<td>Advia 120 cell counter</td>
<td>&gt;10</td>
<td>32 (38)</td>
<td>67</td>
<td>95</td>
<td>0.937 (0.837, 1.00)</td>
<td>0.834 (0.685, 0.983)</td>
<td>NS</td>
<td>0.896 (0.778, 1.014)</td>
<td>NS</td>
<td>NR</td>
</tr>
<tr>
<td>Buttarello, 2010 [20472854]</td>
<td>HD CKD</td>
<td>Advia 120</td>
<td>≥ 5.8</td>
<td>59 (NR)</td>
<td>45</td>
<td>87</td>
<td>0.72 (0.58, 0.86)</td>
<td>0.53 (0.38, 0.69)</td>
<td>NR</td>
<td>0.56 (0.40, 0.72)</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Fishbane, 1999 [9211366]</td>
<td>HD CKD</td>
<td>Technicon H*3 hematology analyzer</td>
<td>&gt;10</td>
<td>32 (22)</td>
<td>43</td>
<td>80</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Silva, 1998 [9794562]</td>
<td>HD CKD</td>
<td>Technicon Mod. H2 System</td>
<td>&gt;10</td>
<td>33 (88)</td>
<td>10</td>
<td>100</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Tessitore, 2001 [11427634]</td>
<td>HD CKD</td>
<td>Advia 120 Hematology Analyser</td>
<td>&gt;6</td>
<td>125 (41)</td>
<td>82</td>
<td>95</td>
<td>0.93 (0.884, 0.976)</td>
<td>0.633 (0.514, 0.752)</td>
<td>P&lt;0.001</td>
<td>0.753 (0.669, 0.837)</td>
<td>P&lt;0.05</td>
<td>NR</td>
</tr>
</tbody>
</table>

Δ = Change in blood levels; AUC=area under the curve; CI=95% confidence interval; CKD=chronic kidney disease; CRI=corrected reticulocyte index; Hb=hemoglobin; HD=hemodialysis; IV=intravenous; NR=not reported; NS=not significant; Sens=sensitivity; Spec=specificity; UI=universal identifier/pubmed ID
Soluble Transferrin Receptor

**Key Points (Table 9)**

Two cohort studies, enrolling a total of 157 adult HD CKD patients, evaluated the test performance of sTfR to predict a response to IV iron treatment. Both studies also compared the test performance of sTfR with that of classical laboratory markers (TSAT or ferritin). One study was rated as being at a high risk of bias, and one at a medium risk of bias. The response to IV iron treatment was defined differently in the two studies, either as an increase in Hb concentration ≥1g/dL after intravenous iron treatment, or as an increase in Hb >15 percent from baseline.

Overall, there is a low level of evidence that sTfR has similar overall test accuracy compared with classical markers (TSAT or ferritin) to predict a response to IV iron treatment among HD CKD patients. We did not identify any study evaluated the test performance of sTfR to predict a response to IV iron treatment among adult PD or ND CKD patients, or among pediatric CKD patients.

**Table 9. Overall strength of evidence for the test performance of Soluble Transferrin Receptor (sTfR) comparing with that of classical markers of iron status to predict a response to IV iron treatment**

<table>
<thead>
<tr>
<th>Number of Studies (Total N Analyzed)</th>
<th>Risk of Bias</th>
<th>Consistency</th>
<th>Directness</th>
<th>Precision</th>
<th>Overall Strength of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 (157 HD CKD patients)</td>
<td>1 medium risk</td>
<td>Consistent</td>
<td>Indirect</td>
<td>Imprecise</td>
<td>Low</td>
</tr>
</tbody>
</table>

CKD=chronic kidney disease; HD=hemodialysis

**Detailed Synthesis (Tables 3 and Table 10)**

Two cohort studies, enrolling a total of 157 adult HD CKD patients (32 and 125 patients), evaluated the ability of sTfR to predict the response to IV iron treatment. Both studies also compared the test performance of sTfR with that of classical laboratory markers (TSAT or ferritin). Baseline mean Hb concentrations were 12.3 and 9.9 g/dL, mean ferritin concentrations were 347 and 201 ng/mL, and mean TSAT was 21 and 22 percent, respectively. One study was rated as being at a high risk of bias, and one at a medium risk of bias, due to potential selection bias, inadequate reporting of eligibility criteria, or inadequate descriptions of the study populations. The response to IV iron treatment were defined differently in the two studies, either as an increase in Hb concentration ≥1g/dL after intravenous iron treatment, or as an increase in Hb >15 percent from baseline. This limited our confidence in evaluating the consistency of findings across studies.

Both studies did not show significant differences in the overall test accuracy between sTfR and TSAT or ferritin, based on the AUC values. When sTfR (with a cutoff >1.5 pg) was combined with another newer marker (%HYPO with a cutoff >6 percent), the sensitivity of the test combination was higher than either test alone, but the reported specificity was lesser than that of %HYPO alone and higher than that of sTfR alone.

To aid the indirect comparisons across studies, the ability of sTfR, ferritin, and TSAT to predict a response to IV iron treatment was plotted in ROC space (Panel C of Figure 4, and Panels E and F of Figure 5, respectively). Through visual inspection of the ROC curves for the
three markers, it appears that there was no difference in the test performance between these three markers of iron status in predicting a response to IV iron treatment.
Table 10. Summary results of the ability of sTfR to predict the response to IV iron treatment in HD CKD patients

<table>
<thead>
<tr>
<th>Study, Year [UI]</th>
<th>Testing Platforms or Methods</th>
<th>sTfR Cut-off (pg)</th>
<th>Nanalyzed (% Responders)</th>
<th>Sens, %</th>
<th>Spec, %</th>
<th>AUC (CI)</th>
<th>Ferritin AUC (CI)</th>
<th>Comparison sTfR vs. Ferritin P value</th>
<th>TSAT AUC (CI)</th>
<th>Comparison sTfR vs. TSAT P value</th>
<th>Other Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bovy, 2007[^44] [17237481] Belgium</td>
<td>Enzyme-linked immunosorbent assays (QuantikineTM IVDTM, R&amp;D Systems, Minneapolis, MN, USA)</td>
<td>&gt;6.6 (Best cutoff)</td>
<td>32 (NR)</td>
<td>NR</td>
<td>NR</td>
<td>0.989 (0.922, 1.0)</td>
<td>0.834 (0.685, 0.983)</td>
<td>NS</td>
<td>0.896 (0.778, 1.014)</td>
<td>NS</td>
<td>NR</td>
</tr>
<tr>
<td>Tessitore, 2001[^61] [11427634]</td>
<td>Commercially available automated particle-enhanced immunephelometric (PETIA) assay (Dade Behring, Marburg, Germany), using highly purified sTfR isolated from human serum as a calibrator</td>
<td>&gt;1.5</td>
<td>125 (41)</td>
<td>81</td>
<td>71</td>
<td>0.7834 (0.668, 0.899)</td>
<td>0.633 (0.514, 0.752)</td>
<td>NS</td>
<td>0.753 (0.669, 0.837)</td>
<td>NS</td>
<td>NR</td>
</tr>
</tbody>
</table>

AUC=area under the curve; CI=95% confidence interval; CKD=chronic kidney disease; IV=intravenous; NR=not reported; NS=not significant; Sens=sensitivity; Spec=specificity; sTfR=soluble transferrin receptor; TSAT=transferrin saturation; UI=universal identifier/pubmed ID
Erythrocyte Zinc Protoporphyrin

Key Points (Table 11)

Two cohort studies, enrolling a total of 187 adult HD CKD patients, evaluated the test performance of ZPP in predicting a response to IV iron treatment.\textsuperscript{61,65} Both studies also compared the test performance of ZPP with that of classical laboratory markers (TSAT or ferritin). However, because the reference standards (Hb versus Hct/decrease in EPO dose) were not comparable, the two studies should be evaluated separately. Therefore, the strength of evidence is insufficient to draw conclusions regarding the test performance of ZPP compared with that of classical laboratory markers (TSAT or ferritin). When the three markers were assessed in a multivariate regression analysis in one study, the test performance of ZPP was comparable with TSAT and ferritin, and none of the three markers was a significant predictor of response to IV iron treatment.\textsuperscript{61}

We did not identify any study evaluated the test performance of ZPP to predict a response to IV iron treatment among adult PD or ND CKD patients, or among pediatric CKD patients.

Table 11. Overall strength of evidence for the test performance of ZPP comparing with that of classical markers of iron status to predict a response to IV iron treatment

<table>
<thead>
<tr>
<th>Number of Studies (Total N Analyzed)</th>
<th>Risk of Bias</th>
<th>Consistency</th>
<th>Directness</th>
<th>Precision</th>
<th>Overall Strength of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 (187 HD CKD patients)</td>
<td>2 high risk</td>
<td>Not applicable (different reference standards)</td>
<td>Indirect</td>
<td>Imprecise</td>
<td>Insufficient</td>
</tr>
</tbody>
</table>

CKD=chronic kidney disease; HD=hemodialysis

Detailed Synthesis (Tables 3 and 12)

Two cohort studies, enrolling a total of 187 adult HD CKD patients (62 and 125 patients), evaluated the test performance of ZPP in predicting a response to IV iron treatment.\textsuperscript{61,65} Both studies also compared the predictive ability of ZPP with that of classical laboratory markers (TSAT or ferritin). One study did not report any information on the anemia or iron status of the study population at baseline.\textsuperscript{65} The other study reported a mean Hb concentration of 9.9 g/dL, mean ferritin concentration of 201 ng/mL, and mean TSAT of 22 percent at baseline.\textsuperscript{61} Both studies were rated as being at a high risk of bias, due to a potential for selection bias or an inadequate description of the study population. The two studies used very different definitions to define a response to IV iron therapy: a 15 percent or more increase in Hb,\textsuperscript{61} or a 5 percent increase in Hct or a decrease in erythropoietin dose of more than 2000 units.\textsuperscript{65} Because the reference standards (Hb versus Hct/decrease in EPO dose) were not comparable, the two studies should be evaluated separately.

Both studies showed that ZPP and ferritin had a similar overall test accuracy, based on the AUC values. However, the studies also showed different findings comparing the test accuracy of ZPP with that of TSAT. Specifically, in predicting a response to IV iron treatment, one study reported a higher sensitivity and specificity for ZPP as compared with TSAT,\textsuperscript{65} and the other study reported a higher sensitivity and lower specificity for ZPP as compared with TSAT.\textsuperscript{61} When the three markers were assessed in a multivariate regression analysis in the latter study, the test accuracy of ZPP was comparable with TSAT and ferritin, and none of the three markers was
a significant predictor of response to IV iron treatment. This same study also assessed the test accuracy of ZPP combined with another newer marker (%HYPO) to predict a response to IV iron treatment, as compared with classical markers (TSAT or ferritin), and found that the test accuracy of the combination of newer markers (ZPP>52 pg or %Hypo >6 percent) was better than TSAT<20% or ferritin <100 ng/mL (either alone or in combination). The other study reported that utility of ZPP in predicting the need for IV iron is better than that of TSAT and ferritin.

To aid the indirect comparisons across studies, the ability of ZPP, ferritin, and TSAT to predict a response to IV iron treatment was plotted in ROC space (Panels D, E, and F of Figure 5, respectively). Through visual inspection of the ROC curves for the three markers, it appears that there was no difference in the overall test accuracy between these three markers of iron status in predicting a response to IV iron treatment.
Table 12. Summary results of the test performance of ZPP in predicting a response to IV iron treatment in adult HD CKD patients

<table>
<thead>
<tr>
<th>Study, Year [UI]</th>
<th>Testing platforms or methods</th>
<th>N&lt;sub&gt;analyzed&lt;/sub&gt; (% Responders)</th>
<th>ZPP Cut-off (pg)</th>
<th>Sens, %</th>
<th>Spec, %</th>
<th>AUC (CI)</th>
<th>Ferritin AUC (CI)</th>
<th>ZPP vs. Ferritin P value</th>
<th>TSAT AUC (CI)</th>
<th>ZPP vs. TSAT P value</th>
<th>Other Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fishbane, 1995[7872320]</td>
<td>Hematofluorometer, AVIV Biomedicals</td>
<td>62 (62)</td>
<td>&gt;52</td>
<td>100</td>
<td>17</td>
<td>0.853 (0.760, 0.946)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.785 (0.672, 0.897)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>NR</td>
<td>0.665 (0.53, 0.80)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&gt;66</td>
<td>90</td>
<td>35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&gt;90</td>
<td>87</td>
<td>83</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&gt;103</td>
<td>80</td>
<td>91</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&gt;107</td>
<td>70</td>
<td>91</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&gt;109</td>
<td>60</td>
<td>91</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&gt;112</td>
<td>50</td>
<td>96</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&gt;122</td>
<td>40</td>
<td>96</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&gt;138</td>
<td>30</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&gt;140</td>
<td>20</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&gt;177</td>
<td>10</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&gt;190</td>
<td>0</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tessitore, 2001[11427634]</td>
<td>Fluorometer, Shimadzu, Rf-551</td>
<td>125 (41)</td>
<td>&gt;52</td>
<td>81</td>
<td>69</td>
<td>0.77 (0.63, 0.91)</td>
<td>0.633 (0.51, 0.75)</td>
<td>NS</td>
<td>0.753 (0.67, 0.84)</td>
<td>NS</td>
<td>Not a significant predictor in the multivariate regression analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&gt;90</td>
<td>14</td>
<td>97</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ZPP &gt;52 or %HYPO &gt;6%</td>
<td>94</td>
<td>72</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AUC=area under the curve; CKD=chronic kidney disease; CI=confidence interval; EPO=erythropoietin; Hb=hemoglobin; HD=hemodialysis; Hct=hematocrit; IV=intravenous; NR=not reported; NS=not significant; Sens=sensitivity; Spec=specificity; TSAT=transferrin saturation; UI=universal identifier/pubmed ID; ZPP=erythrocyte zinc protoporphyrin

Hepcidin

Key Points
One prospective cohort study evaluated the test performance of both isoforms of hepcidin (hepcidin-20 and hepcidin-25) to predict iron deficiency among 56 older adult HD CKD patients who were on maintenance ESA treatment. The study was rated as being at a low risk of bias. The strength of evidence is insufficient to draw conclusions regarding the overall test accuracy or test accuracy of hepcidin-20 or hepcidin-25 comparing with that of classical markers of iron status among adult HD CKD patients.

We identified no study evaluating the test performance of hepcidin to predict a response to IV iron treatment among adult PD or ND CKD patients, or among pediatric CKD patients.

Detailed Synthesis (Tables 3 and 13)
One prospective cohort study evaluated the test performance of hepcidin-20 and hepcidin-25 to predict iron deficiency, defined by a response to IV iron treatment among 56 older adult HD CKD patients (mean age of 67 years). All enrolled patients were on maintenance ESA treatment, aiming at target Hb within the range of 10.5 to 12.5 g/dL. Baseline mean Hb concentration was 11.6 g/dL, mean ferritin concentration was 146 ng/mL, and mean TSAT was 20 percent. The study was rated as being at a low risk of bias. A response to IV iron treatment was defined as an increase in Hb concentration ≥1 g/dL after treatment with 62.5 mg ferric gluconate over 16 consecutive dialysis sessions.

The overall test accuracy to predict a response to IV iron treatment for hepcidin-20 or hepcidin-25 was no better than chance (AUC= 0.54 and 0.52, respectively), and was not significantly different from that of TSAT or ferritin.
Table 13. Summary results of the ability of serum hepcidin to predict the response to IV iron treatment in HD CKD patients

<table>
<thead>
<tr>
<th>Study, Year [UI]</th>
<th>Testing Platforms or Methods</th>
<th>( N_{\text{analyzed (% Responders)}} )</th>
<th>Index Test</th>
<th>Sens, %</th>
<th>Spec, %</th>
<th>AUC (CI)</th>
<th>Ferritin AUC (CI)</th>
<th>Comparison Hepcidin vs. Ferritin P value</th>
<th>TSAT AUC (CI)</th>
<th>Comparison Hepcidin vs. TSAT P value</th>
<th>Other Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tessitore, 2010[1]</td>
<td>PBSCIIc mass spectrometer and copperloaded immobilized metal-affinity capture ProteinChip arrays (IMAC30-Cu2+)</td>
<td>56 (NR)</td>
<td>Hepcidin-20</td>
<td>NR</td>
<td>NR</td>
<td>0.541 (0.373, 0.710)</td>
<td>0.552 (0.391, 0.713)</td>
<td>NS</td>
<td>0.593 (0.431, 0.754)</td>
<td>NS</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Hepcidin-25</td>
<td>NR</td>
<td>NR</td>
<td>0.517 (0.330, 0.672)</td>
<td>0.552 (0.391, 0.713)</td>
<td>NS</td>
<td>0.593 (0.431, 0.754)</td>
<td>NS</td>
<td>NR</td>
</tr>
</tbody>
</table>

AUC=area under the curve; CI=95% confidence interval; CKD=chronic kidney disease; HD=hemodialysis; IV=intravenous; NR=not reported; NS=not significant; Sens=sensitivity; Spec=specificity; TSAT=transferrin saturation; UI=universal identifier/pubmed ID
2b. Adverse Effects or Harms Associated With Testing

Only seven of the 27 identified studies reported information on harms.47,50,59,64-67 Specifically, three studies reported no adverse events associated with iron therapy during the study periods. A total of five deaths were reported across two studies. Studies did not attribute these deaths to either testing or treatment. However, iron testing itself is unlikely to cause deaths, and most of the reported harms were attributed to iron therapy (if reported). Additional details regarding these adverse events are provided in Table 14.

### Table 14. Adverse effects or harms reported in the 27 studies included in Key Question 2

<table>
<thead>
<tr>
<th>Author, Year [PMID]</th>
<th>Adverse Effects or Harms Reported</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chuang, 200347 [12543894]</td>
<td>No adverse reactions were found to be associated with iron therapy.</td>
</tr>
<tr>
<td>Domrongkitchaiporn, 199950 [10401012]</td>
<td>No adverse reaction developed during or immediately after intravenous iron infusion.</td>
</tr>
<tr>
<td>Kaneko, 200366 [12631092a]</td>
<td>Three patients died during this study. 2 patients in the CHr group died; 1 during week 4 (bacterial pneumonia) and one during week 16 (sudden death by unknown cause) of the trial period. 1 patient in the TSAT group died in week 7 because of a liver tumor that was not discovered at patient enrollment and randomization. 1 patient in the TSAT group was prematurely discontinued from the study because of massive bleeding due to a femoral bone fracture and need for blood transfusion. No differences in hospitalizations or infection rate were observed. 1 patient in the CHr group and 1 patient in the TSAT group were hospitalized for infection of renal cysts and internal shunt obstruction, respectively.</td>
</tr>
<tr>
<td>Mittman, 199759 [9398141]</td>
<td>No adverse reactions were found to be associated with iron treatment.</td>
</tr>
<tr>
<td>Silva, 199867 [9794562]</td>
<td>Four of 33 patients (12%)—Metallic taste, when iron administration was too fast; No anaphylactoid reactions; No skin rashes; No intestinal or respiratory allergy; No infectious complications when on IV iron; No hepatic or pancreatic dysfunction related to iron Tx.</td>
</tr>
<tr>
<td>Van Wyck, 200564 [16316362]</td>
<td>No serious adverse effects (hypersensitivity reaction, hospitalization or deaths) were reported associated with iron treatment. Gastrointestinal disturbances, constipation, nausea, vomiting and dyspepsia associated with oral iron therapy. Gastrointestinal disturbances, constipation, nausea/vomiting, dyspepsia, transient taste disturbance (dysgeusia), headache, myalgia and hypotension associated with IV iron treatment.</td>
</tr>
</tbody>
</table>

CHr=content of hemoglobin in reticulocytes; IV=intravenous; TSAT=transferrin saturation; Tx=treatment

a This study was also included in Key Question 3, and thus the same data on harms are also reported there.

### Key Question 3. Intermediate Outcomes Comparing Iron Management Guided by Newer Laboratory Markers With Those of Iron Management Guided by Older Laboratory Markers

#### Key Points (Table 15)

Two short-term RCTs (4 and 6 months), enrolling a total of 354 adult CKD patients (mean age of 60 years old) undergoing hemodialysis, compared the intermediate outcomes of iron management guided by classical markers of iron status (TSAT and/or ferritin) with those of iron management guided by a newer marker of iron status (CHr). It should be noted that the two trials
(one in U.S. and one in Japan) employed different protocols for initiating intravenous iron therapy and anemia management, which affect the applicability of the trial findings.

The two trials showed different findings in terms of the doses of epoetin required to maintain hematocrit (Hct) targets. Specifically, the U.S. trial showed that guiding iron management via CHr resulted in similar epoetin dosing compared with iron management guided by ferritin or TSAT. In contrast, the Japanese trial found the doses of epoetin were significantly decreased (lower by 36 percent) in the group guided by TSAT, but did not change significantly in the group guided by CHr. However, it should be noted that the Hct target was higher in the U.S. trial, which may explain that the U.S. trial used much higher doses of epoetin than the Japanese trial during the trial period. Despite the differences in the protocols for initiating intravenous iron therapy, both trials reported a significant decrease in the intravenous iron doses administered to patients whose iron management was guided by CHr compared with those guided by TSAT or ferritin. Only the Japanese trial specifically monitored the adverse events associated with study medication; no differences in the hospitalization or infection rates between the two iron management groups were reported.

In conclusion, there is a low level of evidence for a reduction in the number of iron status tests and resulting intravenous iron treatments (a post-hoc intermediate outcome) administered to patients whose iron management was guided by CHr compared with those guided by TSAT or ferritin. Both RCTs reported that Hct remained in the targeted ranges (an indication for the adequacy of anemia management) throughout the study period in all randomized arms, although the Hct target was higher in the U.S. trial than the Japanese trial. We identified no study comparing iron management guided by classical markers with that guided by newer markers (%HYPO, sTfR, Ret-He, ZPP, or hepcidin).

**Table 15. Overall strength of evidence for intermediate outcomes comparing iron management guided by newer laboratory markers with those of iron management guided by older laboratory markers**

<table>
<thead>
<tr>
<th>Number of Studies (Total N Analyzed)</th>
<th>Risk of Bias</th>
<th>Consistency</th>
<th>Directness</th>
<th>Precision</th>
<th>Overall Strength of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 RCTs (354 adult CKD HD patients)</td>
<td>2 medium risk</td>
<td>Inconsistent (dose of epoetin treatment) Consistent (dose of iron treatment—post hoc intermediate outcome)</td>
<td>Indirect</td>
<td>Imprecise</td>
<td>Low</td>
</tr>
</tbody>
</table>

CKD=chronic kidney disease; HD=hemodialysis; RCT=randomized controlled trial

**Detailed Synthesis**

**Description of Included Studies (Table 16)**

Two RCTs, with a total of 354 adult CKD patients (mean age: 60 years), undergoing hemodialysis were included. One trial was conducted in the U.S., with a followup duration of 6 months, and the other in Japan, with a followup duration of 4 months. Both trials compared the intermediate outcomes of iron management guided by classical markers (TSAT and/or ferritin) with those of iron management guided by a newer marker of iron status (CHr); however, the two trials employed different protocols for initiating intravenous iron therapy and anemia management. Both trials were rated as being at a medium risk of bias, as the analyses were conducted among trial completers only, and allocation concealment and the methods of randomization were not clearly reported.
Results
Both RCTs reported that the mean Hct remained in the targeted ranges throughout the study period in all randomized arms, suggesting that the anemia management protocols were adequate in both trials.

Dose of Epoetin (Table 17)
Both RCTs reported the dose of epoetin required to maintain the Hct target as the primary outcome. The epoetin dose adjustment schedule was more frequent in the U.S. trial (every 2 weeks) than the Japanese trial (twice per month, 3 days after the previous hemodialysis therapy). The Hct target was higher in the U.S. trial (Hct target between 33 and 36 percent) than the Japanese trial (Hct target between 29.5 and 32.5 percent). The protocols for initiating intravenous iron therapy also differed between the two trials. Generally, the U.S. trial used much higher doses of epoetin than the Japanese trial at baseline (12,232 vs. 4121 IU/week) and at the end of trial period (10,949 vs. 3606 IU/week).

The U.S. trial analyzed the change in the doses of epoetin administered among 138 patients who completed the 28-week trial. The investigators found a decreasing trend in the mean epoetin dose requirement for both iron management groups, but these trends were not statistically significant. Specifically, the mean epoetin dose decreased from 12,237 to 10,949 IU per week in the iron management group guided by the newer marker (CHr <29 pg), and decreased from 12,232 to 11,772 IU per week in the iron management group guided by the classical markers (ferritin <100 ng/mL or TSAT < 20 percent). The authors did not conduct statistical testing for the differences between groups.

The Japanese trial analyzed the change in the doses of epoetin administered among 184 patients who completed the 16-week trial. This trial showed a significant increase in the epoetin dose requirement in the iron management group guided by the newer marker (CHr <32.5 pg) from baseline (4121 IU/week) to 4-week followup (5426 IU/week, P<0.05). During later followup time points, a decreasing trend in the epoetin dose requirement (3957 and 3606 IU/week at 9 and 16 weeks, respectively) was observed; however, these doses did not differ significantly from the baseline dose. A similar trend was observed in the iron management group guided by the classical marker (TSAT < 20 percent). However, the dose of epoetin requirement was significantly lower in the iron management group guided by the classical marker from 11 weeks to the end of the 16-week trial (2528 and 2629 IU/week, respectively), compared with the doses in the iron management group guided by the newer marker.

Iron Testing and Resulting Iron Treatment
Total iron dose requirement was the primary outcome in the U.S. trial and the secondary outcome in the Japanese trial. The U.S. trial initiated 100 mg intravenous iron dextran treatment for 10 consecutive hemodialysis therapies, either when ferritin was <100 ng/mL or TSAT was < 20 percent (the group guided by classical markers) or when CHr was <29 pg (the group guided by the newer laboratory marker). Intravenous iron was not administered if ferritin was > 800 ng/mL or TSAT > 50 percent. Patients in the Japanese trial were treated with 40 mg iron colloid with chondroitin sulfate 3 times per week for 2 weeks at the end of each hemodialysis therapy when either TSAT was < 20 percent (the group guided by the classical marker) or CHr was <32.5 pg (the group guided by the newer laboratory marker).

The U.S. trial compared the number of courses of intravenous iron triggered, the number of patients in whom testing triggered a course of intravenous iron treatment, and the mean weekly...
dose of intravenous iron between the two iron management groups during the 28-week trial. Of the 64 patients in the newer marker group, CHr was tested a total of 369 times, resulting in 27 (42 percent) patients receiving 42 courses of intravenous iron; the weekly dose of intravenous iron dextran was 22.9 (±20.5 SD) mg. Of the 74 patients in the classical markers group, ferritin and TSAT were tested a total of 419 times, resulting in 59 (80 percent) patients receiving 104 courses of intravenous iron; the weekly dose of intravenous iron dextran was 47.7 (±35.5 SD) mg. The number of iron status tests and resulting treatments were significantly higher in the classical markers group.

The Japanese trial compared the total dose of iron colloid administered between the two iron management groups during the 16-week trial. There was a 4-week run-in period before the start of the RCT during which oral and intravenous iron administration was suspended. The total dosage of iron colloid administered was significantly higher in the classical marker group (as compared with the newer marker group) from 13 weeks to the end of the trial (mean total dose 377.5 vs. 267.7 mg, P<0.05).

Both RCTs compared the changes in iron status markers between the two iron management groups. In both RCTs, the CHr test displayed much less test variability, expressed as coefficient of variation (CV), in comparison with the ferritin or TSAT tests. The reported CVs for CHr, ferritin, and TSAT were 3.4, 43.6, and 39.5 percent, respectively, in the U.S. trial; and 6.3, 130.5, and 48.9 percent, respectively in the Japanese trial. In both trials, none of the iron status markers differed significantly between the two iron management groups at baseline; however, changes in markers after iron treatments were inconsistent across the two trials (Table 18).

**Adverse Events**

In the U.S. trial, 19 (12 percent) patients were withdrawn during the study period. Reasons for withdrawal included prolonged hospitalization (8 patients), bleeding requiring blood transfusion (3 patients), transplant (1 patient), withdrawal of consent (1 patient), protocol violation (4 patients), and death (2 patients). However, any association of these events with iron testing or study medication is unclear. It was also not clear whether the dropout rate was unbalanced between the two randomized groups, but it is likely that more patients dropped out from the iron management group guided by CHr than the group guided by classical markers, based on the number of completers.

The Japanese trial specifically monitored the adverse events associated with study medication during the trial. Signs and symptoms were evaluated during and after each hemodialysis session, and the rates of incidence of hospitalization, infections, and deaths were recorded. There were a total of three deaths (2 patients in the CHr group; 1 patient in the TSAT group) due to bacterial pneumonia (at week 4 in the CHr group), sudden death by unknown cause (at week 16 in the CHr group), and liver tumor (at week 7 in the TSAT group). One patient in the TSAT group dropped out because of massive bleeding due to a femoral bone fracture and need for blood transfusion. There were no significant differences in the hospitalization or infection rates of the two iron management groups. Overall, two patients were hospitalized: one due to infection of renal cysts (1 patient in the CHr group) and one due to internal shunt obstruction (1 patient in the TSAT group).
Table 16. Characteristics of randomized controlled trials comparing intermediate outcomes of iron management guided by classical laboratory markers with those of iron management guided by newer laboratory markers of iron status in CKD patients undergoing hemodialysis

<table>
<thead>
<tr>
<th>Study, Year [UI] Country</th>
<th>Nenrolled / Nanalyzed</th>
<th>Demographics</th>
<th>Duration of HD</th>
<th>Anemia and Iron Status Indices</th>
<th>Intervention</th>
<th>Comparator</th>
<th>Iron Treatment Regimen</th>
<th>Anemia Management Protocol Targets</th>
<th>Followup Months</th>
<th>Risk of Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fishbane, 2001 U.S.</td>
<td>157/138</td>
<td>Male (%):54 Age (yr): 60 Race (%): -Caucasian 46 -African American 44 -Hispanic 7 Other 3</td>
<td>≥3 months</td>
<td>Hb (g/dL):NR Hct (%):35.6 ferritin (ng/mL):240.6 TSAT (%) 23.5</td>
<td>Iron management based on serum Chr measured every 4 wks</td>
<td>Iron management based on ferritin or TSAT measured every 4 wks</td>
<td>IV iron dextran 100 mg for 10 consecutive treatments if presence of iron Tx trigger</td>
<td>Hb and Hct every 2 wks; the dose of EPO adjusted to maintain Hct 33–36%&lt;sup&gt;a&lt;/sup&gt;</td>
<td>6</td>
<td>Medium</td>
</tr>
<tr>
<td>Kaneko, 2003 Japan</td>
<td>197/183</td>
<td>Male (%):61 Age (yr): 59 Race (%): NR</td>
<td>2 months to 26 years</td>
<td>Hb (g/dL):NR Hct (%):31.3 ferritin (ng/mL):247.4 TSAT (%) 25.8</td>
<td>Iron management based on serum Chr measured twice a month</td>
<td>Iron management based TSAT measured twice a month</td>
<td>IV iron colloid with chondroitin sulfate 40 mg 3 times per wk for 2 wks if presence of iron Tx trigger</td>
<td>Hct twice per mo; the dose of EPO adjusted to maintain Hct 29.5–32.5%&lt;sup&gt;b&lt;/sup&gt;</td>
<td>4</td>
<td>Medium</td>
</tr>
</tbody>
</table>

Chr=reticulocyte hemoglobin content; EPO=erythropoietin; Hb=hemoglobin; Hct=hematocrit; HD=hemodialysis; NR=not reported; TSAT=transferrin saturation

<sup>a</sup> Protocol called for 25% dose reductions for Hct >36% and holding doses if Hct >40%, or 50% dose increases for Hct <33%.

<sup>b</sup> Doses of EPO administered were categorized as 0, 750, 1500, 2250, 3000, 4500, 6000, or 9000 IU/week and modified as follows: (1) dose was raised by 200% if Hct <26%; (2) dose was raised by 100% if 26% ≤ Hct <29.5%; (3) dose was raised by 50% if 29% ≤ Hct <32.5%; (4) dose was reduced by 33% if 32.5% ≤ Hct <33%; (5) dose was reduced by 50% if 33% < Hct ≤36%; (6) if Hct >36%, administration of EPO was suspended. When the administration of EPO had been suspended and Hct was <29.5%, 2250 IU/week of EPO was resumed. If the modified EPO dose in accordance with rule mentioned earlier did not apply to any of the categories, the nearest dose category was adopted.
Table 17. Dose of epoetin required to maintain hematocrit targets

<table>
<thead>
<tr>
<th>Study, Year [Ul] Country</th>
<th>Arms (Trigger for Iron Tx)</th>
<th>N</th>
<th>Unit</th>
<th>Baseline</th>
<th>4 Wks</th>
<th>8 Wks</th>
<th>9 Wks</th>
<th>11 Wks</th>
<th>16 Wks</th>
<th>24 Wks</th>
<th>28 Wks</th>
<th>P within</th>
<th>P between</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fishbane, 2001[17]</td>
<td>Iron management guided by serum CHr measured every 4 wks (CHr &lt;29 pg)</td>
<td>64</td>
<td>Mean (SD), IU/week</td>
<td>12237 (12001)</td>
<td>NR</td>
<td>12200 (12049)</td>
<td>NR</td>
<td>NR</td>
<td>11300 (11785)</td>
<td>10933 (12095)</td>
<td>10949 (12154)</td>
<td>NS</td>
<td>NR</td>
</tr>
<tr>
<td>U.S.</td>
<td>Iron management guided by ferritin or TSAT measured every 4 wks (ferritin &lt;100 ng/mL or TSAT &lt; 20%)</td>
<td>74</td>
<td>Mean (SD), IU/week</td>
<td>12232 (11029)</td>
<td>NR</td>
<td>12077 (11444)</td>
<td>NR</td>
<td>NR</td>
<td>12100 (11029)</td>
<td>11902 (11320)</td>
<td>11772 (11780)</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>Kaneko, 2003[66]</td>
<td>Iron management guided by serum CHr (CHr &lt;32.5 pg) twice a month</td>
<td>94</td>
<td>Mean (SD), IU/week</td>
<td>4121 (2922)</td>
<td>NR</td>
<td>5426 (3481)</td>
<td>NR</td>
<td>3957 (3320)</td>
<td>3638 (3276)</td>
<td>3606 (3347)</td>
<td>NA</td>
<td>NA</td>
<td>&lt;0.05 from 11 wks</td>
</tr>
<tr>
<td>Japan</td>
<td>Iron management guided by TSAT (TSAT &lt;20%) twice a month</td>
<td>89</td>
<td>Mean (SD), IU/week</td>
<td>4081 (3123)</td>
<td>NR</td>
<td>4803 (3325)</td>
<td>NR</td>
<td>3051 (2730)</td>
<td>2528 (2730)</td>
<td>2629 (2640)</td>
<td>NA</td>
<td>NA</td>
<td>&lt;0.01 from 11 wks</td>
</tr>
</tbody>
</table>

CHr= reticulocytes hemoglobin content; NA=not applicable; NR=not reported; NS=not statistically significant; SD=standard deviation; TSAT=transferrin saturation; Tx=treatment; UI=universal identifier; wk=week
<table>
<thead>
<tr>
<th>Study, Year [UI]</th>
<th>Country</th>
<th>Arms (Trigger for Iron Tx)</th>
<th>N</th>
<th>Followup Duration (wk)</th>
<th>Outcome</th>
<th>Unit</th>
<th>Baseline</th>
<th>Final</th>
<th>Pwithin</th>
<th>Pbetween</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fishbane, 2001</td>
<td>U.S.</td>
<td>Iron management guided by serum CHr measured every 4 wks (CHr &lt;29 pg)</td>
<td>64</td>
<td>28</td>
<td>ferritin</td>
<td>Mean (SD), ng/mL</td>
<td>251.7 (231.3)</td>
<td>304.7 (290.6)</td>
<td>NS</td>
<td>0.05 at final</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Iron management guided by ferritin or TSAT measured every 4 wks (ferritin &lt;100 ng/mL or TSAT &lt; 20%)</td>
<td></td>
<td></td>
<td>ferritin</td>
<td>Mean (SD), ng/mL</td>
<td>229.6 (178.8)</td>
<td>399.5 (247.6)</td>
<td>&lt;0.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Iron management guided by serum CHr measured every 4 wks (CHr &lt;29 pg)</td>
<td>64</td>
<td>28</td>
<td>TSAT</td>
<td>Mean (SD), %</td>
<td>22.3 (11.7)</td>
<td>25.8 (16.6)</td>
<td>NS</td>
<td>0.04 at final</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Iron management guided by ferritin or TSAT measured every 4 wks (ferritin &lt;100 ng/mL or TSAT &lt; 20%)</td>
<td></td>
<td></td>
<td>TSAT</td>
<td>Mean (SD), %</td>
<td>24.7 (12.7)</td>
<td>29.4 (17.8)</td>
<td>&lt;0.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Iron management guided by serum CHr measured every 4 wks (CHr &lt;29 pg)</td>
<td>64</td>
<td>28</td>
<td>CHr</td>
<td>Mean (SD), pg</td>
<td>30.8 (1.7)</td>
<td>30.8 (1.8)</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Iron management guided by ferritin or TSAT measured every 4 wks (ferritin &lt;100 ng/mL or TSAT &lt; 20%)</td>
<td></td>
<td></td>
<td>CHr</td>
<td>Mean (SD), pg</td>
<td>31.1 (1.8)</td>
<td>31.1 (1.8)</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Kaneko, 2003</td>
<td>Japan</td>
<td>Iron management guided by serum CHr (CHr &lt;32.5 pg) twice a month</td>
<td>94</td>
<td>16</td>
<td>ferritin</td>
<td>Mean (SD), ng/mL</td>
<td>234.5 (307.0)</td>
<td>279.5 (326.9)</td>
<td>&lt;0.01</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Iron management guided by TSAT (TSAT &lt;20%) twice a month</td>
<td>89</td>
<td></td>
<td>ferritin</td>
<td>Mean (SD), ng/mL</td>
<td>257.0 (453.4)</td>
<td>372.6 (518.1)</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Iron management guided by serum CHr (CHr &lt;32.5 pg) twice a month</td>
<td>94</td>
<td>16</td>
<td>TSAT</td>
<td>Mean (SD), %</td>
<td>25.5 (12.6)</td>
<td>28.2 (14.3)</td>
<td>NS</td>
<td>&lt;0.05 at final</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Iron management guided by TSAT (TSAT &lt;20%) twice a month</td>
<td>89</td>
<td></td>
<td>TSAT</td>
<td>Mean (SD), %</td>
<td>25.7 (15.6)</td>
<td>32.7 (14.9)</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Iron management guided by serum CHr (CHr &lt;32.5 pg) twice a month</td>
<td>94</td>
<td>16</td>
<td>CHr</td>
<td>Mean (SD), pg</td>
<td>33.2 (2.2)</td>
<td>34.4 (1.6)</td>
<td>&lt;0.0001</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Iron management guided by TSAT (TSAT &lt;20%) twice a month</td>
<td>89</td>
<td></td>
<td>CHr</td>
<td>Mean (SD), pg</td>
<td>32.8 (2.4)</td>
<td>34.3 (1.9)</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
</tbody>
</table>

CHr= reticulocytes hemoglobin content; NA=not applicable; NR=not reported; NS=not statistically significant; TSAT=transferrin saturation; Tx=treatment; UI=universal identifier; wk=week
Key Question 4. Factors Affecting Test Performance and Clinical Utility

Key Points
Only single studies or indirect comparisons across studies provided data on the potential impacts of some factors (i.e., interactions between iron and ESA treatment, route of iron administration, and treatment regimen) on the test performance of newer or classical laboratory markers of iron status. Therefore, the strength of evidence is insufficient to draw conclusions regarding factors that may affect the test performance or clinical utility of laboratory markers of iron status.

Detailed Synthesis (Tables 19 and 20)
Although only two studies included in this section, from all 27 studies included in Key Questions 2, relevant data on factors that may affect the test performance of laboratory markers of iron status were also reported here.

Interactions Between Iron and ESA Treatment
One trial randomized 134 HD CKD patients to either no IV iron or IV iron (1 gram of ferric gluconate) group.68 This trial was rated as being at a medium risk of bias. This trial enrolled a special population of HD CKD patients with high ferritin (500-1200 ng/mL) and a low TSAT levels (≤ 25%), possibly due to functional iron deficiency. The test accuracy of baseline laboratory biomarkers of iron status in predicting a response to ESA treatment, defined as a Hb increase ≥2 g/dL, was assessed in both groups (IV iron or no IV iron group). Baseline epoetin doses were raised by 25 percent in both groups, starting with the first hemodialysis session of week 1 and then maintained for the entire study until the first hemodialysis session of week 6. Laboratory biomarkers were obtained weekly.

Within the no intravenous iron group (25% epoetin dose increase alone), the sensitivity and specificity pairs for a TSAT cutoff of ≥19 percent and a ferritin cutoff of ≥726 ng/mL were 29 and 70 percent, and 27 and 69 percent, respectively. The sensitivity and specificity pairs for a CHr cutoff of ≥31.2 pg and a sTfR cutoff of ≥5.9 mg/L were 27 and 69 percent, and 35 and 77 percent, respectively. Multivariable logistic regression analysis showed that none of response markers (including TSAT, ferritin, CHr, sTfR, c-reactive protein, and epoetin) other than absolute value of epoetin dose increase predicted a statistically and clinically significant response to anemia treatment.

In contrast, in the intravenous iron group, a cutoff of CHr of ≥31.2 pg had a higher sensitivity (64 percent) and specificity (75 percent) in predicting treatment response. However, the test accuracies were lower for sTfR, TSAT, and ferritin. Multivariable logistic regression analysis showed that a higher baseline CHr and a lower baseline c-reactive protein predicted greater likelihood of a response to anemia and iron treatment. In the intention to treat population, the odds ratio of achieving a ≥2 g/dL Hb response in patients with baseline CHr ≥31.2 pg relative to those with lower values was 5.3 (95 percent CI, 1.78, 15.83).

Biological Variation in Diagnostic Indices
No study examined the impacts of biological variation in diagnostic indices on the test performance or clinical utility of laboratory markers of iron status.
Use of Different Diagnostic Reference Standards

Included in Key Question 2a, one study examined the test performance of RetHe using two different reference standards, and showed that the test performance of RetHe was less favorable for assessing “functional iron deficiency” (TSAT<20%, ferritin 100-800 ng/mL, and Hb <11 g/dL) than for assessing “traditional parameters for iron deficiency” (serum iron < 40 µg/dL, TSAT<20%, ferritin <100 ng/mL, and Hb <11 g/dL) in HD CKD patients. In addition, the heterogeneity in the definitions for the reference standard (a response to IV iron treatment) may explain the differences in study findings.

Type of Dialysis (i.e., Peritoneal or Hemodialysis)

No study examined the impacts of type of dialysis on the test performance or clinical utility of laboratory markers of iron status.

Patient Subgroups

No study performed analyses by patient subgroups.

Route of Iron Administration (i.e., Oral or Intravenous) or Treatment Regimen (i.e., Repletion or Continuous Treatment)

No study examined the impacts of route of iron administration or treatment regimen on test performance or clinical utility of laboratory markers of iron status. Indirect comparisons between studies included in the Key Question 2 and the studies included in this section suggest potential impacts of these factors on the test accuracy of newer and classical laboratory markers of iron status.

Most studies included in the Key Question 2 reported that patients were on maintenance ESA (i.e., no change in ESA dose during study) and received IV iron treatment. This is in contrast to the study included in this section. Two cohorts (reported in one article, rated as being at a medium risk of bias) assessed test performance of sTfR in predicting a Hb response to initiation of ESA treatment (> 2 g/dL increase in Hb at 3 months after initiation of ESA therapy, study 1 in the article), and in predicting a response to an increase in ESA treatment dose (> 1 g/dL increase in Hb 4 weeks from baseline, study 4 in the article). Both cohorts also treated patients with oral iron. The results from the first cohort showed that a sTfR cutoff of <6 mg/L had better specificity, but the same sensitivity, than a ferritin cutoff of >50 µg/L in predicting an Hb response to initiation of ESA treatment in 17 adult HD CKD patients. In the second study (16 adult HD CKD patients), the results showed that the change in sTfR >20 percent from baseline to week 1 had perfect specificity but a lower sensitivity in predicting a Hb response to an increase in ESA treatment dose.
Table 19. Characteristics of studies evaluating factors affecting test performance and clinical utility

<table>
<thead>
<tr>
<th>Study, Year [UI] Country, Design</th>
<th>Study Population</th>
<th>Groups</th>
<th>Intervention</th>
<th>N</th>
<th>Demographics</th>
<th>Anemia and Iron Status Indices</th>
<th>Followup Duration (wk)</th>
<th>Risk of Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singh, 200768 [17396118] U.S. RCT</td>
<td>HD CKD</td>
<td>IV iron</td>
<td>IV ferric gluconate 1 g &amp; 25% increase in weekly epoetin dose for 6 weeks</td>
<td>64</td>
<td>Male (%):58 Age (yr): 61 Race (%): -Caucasian 31 -African American 47 -Hispanic 14 -Asian/Pacific islander 8</td>
<td>Hb (g/dL): 10.4 Hct (%): NR Ferritin (ng/mL): 759 TSAT (%): 18.5</td>
<td>6</td>
<td>Medium</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No IV iron</td>
<td>25% increase in weekly epoetin dose for 6 weeks</td>
<td>65</td>
<td>Male (%):43 Age (yr): 59 Race (%): -Caucasian 31 -African American 51 -Hispanic 14 -Asian/Pacific islander 3 Other 2</td>
<td>Hb (g/dL): 10.2 Hct (%): NR Ferritin (ng/mL): 765 TSAT (%): 19</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Ahluwalia, 1997(study 1; study 4)69 [9328369] U.S. Prospective cohorts</td>
<td>HD CKD</td>
<td>ESA naïve patient starting ESA treatment</td>
<td>One ferrous sulfate tablet (containing 50 mg elemental iron) per day with mean ESA dose of 162 IU/kg/week</td>
<td>17</td>
<td>Male (%):NR Age (yr): 46 Race (%):NR</td>
<td>Hb (g/dL): 7.1 Hct (%):NR Ferritin (ng/mL):98.5 TSAT (%):NR</td>
<td>12</td>
<td>Medium</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Increase in ESA dose for patient on maintenance ESA treatment</td>
<td>One 65 mg ferrous sulfate tablet per day with mean ESA dose of 121 IU/kg/week</td>
<td>16</td>
<td>Male (%):NR Age (yr): 46 Race (%):NR</td>
<td>Hb (g/dL): 7.8 Hct (%):NR Ferritin (ng/mL):59.5 TSAT (%):17</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

CKD=chronic kidney disease; ESA=erythropoiesis stimulating agents; Hb=hemoglobin; Hct=hematocrit; HD=hemodialysis; IV=intravenous; IU=international units; NR=not reported; TSAT=transferrin saturation; wk=week; yr=year
Table 20. Test accuracy of TSAT, ferritin, CHr, and sTfR for predicting change in hemoglobin in subgroups of IV iron and no IV iron treatment

<table>
<thead>
<tr>
<th>Study, Year [UI] Country Design</th>
<th>Group</th>
<th>N</th>
<th>Reference Standard—Dx of Iron Deficiency</th>
<th>Index Test</th>
<th>Cutoff for Index Test</th>
<th>Sensitivity, %</th>
<th>Specificity, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singh, 2007 [17396118] U.S. RCT</td>
<td>IV Iron</td>
<td>64</td>
<td>Hb change of ≥ 2 g/dL</td>
<td>TSAT</td>
<td>≥ 19%</td>
<td>48.5</td>
<td>54.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ferritin</td>
<td>≥ 726 ng/mL</td>
<td>46.9</td>
<td>53.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CHr</td>
<td>≥ 31.2 pg/cell</td>
<td>63.9</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>sTfR</td>
<td>≥ 5.9 mg/L</td>
<td>42.4</td>
<td>48.4</td>
</tr>
<tr>
<td></td>
<td>No IV Iron</td>
<td>65</td>
<td>Hb change of ≥ 2 g/dL</td>
<td>TSAT</td>
<td>≥19%</td>
<td>28.6</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ferritin</td>
<td>≥ 726 ng/mL</td>
<td>27.3</td>
<td>66.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CHr</td>
<td>≥ 31.2 pg/cell</td>
<td>26.7</td>
<td>68.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>sTfR</td>
<td>≥ 5.9 mg/L</td>
<td>35.3</td>
<td>77.4</td>
</tr>
<tr>
<td>Ahluwalia, 1997 (study 1; study 4) [9328369] U.S. Prospective cohorts</td>
<td>ESA naïve patient starting ESA treatment</td>
<td>16</td>
<td>&gt; 2 g/dL increase in Hb at 3 months after initiation of rHuEPO therapy</td>
<td>sTfR</td>
<td>&lt;6 mg/L</td>
<td>88</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ferritin</td>
<td>&gt;50 µg/L</td>
<td>88</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ferritin &amp; sTfR</td>
<td>&gt;50 µg/L &amp; &lt;6 mg/L</td>
<td>75</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>Increase in ESA dose for patient on ESA treatment</td>
<td>17</td>
<td>&gt; 1 g/dL increase in Hb 4 weeks over the baseline level</td>
<td>sTfR</td>
<td>&gt; 20 % increase in sTfR at 1 week</td>
<td>69</td>
<td>100</td>
</tr>
</tbody>
</table>

CHr=reticulocytes hemoglobin content; Dx=diagnosis; ESA=erythropoiesis stimulating agents; Hb=hemoglobin; Hct=hematocrit; IV=Intravenous; NR=not reported; rHuEPO=recombinant human erythropoietin; sTfR=soluble transferrin receptor; TSAT=transferrin saturation; Tx=treatment
**Discussion**

**Key Findings and Strength of Evidence**

We did not identify any study that provided data directly addressing our overarching question regarding the impact on patient-centered outcomes (mortality, morbidity, quality of life, and adverse effects) of using newer laboratory biomarkers. In the absence of direct evidence, the overarching question could be answered by the component questions (Key Questions 2, 3, and 4). A number of studies addressing these component questions were identified. A summary of the strength of evidence addressing each Key Question is provided in Table 21.

<table>
<thead>
<tr>
<th>Key Question</th>
<th>Summary, Comments, and Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>2a. What reference standards are used for the diagnosis of iron status in studies evaluating test accuracy?</td>
<td>- There is a lack of generally accepted reference standard tests for determining iron deficiency in the setting of CKD. This is reflected by the fact that current studies use two distinct methods to operationalize a reference standard for assessing test performance: 1) a response to intravenous (IV) iron treatment, often referred as “functional iron deficiency”; 2) classical laboratory biomarkers, alone or in combination with each other, often referred as “absolute iron deficiency.” However, across studies, there are large variations in the definitions of these reference standards.</td>
</tr>
</tbody>
</table>

| Key Question 2. What is the diagnostic test accuracy of newer markers of iron status as a replacement for or an add-on to classical laboratory markers? | Low / Insufficient (depending on the test comparisons, study populations, or test performance outcomes) • Among adult HD CKD patients, there is a low level of evidence that: o Content of hemoglobin in reticulocytes (CHr) has similar or better overall test accuracy compared with TSAT or ferritin to predict a response to IV iron treatment. Data from a few studies suggest that CHr (with cutoff values of <27 or <28 pg) has better sensitivity and specificity to predict iron deficiency than classical markers (TSAT <20 or ferritin <100 ng/mL). o Percent hypochromic red blood cells (%HYPO) has similar or better overall test accuracy compared with TSAT, and better overall test accuracy compared with ferritin to predict a response to IV iron treatment. Data suggest that %HYPO (with cutoff values of >6% or >10%) has a better sensitivity and specificity to predict iron deficiency (as defined by a response to IV iron treatment) than classical markers (TSAT <20% or ferritin <100 ng/mL). o Soluble transferring receptor (sTfR) has a similar test performance compared with classical markers (TSAT or ferritin) to predict a response to IV iron treatment. • There is insufficient evidence regarding: o Test performance of newer markers of iron status as an add-on to older markers. o Test performance comparing ZPP and hepcidin to predict a response to IV iron treatment in adult HD CKD patients. o Test performance comparing newer with classical laboratory markers to predict a response to IV iron treatment, in adult PD CKD and ND CKD patients, and in pediatric CKD patients. |

<table>
<thead>
<tr>
<th>Key Question</th>
<th>Strength of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>2a. What reference standards are used for the diagnosis of iron status in studies evaluating test accuracy?</td>
<td>Not rated (descriptive data)</td>
</tr>
</tbody>
</table>
### Table 21. Summary of the strength of evidence addressing Key Questions (continued)

<table>
<thead>
<tr>
<th>Key Question</th>
<th>Strength of Evidence</th>
<th>Summary, Comments, and Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>2b. What are the adverse effects or harms associated with testing using newer and/or older markers of iron status?</td>
<td>Insufficient</td>
<td>• Only 7 of the 27 studies reported information: &lt;br&gt;  ◦ 3 studies reported no adverse events associated with iron therapy during the study periods. &lt;br&gt;  ◦ A total of 5 deaths reported. Studies did not attribute these deaths to either testing or any treatment. &lt;br&gt;  ◦ Most of the reported harms were attributed to iron therapy.</td>
</tr>
<tr>
<td>Key Question 3. What is the impact of managing iron status based on newer laboratory biomarkers either alone or in addition to older laboratory biomarkers on intermediate outcomes?</td>
<td>Low</td>
<td>• Two short-term RCTs (4 and 6 months) showed a reduction in the number of iron status tests and resulting intravenous iron treatments (a post-hoc intermediate outcome) administered to patients whose iron management was guided by CHr compared with those guided by TSAT or ferritin. &lt;br&gt;  • Both RCTs reported that Hct remained in the targeted ranges (an indication for the adequacy of anemia management) throughout the study period in all randomized arms, although the Hct target differed between the two trials. &lt;br&gt;  • One trial showed that guiding iron management via CHr resulted in similar epoetin dosing compared with iron management guided by ferritin or TSAT. In contrast, the other trial found the doses of epoetin were significantly decreased (lower by 36 percent) in the group guided by TSAT, but did not change significantly in the group guided by CHr. &lt;br&gt;  • No study compared iron management guided by classical markers with that of newer markers (%HYPO, sTfR, Ret-He, ZPP, or hepcidin).</td>
</tr>
<tr>
<td>3a. What are the adverse effects or harms associated with the treatments guided by tests of iron status?</td>
<td>Insufficient</td>
<td>• Only 1 RCT explicitly monitored the adverse events: &lt;br&gt;  ◦ There were a total of three deaths (2 patients in the CHr group; 1 patient in the TSAT group) due to bacterial pneumonia (at week 4 in the CHr group), sudden death by unknown cause (at week 16 in the CHr group), and liver tumor (at week 7 in the TSAT group). &lt;br&gt;  ◦ One patient in the TSAT group dropped out because of massive bleeding due to a femoral bone fracture and need for blood transfusion. &lt;br&gt;  ◦ There were no significant differences in the hospitalization or infection rates of the two iron management groups.</td>
</tr>
<tr>
<td>Key Question 4. What factors affect the test performance and clinical utility of newer markers of iron status?</td>
<td>Insufficient</td>
<td>• Only single study or indirect comparisons across studies provided data on the potential impacts of some factors on the test performance of newer or classical laboratory markers of iron status: &lt;br&gt;  ◦ One RCT found an interaction between iron and ESA treatment on test accuracy of CHr. A higher baseline CHr predicted greater likelihood of a response to anemia and iron treatment only in the IV iron (plus epoetin) treatment group, but not in the no IV iron (epoetin only) treatment group. &lt;br&gt;  ◦ One study showed that the test accuracy of RetHe was lower for assessing “functional iron deficiency” (TSAT&lt;20%, ferritin 100-800 ng/mL, and Hb &lt;11 g/dL) than for assessing “traditional parameters for iron deficiency” (serum iron &lt; 40 µg/dL, TSAT&lt;20%, ferritin &lt;100 ng/mL, and Hb &lt;11 g/dL) in HD CKD patients. &lt;br&gt;  ◦ Indirect comparisons across studies suggested potential impacts of route of iron administration and treatment regimen on the test accuracy of newer and classical laboratory markers of iron status. &lt;br&gt;  • No study performed analyses by patient subgroups. &lt;br&gt;  • No study examined the impacts of biological variation or type of dialysis in diagnostic indices on the test performance or clinical utility of laboratory markers of iron status.</td>
</tr>
</tbody>
</table>

CHr=reticulocytes hemoglobin content; CKD=chronic kidney disease; Hb=hemoglobin; Hct=hematocrit; HD=hemodialysis; IV=intravenous; PD=peritoneal dialysis; RetHe=reticulocyte hemoglobin equivalent; sTfR= soluble transferrin receptor; TSAT=transferrin saturation; ZPP=erythrocyte zinc protoporphyrin
We synthesized 27 studies to answer Key Question 2 (test performance of newer markers compared with the older markers of iron status), of which 12 evaluated the test performance of newer or classical laboratory markers of iron status in predicting a response to intravenous iron treatment. Most studies enrolled only adult HD CKD patients, though a few examined adult PD and ND CKD patients. Only one study enrolled pediatric CKD patients. Although the reviewed studies evaluated many newer markers, such as CHr, %HYPO, RetHe, sTfR, hepcidin, and ZPP, the majority assessed CHr or %HYPO among adult HD CKD patients.

Based on our analysis, we concluded that there is a low level of evidence that both CHr and %HYPO have a similar or better overall test accuracy compared with classical markers (TSAT or ferritin) to predict a response to IV iron treatment (as the reference standard for iron deficiency). In addition, data suggest that CHr (with cutoff values of <27 or <28 pg) and %HYPO (with cutoff values of >6% or >10%) have better sensitivities and specificities to predict “functional” iron deficiency than classical markers (TSAT <20 or ferritin <100 ng/mL). Across studies, there exists a high degree of heterogeneity in the test comparisons, definitions for the reference standard (a response to IV iron treatment), iron status of the study populations (assessed by TSAT or ferritin), and background treatment across studies. This heterogeneity may limit the comparability of findings across studies.

A response to IV iron treatment is considered by many clinicians as the reference method for diagnosing iron deficiency. The most commonly used definition for a response to IV iron treatment was an increase in hemoglobin (Hb) concentration ≥1 g/dL; however, a consensus does not yet exist. We found no uniform regimen of IV iron in terms of dosage, duration, or iron formulation across these studies. The potential effects of IV iron treatment regimen on the test performance of newer or classical laboratory markers of iron status remain unknown.

For Key Question 3 (the impact of managing iron status based on newer laboratory biomarkers, either alone or in addition to older laboratory biomarkers, on intermediate outcomes compared with managing iron status based on older laboratory biomarkers alone), we identified only two short-term RCTs (4 and 6 months), enrolling a total of 354 adult HD CKD patients. Although both (one conducted in the U.S. and one in Japan) compared intermediate outcomes of iron management guided by CHr with those of iron management guided by classical markers of iron status (TSAT and/or ferritin), the two RCTs employed different protocols for initiating intravenous iron therapy and anemia management, which affect the applicability of their findings.

We concluded that there is a low level of evidence for a reduction in the number of iron status tests and resulting intravenous iron treatments (a post-hoc intermediate outcome) administered to patients whose iron management was guided by CHr compared with those guided by TSAT or ferritin. Both RCTs reported that Hct remained in the targeted ranges (an indication for the adequacy of anemia management) throughout the study period in all randomized arms, though the Hct target was higher in the U.S. trial than the Japanese trial. Only the Japanese trial specifically monitored the adverse events associated with study medication; no differences in the hospitalization or infection rates between the two iron management groups were reported.

For Key Question 4 (factors affecting the performance or clinical utility of newer markers of iron status), we included three studies (1 RCT and 2 prospective cohorts) as well as relevant data from all 27 studies included in Key Questions 2. Nevertheless, we found insufficient evidence to draw any conclusions, as only single studies or indirect comparisons across studies provided data on the potential impacts of some factors (i.e., interactions between iron and ESA treatment, route
of iron administration, and treatment regimen) on the test performance of newer or classical laboratory markers of iron status.

Findings in Relationship to What Is Already Known

Our findings are consistent with the recommendations in the Kidney Disease Outcome Quality Initiative (KDOQI) and the National Institute for Health and Clinical Excellence (NICE) guidelines for anemia management in CKD. The guidelines recommend that the initial assessment of iron deficiency anemia include ferritin to assess iron stores, and serum TSAT or CHr (KDOQI) or %HYPO (NICE) to assess adequacy of iron for erythropoiesis. We found that there is a low level of evidence that both CHr and %HYPO have a similar or better overall test accuracy compared with classical markers (TSAT or ferritin) to predict a response to IV iron treatment among HD CKD patients. In addition, data suggest that CHr (with cutoff values of <27 or <28 pg) and %HYPO (with cutoff values of >6% or >10%) have better sensitivities and specificities to predict “functional” iron deficiency than classical markers (TSAT <20% or ferritin <100 ng/mL). Together, these findings suggest that CHr or %HYPO can be used to monitor iron deficiency in place of the classical markers among HD CKD patients receiving erythropoietin.

Our confidence in the totality of evidence, however, was limited by a high degree of heterogeneity and the large potential risk of bias in the body of literature (see “Limitation of the Evidence Base” for more details). Many important questions remain unanswered, such as the test performance of newer markers of iron status as an add-on to older markers, and the factors that might affect the test performance or clinical utility of laboratory markers of iron status.

We identified one study showing an improvement in test performance by using a combination of laboratory biomarkers, such as % HYPO >6 with TSAT ≤20%, %HYPO >6% with CHr ≤29 pg, and % HYPO >6 with ZPP >52 µmol/mol. However, there are potentially endless test combinations to be evaluated, and without a widely accepted definition of reference standard for the diagnosis of iron deficiency in the context of CKD, new studies are unlikely to significantly contribute to what is already known, or change existing clinical practice.

Applicability and Implications for Clinical and Policy Decisionmaking

We assessed the applicability of the included studies by organizing them according to each patient population of interest, that is, nondialysis patients with stage 3, 4, or 5 CKD, patients with CKD undergoing HD or PD, or patients with a kidney transplant. We evaluated studies of pediatric, adult, and elderly adults separately. Among all the studies included in our review, not one enrolled patients with a kidney transplant or elderly adults exclusively. Only one small study enrolled pediatric CKD patients (16 pediatric PD CKD patients and 11 pediatric HD CKD patients; both groups were analyzed separately). A majority of this review’s findings are thus applicable to only adult HD CKD patients.

The available data are limited due to a high degree of heterogeneity, and are at high risk of bias, limiting their utility in informing clinical practice. However, some clinical implications can be drawn.

We identified two RCTs that compared intermediate outcomes of iron management guided by CHr with those of iron management guided by classical markers of iron status (TSAT and/or ferritin). These two trials (one conducted in the U.S. and one in Japan) employed different
protocols for initiating IV iron therapy and anemia management. Specifically, the epoetin dose adjustment schedule was more frequent in the U.S. trial (every 2 weeks)\textsuperscript{70} as compared with the Japanese trial (twice per month, 3 days after the previous HD therapy).\textsuperscript{66} The U.S. trial also used much higher doses of epoetin than the Japanese trial at baseline (12,232 vs. 4,121 IU/week), and the Hct target was higher as well (between 33 and 36 percent, and 29.5 and 32.5 percent, respectively). The protocols for initiating IV iron therapy also differed between the two trials. These differences may reflect disparities in the healthcare systems of their respective countries, and should be considered as part of clinical decisionmaking.

Also worth noting is a trend in decreasing ESA doses and Hb levels in the U.S.\textsuperscript{71} This trend began after the publication of the Correction of Hemoglobin and Outcomes in Renal Insufficiency (CHOIR) study in 2006, which demonstrated increased mortality among patients treated with ESAs to higher target Hb levels.\textsuperscript{72} None of the included studies in the U.S. were conducted after 2006, which may indicate that the literature does not reflect the current paradigm of anemia management.

Considering our findings with respect to test performance of newer markers versus classical markers together, we can conclude that no single test (using either newer or classical markers) was adequate to determine iron status. Most studies did not show adequate predictive ability (defined as LR+ ≥10 and LR- ≤10) of the marker’s test result (Figures 4 and 5). Classical markers of iron status (ferritin and TSAT) are widely available but have poor sensitivity and specificity. On the other hand, although CHr and %HYPO may have better test performance, neither test is widely available. It should also be noted that test results are invalid for %HYPO when blood samples are stored, as sample storage causes RBC swelling and an incorrect estimation of hypochromic RBCs. This drawback can be prevented by assessing %HYPO immediately after the blood draw. In this context, the site of the blood draw has to be attached to the laboratory setting. This limitation should be weighed when considering the use of %HYPO for assessing iron status.

**Limitations of the Evidence Base**

The available data are very limited due to a high degree of heterogeneity. Many definitions of a response to IV iron treatment as the reference standard for iron deficiency were used across studies. Moreover, there is a lack of a uniform regimen of intravenous iron treatment in terms of dosage, iron formulation, treatment frequency, and followup duration for the iron challenge test (to define a response) across studies.

Many studies included in our review were rated as being at a high risk of bias, limiting their utility in informing clinical practice. Detailed quality appraisals of the included studies are described in Appendix E. In brief, because the demographic details of study populations, including racial breakdown and comorbid conditions, were often not reported, there are potentially several types of biases in the included studies. For example, selection bias could occur if patients were not recruited consecutively. A related source of bias in this context is spectrum bias, in which the reported sensitivity and specificity may be exaggerated in populations with increased disease severity. Incorporation bias is often difficult to eliminate, because the result from the index test is used to determine who will receive iron treatment. Some measures recommended to maximize the quality of test interpretation include repeat testing, targeted followup of false positives, and blinding of the diagnosis or test group to diminish the likelihood of misclassification bias. Such safeguards, however, were not reported in the reviewed studies.
Research Gaps

The most directly applicable study designs for clinical decisionmaking would be studies that compare two or more iron and anemia management strategies, follow the patients through decisions and treatments, and then report on patient outcomes. However, none of the comparative studies identified in this review were of such a design. In truth, it is unlikely such studies can be conducted, due to the high patient and resource requirements. Typically, the assessment of diagnostic tests typically follows the Fryback approach,73 progressing from the establishment of technical and clinical validity, to the assessment of test impact on clinicians’ diagnostic thinking and therapeutic decisionmaking, as well as clinical outcomes. Finally, a global assessment of the test from a societal perspective can be performed. Thus, we suggest that future research address the gaps that we identified for each of the component questions in this review. We also identified several cross-cutting methodological issues that affected all of the Key Questions, and that should be addressed. Ultimately, when a reference standard of iron deficiency is finally established, and test performance data are sufficient and reliable, decision analysis could be used to assess how using combinations of different markers to guide iron management strategies might influence clinical outcomes.

A summary of the research gaps we identified, as well as our suggestions for future research, are provided in Table 22.
Table 22. Research gaps and suggestions for future research

<table>
<thead>
<tr>
<th>Key Question</th>
<th>Research Gaps</th>
<th>Suggestions for Future Research</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Key Question 2.</strong> What is the diagnostic test accuracy of newer markers of iron status as a replacement for or an add-on to classical laboratory markers?</td>
<td>Insufficient evidence for the test performance of newer markers of iron status as an add-on to older markers</td>
<td>• It is important to use an independent reference standard when assessing the test performance. See “Cross-cutting issues” for the research gaps for establishing a reference standard for iron deficiency.</td>
</tr>
<tr>
<td></td>
<td>Many existing studies are at a high risk of bias, limiting their utility in informing clinical practice</td>
<td>• General principles for the design of studies of diagnostic tests include the use of an appropriate reference standard, adequate description of the index and reference tests, blinded interpretation of test results, and independence of the index and reference standard tests.(^7)</td>
</tr>
<tr>
<td></td>
<td>There is no uniform iron management algorithms across studies</td>
<td>• Studies assessing diagnostic accuracy should instead aim to enroll patients representative of the spectrum of disease typically seen in clinical practice.</td>
</tr>
<tr>
<td></td>
<td>Insufficient evidence to draw conclusions regarding factors that may affect the test performance or clinical utility of laboratory markers of iron status</td>
<td>• Future observational studies should assess the outcomes of different iron management algorithms or test-and-treat protocols, considering differences in CKD populations, clinical settings, and potential harms or burden to the patients.</td>
</tr>
<tr>
<td></td>
<td>Whether test performance and clinical utility of newer or classical markers of iron status vary by different CKD populations are not known</td>
<td>• Assessing impact of the most promising iron management algorithms on both intermediate and patient outcomes through prospective observational studies or RCTs.</td>
</tr>
</tbody>
</table>
| | Almost all existing studies enrolled only single CKD population (ND, HD, or PD CKD patients). Future studies should include wider CKD populations, and plan for subgroup analyses. | • Future studies are need to evaluated the following factors, suggested by the experts:  
  o Biological variation in diagnostic indices  
  o Use of different diagnostic reference standards  
  o Type of dialysis (i.e., peritoneal or hemodialysis)  
  o Patient subgroups (i.e., age, sex, comorbid conditions, erythropoiesis-stimulating agent resistance, protein energy malnutrition secondary to an inflammatory state, hemoglobinopathies [e.g., thalassemia and sickle cell anemia])  
  o Route of iron administration (i.e., oral or intravenous)  
  o Treatment regimen (i.e., repletion or continuous treatment)  
  o Interactions between treatments (i.e., patients treated with versus without ESA, patients treated with versus without iron-replacement therapy) |

66
Table 22. Research gaps and suggestions for future research (continued)

<table>
<thead>
<tr>
<th>Key Question</th>
<th>Research Gaps</th>
<th>Suggestions for Future Research</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross-cutting issues (for Key Question 2, 3, and 4)</td>
<td>There is no reference standard for determining iron deficiency in CKD patients</td>
<td>• A response to IV iron treatment is considered by many clinicians as the reference method for diagnosing iron deficiency but future research is needed to establish a standardized definition for appropriate CKD populations, and a standardized testing protocol specifying the regimen of IV iron challenge in terms of dosage and iron formulation and proper duration of iron challenge testing.</td>
</tr>
<tr>
<td></td>
<td>Existing studies were underpowered leading to imprecise estimates</td>
<td>• Future studies should be larger, ideally designed based on power calculations, to be able to reliably detect plausible effect sizes and provide precise estimates of diagnostic accuracy.</td>
</tr>
</tbody>
</table>
| | There is no decision analysis to assess how using combinations of different markers to guide iron management strategies might influence clinical outcomes | • Patient outcomes of interest are:  
  - Mortality  
  - Morbidity (e.g., cardiac or liver toxicity and infection)  
  - Quality of life, measured using standardized scales, including: Kidney Disease Quality of Life (KDQOL), Health Related Quality of Life (HRQOL), Medical Outcomes Study Short Form-36 (SF-36), and Pediatric Quality of Life Inventory (PQLI)  
  - Adverse effects or harms associated with testing and associated treatments (e.g., test-related anxiety, adverse events secondary to venipuncture, effects of iron overload with iron treatments, and cardiovascular complications from use of erythropoietin at higher Hb levels)  
• For studies assessing clinical outcomes, blinding to test results to the outcome assessors is essential to avoid bias. |

CKD=chronic kidney disease; HD=hemodialysis; IV=intravenous; ND=nondialysis; PD=peritoneal dialysis; RCT=randomized controlled trial

Conclusions

Combining the evidence addressing Key Questions 2, 3, and 4, we can conclude that all currently available laboratory biomarkers of iron status (either newer or classical markers) do not have an ideal predictive ability when used singly to determine iron deficiency as defined by a response to iron challenge test. Furthermore, there is insufficient evidence to determine the test performance of the combinations of newer biomarkers, or combinations of newer and classical biomarkers, for diagnosing iron deficiency. However, it may be that CHr and %HYPO have better predictive ability for a response to IV iron treatment than classical markers (TSAT <20 or ferritin <100 ng/mL) in HD CKD patients. In addition, results from two RCTs showed a reduction in the number of iron status tests and resulting IV iron treatments administered to patients whose iron management was guided by CHr compared with those guided by TSAT or ferritin. These results suggest that CHr may reduce potential harms from IV iron treatment by lowering the frequency of iron testing, although the evidence for the potential harms associated with testing or test-associated treatment is insufficient.

Nevertheless, the strength of evidence supporting these conclusions is low and there remains considerable clinical uncertainty regarding the use of newer markers in the assessment of iron status and management of iron deficiency in stages 3–5 CKD patients (both nondialysis and dialysis). In addition, factors that may affect the test performance and clinical utility of newer laboratory markers of iron status remain largely unexamined.
References


Abbreviations and Acronyms

%HYPO Percentage of Hypochromic Erythrocytes
AAAC American Association for Clinical Chemistry
AHRQ Agency for Healthcare Research and Quality
AUC Area Under the Curve
CHr Hemoglobin Content of Reticulocytes
CI Confidence Interval
CKD Chronic Kidney Disease
EPC Evidence-based Practice Center
ESA Erythropoiesis Stimulating Agents
FDA Food and Drug Administration
Hb Hemoglobin
Hct Hematocrit
HD Hemodialysis
IV Intravenous
KDOQI Kidney Disease Outcome Quality Initiative
ND CKD Nondialysis Chronic Kidney Disease
NKF National Kidney Foundation
PD Peritoneal dialysis
PICO Populations, interventions, comparators, and outcomes
PRISMA Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RBC Red Blood Cell
RCT Randomized Controlled Trial
Ret He Reticulocyte Hemoglobin Equivalent
ROC Receiver Operating Characteristic
SQUID Superconducting Quantum Interference Devices
sTfR soluble Transferrin Receptor
TEP Technical Expert Panel
TOO Task Order Officer
TSAT Transferrin Saturation
ZPP Erythrocyte Zinc Protoporphyrin
Appendix A. Literature Search Strategy

Database(s):  Ovid MEDLINE(R) 1948 to July Week 2 2011, Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations July 21, 2011, EBM Reviews—Cochrane Central Register of Controlled Trials 3rd Quarter 2011

1)  percentage of hypochromic erythrocytes.mp. [mp=ps, rs, ti, ot, ab, nm, hw, ui, sh, kw] (15)
2)  (percentage adj2 hypochromic erythrocytes).tw. (15)
3)  *Reticulocytes/pa [Pathology] (77)
4)  (%HYPO and CHr).mp. [mp=ps, rs, ti, ot, ab, nm, hw, ui, sh, kw] (11)
5)  *Erythrocyte Indices/ (783)
6)  Erythrocyte Count.mp. [mp=ps, rs, ti, ot, ab, nm, hw, ui, sh, kw] (10128)
7)  *Erythrocytes/an, du, me [Analysis, Diagnostic Use, Metabolism] (18919)
8)  *Erythrocytes/pa (994)
9)  Erythropoiesis/ph (1271)
10)  *Reticulocytes/ch (65)
11)  *Reticulocyte Count/ (156)
12)  Ferritins/bl (7434)
13)  *Hemoglobins/an (5118)
14)  *Erythrocyte Indices/ (783)
15)  Reticulocytes/me (4605)
16)  Transferrin/an (3892)
17)  TSAT.tw. (221)
18)  exp Anemia, Hypochromic/ or exp Anemia, Iron-Deficiency/ (15078)
19)  hypochromic an?emia.mp. (802)
20)  hypochromic erythrocytes.mp. (40)
21)  Anemia, Iron-Deficiency/dt (1439)
22)  (transferrin adj saturation).af. (2614)
23)  exp transferrin/ (14770)
24)  hepcidin.af. (1413)
25)  ((soluble or serum) adj transferrin).af. (3648)
26)  ((soluble or serum) adj transferrin adj receptor).af. (822)
27)  (zinc adj protoporhyrin).af. (5)
28)  erythrocyte zinc protoporhyrin.mp. (1)
29)  superconducting quantum interference device.mp. (345)
30)  *Biological Markers/an, bl, me [Analysis, Blood, Metabolism] (11098)
31)  *ferritins/ or *apoferritins/ (6927)
32)  acute-phase proteins/ or exp transferrin/ (19169)
33)  transferrin.mp. (28365)
34)  Transferrin/ad, an, bl, du, de, me, pk, tu [Administration & Dosage, Analysis, Blood, Diagnostic Use, Drug Effects, Metabolism, Pharmacokinetics, Therapeutic Use] (10637)
35)  or/1-34 (104459)
36)  exp "sensitivity and specificity"/ (347657)
37)  exp Predictive Value of Tests/ (116486)
38)  exp ROC CURVE/ (20708)
39)  exp Mass Screening/ (88916)
40)  exp diagnosis/ (5609927)
41)  exp REPRODUCIBILITY OF RESULTS/ (224874)
42)  exp false negative reactions/ or false positive reactions/ (31445)
43)  predictive value.tw. (47133)
44)  (sensitivity or specificity).tw. (616885)
45)  accuracy.tw. (179753)
46)  screen$.tw. (378934)
47)  diagno$.tw. (1399885)
48)  roc.tw. (15058)
49)  reproducib$.tw. (95276)
50)  (false positive or false negative).tw. (41652)
51)  likelihood ratio.tw. (4915)
52)  accuracy.tw. (179753)
53)  di.fs. (1755561)
54)  biological variability.mp. (700)
55)  reference values.tw. (8341)
56)  reference standard$.tw. (6786)
57)  or/36-56 (7398453)
58)  (NeoRecormon or Aranesp or Methoxy Polyethylene Glycol Epoetin Beta or MIRCERA or Epoetin or Dynepo or PDpoetin).af. or (NeoRecormon or Aranesp or Methoxy Polyethylene Glycol Epoetin Beta or MIRCERA or Epoetin or Dynepo or PDpoetin).tw. (2810)
59)  (epogen or epotin or betapoietin or relpoietin or epokine or procrict or eprex or darbopoietin).af. or (epogen or epotin or betapoietin or relpoietin or epokine or procrict or eprex or darbopoietin).tw. (252)
60)  (Ferumoxytol or Feraheme or Iron Dextran or DexFerrum or INFeD or Ferrous fumarate or ferrous gluconate or ferrous sulfate or ferrous sulphate or carbonyl iron or polysaccharide iron complex or Icar or Feosol or Ircon or Hemocyte or Nephro-Fer or Feostat or Ferro-DSS or Ferro-Sequels or Fergon or Fer-Gen-Sol or Fer-In-Sol or Mol-Iron or Feratab or Ferrex or Niferex or Hytinic or Fe-Tinic or Iron sucrose or Venofer or Sodium ferric gluconate or Ferrlecit).mp. or exp Iron/ or exp Ferric Compounds/ or exp Ferrous Compounds/ or exp Anemia, Iron-Deficiency/ or ferrous.mp. (105161)
61)  exp recombinant erythropoietin/ or recombinant erythropoietin.mp. (6065)
62)  or/58-61 (111284)
63)  exp Renal Replacement Therapy/ or exp Renal Dialysis/ or exp Kidney Transplantation/ or exp Kidney Function Tests/ or renal.mp. or nephro$.mp. or kidney.mp. or ur?emia.tw. or h?emodialysis.tw. (827443)
64)  hemodialysis.af. (45683)
65)  peritoneal dialysis.mp. or exp Peritoneal Dialysis/ (24886)
66)  exp Kidney Diseases/ or exp Kidney Failure, Chronic/ or chronic kidney disease.mp. or exp Chronic Disease/ or exp Kidney Glomerulus/ (584160)
67)  or/63-66 (1060229)
68)  35 and 57 and 67 (5403)
69)  remove duplicates from 68 (5089)
70)  35 and 62 and 67 (2619)
71)  remove duplicates from 70 (2409)
72)  69 and 71 (1287)
73)  69 not 72 (3802)
74)  73 or 71 (6211)
75)  limit 74 to (addresses or bibliography or biography or case reports or comment or congresses or consensus development conference or dictionary or directory or festschrift or in vitro or interactive tutorial or interview or lectures or legal cases or legislation or news or newspaper article or overall or patient education handout or periodical index or portraits or "scientific integrity review" or twin study) [Limit not valid in CCTR; records were retained] (458)
76)  74 not 75 (5753)
Appendix B. Excluded Studies

Of the 694 articles obtained for full-text screening, 30 were included and 664 were excluded on double, independent full-text screening because they did not meet one or more of the PICO criteria for a particular key question. The two most common reasons for rejection were: (1) no diagnostic outcomes reported (studies reported only correlations between markers or the measurements of levels of markers before and after treatment); (2) no comparative data for the outcomes of management strategies where treatment decisions were guided by test results (newer versus classical markers). The 664 excluded references are listed below, in alphabetic order of first author’s surname, along with the reason for exclusion for each.

1) [No authors listed]. Serum ferritin concentrations after intravenous iron-dextran. Lancet. 1(8279):1017-8, 1982 May 1. PMID: 6122828 Not biomarkers of interest—older markers only (KQ2)


6) Agarwal R, Davis IL, Hamburger RJ. A trial of two iron-dextran infusion regimens in chronic hemodialysis patients. Clin Nephrol. 54(2):105-11, 2000 Aug. PMID: 10968685 Only older markers studied in terms of analytic validity data


11) Aggarwal HK, Tziviskou E, Bellizzi V, Khandelwal M, Moupas L, Bargman JM, Jassal SV, Oreopoulos DG. Prolonged administration over six hours of large doses of intravenous iron saccharate (500 mg) prevents severe adverse reactions in peritoneal dialysis patients. Perit Dial Int. 22(5):636-7, 2002 Sep-Oct. PMID: 12455582 Treatment no based on biomarker

N<10, Single arm treatment cohort, analytic validity data for older markers only

Analytic validity data for only older markers studied

Comparison of Different cutoffs

Treatment based on single marker and analytic validity data for only older markers studied

not test-directed Tx (KQ3)

Article in Russian; appears to be analytic validity data only; the google translate was not clear;

Analytic validity data for only older markers studied

Analytic validity data only for older markers

Analytic validity data only for older markers

Analytic validity data for older markers only

Not biomarkers of interest—older markers only & Analytical validity data only (KQ2)

Article in Italian; Not biomarkers of interest—older markers only & Analytical validity data only (KQ2)

Analytic validity data for older markers only

Single arm Tx cohort (KQ3)

Analytic validity data for only older markers studied


41) Ashby DR, Gale DP, Busbridge M, Murphy KG, Duncan ND, Cairns TD, Taube DH, Bloom SR, Chapman RS, Maxwell PH, Choi P. Plasma hepcidin levels are elevated but responsive to erythropoietin therapy in renal disease. Kidney Int. 75(9):976-81, 2009 May. PMID: 19212416


No comparison between biomarkers and treatment is fixed.


Treatment not based on markers and only one biomarker studied with no comparison


Sensitivity and specificity comparison only between older markers


Single arm Tx cohort (KQ3)


Single arm treatment cohort and analytic validity data for older markers only


<10 pts with CKD in ACD group


Single marker studied and treatment is not based on marker


Only Analytic validity data (KQ2)


Analytic validity data only (KQ2)


Analytic validity data only (KQ2)

53) Barracloough KA,Noble E,Leary D,Brown F,Hawley CM,Campbell SB,Isbel NM,Mudge DW,van Eps CL,Sturtevant JM,Johnson DW. Rationale and design of the oral HEMe iron polypeptide Against Treatment with Oral Controlled Release Iron Tablets trial for the correction of anaemia in peritoneal dialysis patients (HEMATOCRIT trial). BMC Nephrol. 10:20, 2009. PMID: 19635169

Not test-directed treatment, Tx was fixed


Analytic validity data only (for KQ2)


Non CKD patients.


Single arm Tx cohort (KQ3)


Only older markers studied


Population <10 and analytic validity data for older markers only


Analytic validity data (KQ2)


67) Bezwoda WR, Derman DP, Bothwell TH, MacPail AP, Torrance JD, Milne FJ, Meyers AM, Levin J. Iron absorption in patients on regular dialysis therapy. Nephron. 28(6):289-93, 1981. PMID: 7312083 Not biomarkers of interest—older markers only & Analytical validity data only (KQ2)


69) Bhandari S, Brownjohn AM, Turney JH. Single arm Tx cohort (KQ3)

70) Bhandari S, Owda AK, Kendall RG, Moran N, Norfolk DR, Brownjohn AM, Turney JH. Red cell ferritin, a marker of iron deficiency in hemodialysis patients. Ren Fail. 19(6):771-80, 1997 Nov. PMID: 9415934 Not a biomarker of interest (KQ2)—red cell ferritin is not of interest


75) Blumberg AB, Marti HR, Graber CG. Serum ferritin and bone marrow iron in patients undergoing continuous ambulatory peritoneal dialysis. JAMA. 250(24):3317-9, 1983 Dec 23-30. PMID: 6645029 Analytic validity data for older markers only


Single arm treatment cohort. Only older markers studied in terms of change in mean values.


Analytic validity data only (KQ2)


Not a biomarker of interest


Single arm treatment cohort. Treatment not based on any biomarker


Not test-directed Tx; No Dx information


Analytic validity data only (KQ2)


N<10; No Dx information


Analytic validity data only (KQ2)


Single arm Tx cohort ; No Dx information


No Dx information; Not a biomarker of interest


Not test-directed treatment, Tx was fixed


Analytic validity data only (for KQ2)


Analytic validity data only (KQ2)


Single arm study


Single arm Tx cohort (KQ3); no Dx information


Even though it is test-directed Tx, the tests used to direct Tx are the same in both arms


Non CKD population

94) Brozovich B,Cattell WR,Cottrall MF,Gwyther MM,McMillan JM,Malpas JS,Salsbury A,Trott
Not a biomarker of interest (KQ2)—only older biomarkers evaluated

Not biomarkers of interest—older markers only & Analytical validity data only (KQ2)

Analytical validity data only (KQ2)

Not a diagnostic study

Q&A article about IV iron in PD pts


Single arm Tx cohorts

SQUID is used as a gold standard versus Ferritin; used in evaluating hepatic iron excess not iron deficiency (KQ2)

Analytic validity data (KQ2)

103) Cantaro S,Piva E. [Hematological and iron parameters to predict mortality in ESRD]. [Italian]. NERFROL.. 22 Suppl 31:S135-9, 2005 Jan-Feb. PMID: 15786388
Single arm Tx cohort (KQ3); no Dx information


Not test-directed treatment, Tx was fixed

Not a biomarker of interest; Single arm Tx cohort (KQ3)

Not an intervention of interest—N-carnitine

Not biomarkers of interest—older markers only (KQ2)


Only one biomarker studied

Only older markers compared in terms of analytic validity.

Non-CKD patients

Only older markers studied in terms of mean values across 2 treatment groups

Dx information for older markers only; Single arm Tx cohort (KQ3)

Treatment not based on markers and analytic validity data for only older studied

TfR-ferritin Index is a combination of a new and old marker

Analytic validity data only (KQ2)

Analytic validity looked at for KQ2

TfR-ferritin Index is a combination of a new and old marker

Non CKD-patients

No Dx information; Single arm Tx cohort (KQ3)

Non CKD population and PPV for older markers only

Biomarkers assessed in terms of different ranges of hematocrit values without any comparison with each other.

126) Corazza F, Bergmann P, Dratwa M, Guns M, Fondu P. Responsiveness to recombinant erythropoietin therapy in end-stage renal disease. An analysis of
the predictive value of several biological measurements, including circulating erythroid progenitors. Nephrol Dial Transplant. 7(4):311-7, 1992. PMID: 1317521

Only older markers studied in terms of mean values as per response to treatment


Analytical validity data only (KQ2)


Analytic validity data only (KQ2)


Single arm Tx cohort (KQ3)

130) Coyne D. Challenging the boundaries of anemia management: a balanced approach to i.v. iron and EPO therapy. Kidney Int Suppl. (101):S1-3, 2006 May. PMID: 16830698

Narrative review


Not test-directed Tx; Tx remained the same in both arms

133) Coyne DW, Sims A, Bingel B. Results of an anemia management program to reduce high epoetin doses by targeted use of i.v. ferric gluconate. Nephrol Nurs J. 35(6):583-7, 2008 Nov-Dec. PMID: 19260610

Single arm Tx cohort; No Dx information


Not test-directed Tx (KQ3)


Analytic validity data only (for KQ2)


No biomarker studied


Single arm treatment cohort and only one marker studied


N<10 in one treatment arm and analytic validity data only (KQ2)


n Dx information; Single arm Tx cohort (KQ3)

Treatment fixed and analytic validity data only for one newer marker

Single arm Tx cohorts; N<10 (KQ3)

Not test-directed Tx (KQ3)

not test-directed Tx

146) Deaver K, Bennington L. Adjusting i.v. iron and EPO doses in patients on hemodialysis prior to continuing surgery: can we protect our patients education from iron-deficiency anemia?. Nephrol Nurs J. 33(4):430-7; quiz 438, 2006 Jul-Aug. PMID: 17002001
Narrative review

n<10 in one arm

Analytic validity data for older markers only

Only one biomarker present

N<10, single arm treatment cohort and analytic validity data for older markers only

Single arm Tx cohort (KQ3)

n Dx information; Single arm Tx cohort (KQ3)

Treatment is fixed and analytic validity data only (KQ2)

N<10 and analytic validity data only for older markers

Not test-directed Tx based on our tests of interest

Treatment is fixed and analytic validity data (KQ2)

Treatment not based on biomarkers. Only older markers studied in terms of analytic validity data

Treatment is not based on biomarker and no biomarker comparison present

---


Only one biomarker of interest studied


Marker prohepcidin is not enlisted in the protocol and rest analytic data for older markers only.

El-Khatib M, Duncan HJ, Kant KS. Role of C-reactive protein, reticulocyte haemoglobin content and inflammatory markers in iron and erythropoietin administration in dialysis patients. Nephrology. 11(5):400-4, 2006 Oct. PMID: 17014552

No Dx data


Not a biomarker of interest (KQ2)—only older biomarkers evaluated


Analytic validity data for only older markers studied


N<10 in one group and only older markers studied in terms of mean values in different groups


Not a biomarker of interest (KQ2)—only older biomarkers evaluated; Analytic validity data only (KQ2)


Analytic validity data for only older markers studied


No comparison of biomarkers and treatment not based on markers


Only one biomarker of interest is studied and treatment is also depended on same marker


Not biomarkers of interest—older markers only & Analytical validity data only (KQ2)


Non CKD population


n Dx information; Single arm Tx cohort—postmarket analysis of IV iron (KQ3)


Single arm Tx cohort; No Dx information


Single arm Tx cohort (KQ3)


Single arm Tx cohort (KQ3)

Fernandez-Gallego J, Martin MA, Sujan S, Vega E. [Treatment with intravenous iron and ferritin...

176) Fernandez-Gallego J,Ramos B,Ruiz A,Contreras
J,Alvarez BG,Lopez de NE. [Study of various factors that could have an impact on the treatment with erythropoietin of hemodialysis anemia]. [Spanish]. Nefrologia. 20(2):164-70, 2000 Mar-Apr. PMID: 10853198


Variants in transferrin studied

Single arm Tx cohort (KQ3)


Single arm Tx cohort (KQ3)


Not test-directed Tx


Analytic validity data for only older markers studied


Not test-directed Tx


Single arm study. Change in levels of older markers studied.


Non CKD patients. Analytic validity data (KQ2)


Not a biomarker of interest; CHr was used as the reference gold-standard test. Newer marker Ret HE


No comparison between biomarkers. Mean values of only older markers studied in two different groups of treatment


No comparison between biomakers


N<10 and analytic validity data only for older markers


Single arm treatment cohort


Mean values of older markers studied


Analytic validity data only (KQ2). New marker HRF is used


N<10; single arm treatment cohort and analytic validity data for older markers only


Analytic validity data for older markers only

Only one biomarker studied.

211) Gotloib L, Silverberg D, Fudin R, Shostak A. Iron deficiency is a common cause of anemia in chronic kidney disease and can often be corrected with intravenous iron. JN, J. nephrol.. 19(2):161-7, 2006 Mar-Apr. PMID: 16736414

Analytic validity data only for older markers and single arm treatment cohort.


Changes in mean values of older markers only studied over the course of therapy and treatment not based on markers.


No biomarker studied and treatment is fixed.


Analytic validity data only for older markers.


Not biomarkers of interest—older markers only.


Single arm treatment cohort and no comparison between markers.


Comparison of mean values of older markers only pre and post treatment.


Analytical validity data only (KQ2).


Mean values of older markers measured before and after treatment.


Analytic validity data for older markers only.


Anemia prevalence study.


Changes in mean values of older markers only studied over the course of therapy and treatment not based on markers.


Language other than English.

Not biomarkers of interest—older biomarkers only

Single arm Tx cohort (KQ3); No Dx information

Single arm treatment cohort. Change in mean values of only older markers studied

Single arm Tx cohort (KQ3); No Dx information

Single arm study. Changes in mean levels of only older markers studied

Analytical validity data only (KQ2)

Analytic validity data for only older markers studied.

No comparison between biomarkers. Just the data gathered at baseline

Only one biomarker of interest studied in terms of analytic validity. Also a newer marker erythrocyte ferritin is used.

No biomarker comparison and treatment is not based on biomarker

Relationship of Hb with older markers studied with no comparison between newer versus older.

Biomarker not of interest. Analytic validity data only (KQ2)

Not biomarkers of interest—older markers only & Analytical validity data only (KQ2)

Not biomarkers of interest—older markers only & Analytical validity data only (KQ2)
Not test-directed Tx (KQ3)

Cost effectiveness analysis done

Treatment not based on biomarker

Analytic validity data for older markers only and treatment is fixed.

Only one biomarker studied and treatment is not based on biomarker

Single arm treatment cohort and only one biomarker studied

No comparison between biomarkers and treatment is not based on any biomarker

No Dx information, no test-directed Tx

Not test-directed Tx (KQ3)

Narrative review

Single arm Tx cohorts (KQ3)

Not CKD patients

Analytical validity data only (KQ2)

Case study (n<10)

Single arm treatment cohort and no comparison between biomarkers

256) Javier AM. Weekly administration of high-dose sodium ferric gluconate is safe and effective in peritoneal dialysis patients. Nephrol Nurs J. 29(2):183-6, 2002 Apr. PMID: 11997953
N<10; Not biomarkers of interest—older biomarkers only (KQ2); Not test-directed Tx (KQ3)

Single arm Tx cohort (KQ3); No Dx information

258) Johnson CA, Rosowski E, Zimmerman SW. A prospective open-label study evaluating the efficacy and adverse reactions of the use of Niferex-150 in ESRD patients receiving EPOGEN. Adv Perit Dial. 8:444-7, 1992. PMID:
1361844

Single arm Tx cohort (KQ3); No Dx information


Change in mean values of older markers studied and treatment is not based on biomarker


Narrative Review


No comparison between biomarkers influencing treatments


Analytic validity data only (for KQ2)


Mean values of only older biomarkers studied in different groups


Not biomarkers of interest—older biomarkers only (KQ2); Test-directed Tx (KQ3)


None


None


None


Not biomarkers of interest—older biomarkers only (KQ2); Single arm Tx cohort (KQ3)


Single arm Tx cohort (KQ3)


Only older markers compared in terms of Odds Ratios for predicting risk of death


Not a biomarker of interest; only older markers used

272) Kalantar-Zadeh K, Luft FC, Humphreys MH. Moderately high serum ferritin concentration is not a sign of iron overload in dialysis patients. Kidney Int. 56(2):758-9, 1999 Apr. PMID: 10432420

editorial. Comparison only between 2 older markers in terms of analytic validity data


No Dx information; Not test-directed Tx


N<10; Single arm Tx cohort (KQ3); No Dx information


Not biomarkers of interest—older biomarkers only


Not test-directed Tx (KQ3)


Treatment fixed. Only older markers studied with no comparison with newer ones.

Kapoian T. Sodium ferric gluconate complex reduces epoetin doses in epoetin-resistant hemodialysis patients with elevated ferritin: preliminary results of the DRIVE-II study [abstract no: F-PO671]. No source- cochrain. PMID: 18216316, Title-Ferric gluconate reduces epo requirements in hemodialysis patients with elevated ferritin 2008 -Mean levels of only older markers monitored during treatment and treatment is not dependent on biomarker

Kaskel FJ. Safety and efficacy of sodium ferric gluconate complex in iron-deficient pediatric hemodialysis patients. Nat. clin. pract. nephrol.. 2(5):244-5, 2006 May. PMID: 16932433

Not test-directed Tx (KQ3)


Single arm Tx cohort (KQ3)
Not biomarkers of interest—older markers

Not biomarkers of interest—older markers only

Treatment not based on biomarker

Not biomarkers of interest—older markers only

Single arm Tx cohort (KQ3)

Analytic validity data (KQ2)

Analytical validity data only (KQ2)

Comparison of Different cutoffs

Analytic validity data only (KQ2)

Only older markers compared to each other while looking at their association with mortality.

Not test-directed Tx (KQ3)

301) Krause JR, Stole V. Serum ferritin and bone marrow biopsy iron stores. II. Correlation with low serum iron and Fe/TIBC ratio less than 15%. Am J Clin Pathol. 74(4):461-4, 1980 Oct. PMID: 7424828
Not biomarkers of interest—older markers only (KQ2)

Single arm Tx cohort (KQ3); No Dx information

Not biomarkers of interest—older biomarkers only

Non-CKD patients

Non-CKD patients

Experimental lab study


Analytical validity data only (KQ2); Logistic regression only looks at r-HuEPO independence, not iron deficiency


Single arm Tx cohort (KQ3); Not biomarkers of interest—older markers only (KQ2)


Not test-directed Tx


Single arm treatment cohort. Change in mean values of only older markers studied pre and post treatment


Single arm Tx cohort (KQ3); No Dx information; CKD patients—N<10


Single arm Tx cohort (KQ3); No Dx information

313) Lazarus JM, Hakim RM, Newell J. Recombinant human erythropoietin and phlebotomy in the treatment of iron overload in chronic hemodialysis patients. Am J Kidney Dis. 16(2):101-8, 1990 Aug. PMID: 2382644 N<10; Single arm Tx cohort (KQ3); Not biomarkers of interest—older biomarkers only & Analytical validity data only (KQ2)


Comparison of Different cutoffs


Not test-directed Tx


Treatment fixed and not based on biomarkers. No biomarker comparison

317) Leu MI, Shih LY, Huang MJ. Serum ferritin and bone marrow iron in patient on maintenance hemodialysis. Taiwan I Hsueh Hui Tsai Chih. 80(8):804-14, 1981 Aug. PMID: 6947063

Not biomarkers of interest—older biomarkers only


Only newer biomarker discussed. In vitro study?


Not test-directed Tx (KQ3)


Analytic validity data only for older markers


Changes in mean values of older markers studied pre and post treatment


Treatment not based on biomarker

Change in mean values of only older markers after the treatment is studied


Sensitivity and specificity comparison between older markers only


Changes in mean values of older markers studied pre and post treatment


Single arm Tx cohort (KQ3); No Dx information


Non CKD patients and analytic validity data only between older markers

Lisochenko NM, Idel'son LI. [Comparative evaluation of orferon with other iron preparations in iron deficiency anemias]. [Russian]. Vrach Delo. 5:20-2, 1973 May. PMID: 4757724

Not biomarkers of interest—older markers;


Liscenko NM, Idel'son LI. [Comparative evaluation of orferon with other iron preparations in iron deficiency anemias]. [Russian]. Vrach Delo. 5:20-2, 1973 May. PMID: 4757724

Not biomarkers of interest—older markers;


Not test-directed Tx (KQ3)


Analytic validity data only for older markers


Only one biomarker studied


B-21
PMID: 1739797
Analytic validity data only (for KQ2)

341) Macdougall IC, Chandler G, Elston O, Harchowal J. 
Beneficial effects of adopting an aggressive 
intravenous iron policy in a hemodialysis unit. 
PMID: 10516375
Single arm Tx cohort (KQ3); No Dx information

342) Macdougall IC, Hutton RD, Cavill I, Coles 
GA, Williams JD. Poor response to treatment of 
renal anaemia with erythropoietin corrected by 
iron given intravenously. BMJ. 299(6692):157-8, 
1989 Jul 15. PMID: 2504356
Not test-directed Tx (KQ3); Not biomarkers of 
interest—older markers only

343) Macdougall IC, Tucker B, Thompson J, Tomson 
CR, Baker LR, Raine AE. A randomized controlled 
study of iron supplementation in patients treated 
with erythropoietin. Kidney Int. 50(5):1694-9, 
1996 Nov. PMID: 8914038
Not test-directed Tx (KQ3); Not biomarkers of 
interest—older markers only

344) Macdougall IC. What is the most appropriate 
strategy to monitor functional iron deficiency in 
the dialysed patient on rhEPO therapy? Merits of 
percentage hypochromic red cells as a marker of 
functional iron deficiency. Nephrol Dial 
Transplant. 13(4):847-9, 1998 Apr. PMID: 
9568837
Narrative review

345) Maconi M, Cavalca L, Danise P, Cardarelli F, Brini 
M. Erythrocyte and reticulocyte indices in iron 
deficiency in chronic kidney disease: comparison 
69(3):365-70, 2009. PMID: 19125368
Only newer markers compared

346) Madhan KK, Chamberlain M, Anderson E. 
Anaemia in patients with chronic kidney disease: 
management with epoetin beta in primary care 
setting in New Zealand. Nephrology. 13(5):428- 
32, 2008 Oct. PMID: 18331435
Single arm Tx cohort (KQ3); Not biomarkers of 
interest—older markers only

347) Madore F, Lowrie EG, Brugnara C, Lew 
NL, Lazarus JM, Bridges K, Owen WF. Anemia in 
hemodialysis patients: variables affecting this 
8(12):1921-9, 1997 Dec. PMID: 9402095
Single arm Tx cohort (KQ3); Not biomarkers of 
interest—older markers only

348) Mafra D, Cuppari L, Favaro DI, Cozzolino SM. 
Zinc levels after iron supplementation in patients 
14(3):164-9, 2004 Jul. PMID: 15232795
Analytical validity data only (KQ2); Single arm 
Tx cohort (KQ3)

349) Magid E, Hilden M. Ferrokinetics in patients 
suffering from chronic renal disease and anaemia. 
6031893
Not biomarkers of interest—ferrokinetic 
parameters (KQ2)

350) Majdan M, Ksiązek A, Koziol M, Spasiwiecz 
D, Swatowski A, Solski J. Comparison of plasma 
erthropoietin concentrations and iron status in 
hemodialyzed patients not requiring and requiring 
PMID: 8832602
Comparison of mean values of older markers in 
different groups

351) Malovrh M, Hojs N, Premru V. The influence of 
near-end, continuous, low-dose iron 
replacement on hemoglobin levels in 
hemodialysis patients treated with erythropoiesis- 
Jan. PMID: 20618233
Single arm cohort subject to treatment at two 
different periods but using common markers

352) Malovrh M, Premru V. Subcutaneous compared 
with intravenous epoetin treatment in patients on 
hemodialysis: one center study. Therap Apher 
Dial. 9(3):233-6, 2005 Jun. PMID: 15966996
Not test-directed Tx (KQ3)

353) Malyszko J, Malyszko JS, Hryszko T, Pawlak 
K, Mysliwiec M. Is hepcidin a link between 
anemia, inflammation and liver function in 
hemodialyzed patients?. Am J Nephrol. 
Analytical validity data only (KQ2)

354) Malyszko J, Malyszko JS, Kozminski P, Koc- 
Zorawska E, Mysliwiec M, Macdougall I. Possible 
relationship between neutrophil gelatinase- 
associated lipocalin, hepcidin, and inflammation 
in haemodialysed patients. Nephron. 115(4):c268- 
75, 2010. PMID: 20424477
Analytical validity data only (KQ2)

355) Malyszko J, Malyszko JS, Kozminski P, Mysliwiec 
M. Type of renal replacement therapy and 
residual renal function may affect prohepcidin and 
hepcidin. Ren Fail. 31(7):544-8, 2009. PMID: 
20030521
Analytic validity data only (KQ2)

356) Malyszko J, Malyszko JS, Mysliwiec M. 
Hyporesponsiveness to erythropoietin therapy in 
hemodialyzed patients: potential role of 
prohepcidin, hepcidin, and inflammation. Ren 
Fail. 31(7):544-8, 2009. PMID: 19839848
Only mean values of older and newer markers 
given for two different groups given as per ESA 
response

357) Malyszko J, Malyszko JS, Mysliwiec M. Serum 
prohepcidin and hepcidin in hemodialyzed

Analytic validity data only (KQ2)


Analytic validity data only (KQ2)


Analytic validity data only (KQ2)


Analytic validity data only (for KQ2)


Analytic validity data for older markers only


Most likely reject. No native language screener to confirm.


Non CKD-patients


Comparison of Different cutoffs


Non CKD patients and comparison between only older markers.


Analytical validity data only (KQ2); Single arm Tx cohort (KQ3)


Analytical validity data only (KQ2); Single arm Tx cohort (KQ3)


Single arm Tx cohort (KQ3); Not biomarkers of interest—older markers only


Single arm Tx cohort (KQ3); Not biomarkers of interest—older markers only


Treatment is not based on biomarker. Analytic validity (correlations) seen between 2 older markers only


Case series; N<10; Crosssectional study; Not biomarkers of interest—older markers only


Single arm Tx cohort (KQ3); No Dx information


389) Moreno F, Sanz-Guajardo D, Lopez-Gomez JM, Jofre R, Valderrabano F. Increasing the hematocrit has a beneficial effect on quality of life and is safe in selected hemodialysis patients. Spanish Cooperative Renal Patients Quality of Life Study Group of the Spanish Society of Nephrology. J Am Soc Nephrol. 11(2):335-42,
N<10; Single arm Tx cohort (KQ3); No Dx information

N<10; Single arm Tx cohort (KQ3); No Dx information

Analytic validity data for older markers only

N<10; Single arm Tx cohort (KQ3); Analytical validity data only (KQ2)

Not an intervention of interest; Single arm Tx cohort (KQ3); No Dx information

Not biomarkers of interest—older biomarkers only

Not biomarkers of interest—older biomarkers only

Not-test directed Tx (KQ3); Study protocol only

Single arm Tx cohort (KQ3); Not biomarkers of interest—older biomarkers only

Only one biomarker studied

Analytic validity data only for older markers

Analytic validity data for older markers only

No comparison between markers

N<10; Not test-directed Tx (KQ3);

Single arm Tx cohort (KQ3); Not biomarkers of interest—older biomarkers only

Analytic validity data for older markers only


411) Nguyen TV. Comparison of serum ferritin and transferrin saturation values associated with two i.v. iron formulations in hemodialysis patients. Am J Health-Syst Pharm. 66(12):1101-4, 2009 Jun 15. PMID: 19498125


Treatment fixed and not dependent on markers


No comparison between markers and part of study group is non CKD patients


Analytic validity data for older markers only


Single arm treatment cohort. Analytic validity data for older markers only


Non CKD patients


No comparison between biomarkers and treatment is fixed. New marker- Non transferrin bound iron.


Analytic validity data for older markers only


Only one biomarker studied


Non CKD patients. Newer marker -Soluble transferrin receptor-ferritin index (sTFR-F)


Single arm Tx cohort (KQ3)


Analytic validity data for older markers only


Single arm treatment cohort. Analytic validity data for older markers only


Non CKD patients


No comparison between biomarkers and treatment is fixed. New marker- Non transferrin bound iron.


Analytic validity data for older markers only


Single arm Tx cohort (KQ3)


Narrative review


Treatment is fixed and analytic validity data for older markers only


Analytic validity data for older markers only

No comparison between biomarkers and treatment is fixed

Non CKD patients. Unsaturated Iron binding capacity—newer marker?

Not biomarkers of interest—older markers only (KQ2)

Treatment not based on biomarker. Mostly a cost effectiveness study

Narrative review

Treatment is fixed and not dependent on markers. Analytic validity data for older markers only

No comparison between markers and treatment is fixed

Analytic validity data for older markers only

Treatment is fixed and no comparison of biomarkers.

Single arm Tx cohorts (KQ3)

Analytic validity data for older markers only

Treatment fixed and not based on biomarkers. Analytic validity data only (KQ2)

Analytic validity data only (for KQ2)

Non CKD patients


Comparison of mean values of older markers in different groups


Analytic validity data for older markers only


Analytic validity data for older markers only


Treatment is fixed and analytic validity data for older markers only


Comparison of mean values of older markers in different groups


Not biomarkers of interest—older biomarkers only & Analytical validity data only (KQ2); Single arm Tx cohort (KQ3)


Not biomarkers of interest—older biomarkers only & Analytical validity data only (KQ2); Single arm Tx cohort (KQ3)


non-CKD population; Not biomarkers of interest—older biomarkers only & Analytical validity data only (KQ2);


N<10; Single arm Tx cohort (KQ3); Not biomarkers of interest—older markers only


Single arm Tx cohort (KQ3); Not biomarkers of interest—older markers only


Not test-directed Tx


Single arm Tx cohort (KQ3); Not biomarkers of interest—older markers only & Analytical validity data only (KQ2)


Single arm Tx cohort (KQ3); Not biomarkers of interest—older markers only & Analytical validity data only (KQ2)

same markers for both arms; results presented by age group only;


480) Roger SD, Cooper B. What is the practical conversion dose when changing from epoetin alfa to darbepoetin outside of clinical trials?. Nephrology. 9(4):223-8, 2004 Aug. PMID: 15363054


B-30


491) Saltissi D,Coles GA,Napier JA,Bentley P. The hematological response to continuous ambulatory peritoneal dialysis. Clin Nephrol. 22(1):21-7, 1984 Jul. PMID: 6478659 Single arm Tx cohort (KQ3); Not biomarkers of interest—older markers only & Analytical validity data only (KQ2)


499) Seguiais LF,Zilleruelo G. Transferrin saturation in pediatric hemodialysis. Pediatr Nephrol. 12(7):618, 1998 Sep. PMID: 9761366 Single arm Tx cohort (KQ3); Not biomarkers of interest—older markers only & Analytical validity data only (KQ2)


S Ferritin is used as cutoff for gp 1 & II


Analytic validity data only


Analytic validity data only (KQ2)


Single arm treatment cohort and analytic validity data for older markers only


Only older markers compared with analytic validity data only (does coefficient of variation matter?)


Just one biomarker discussed (ferritin) with no comparison with another in terms of influencing treatment


Analytic validity data for older markers only


No comparison between markers


Letter


Correlations given?


Not biomarkers of interest—older markers only & Analytical validity data only (KQ2)


Analytic validity data only (KQ2)


Single arm treatment cohort and analytic validity data for older markers only


narrative review


Analytic validity data for older marker only


Analytic validity data (KQ2)


Analytic validity data only
Analytic validity data only

Only older markers studied in terms of mean values. No comparison of older with newer marker.

CHOIR study; Not test-directed Tx based on our tests of interest

Not test-directed Tx

Analytic validity data for older markers only

Not test-directed Tx

Analytic validity data for older markers only

526) Sirover WD, Siddiqui AA, Benz RL. Beneficial hematologic effects of daily oral ascorbic acid therapy in ESRD patients with anemia and abnormal iron homeostasis: a preliminary study. Ren Fail. 30(9):884-9, 2008. PMID: 18925528
Only older markers studied in terms of analytic validity

Single arm Tx cohort (KQ3)

Single arm treatment cohort and analytic validity data only

Treatment not dependent on markers and comparison of markers

Only analytic validity (mean values of biomarkers) seen

Analytic validity data only (for KQ2)

No CKD population

No comparison between markers and treatment is not dependent on markers. Newer marker ferrum

Not a direct study but a questionnaire

No marker used and treatment is not dependent on marker


No comparison between older and newer markers but only older markers mean levels studied with respect to two different treatments


Not test-directed Tx


Sensitivity, specificity and predictive values comparison only between older markers


Sensitivity and specificity comparison only between older markers


Analytic validity data only (KQ2)


not test-directed Tx


Only analytic validity data. ISAT newer marker used


Non CKD patients


Analytic validity data for older markers only


Analytic validity data for older markers only


n<10


Non CKD patients

Sunder-Plassmann G,Horl WH. Importance of iron supply for erythropoietin therapy. Nephrol Dial Transplant. 10(11):2070-6, 1995 Nov. PMID: 8643170

No PDF


Single arm treatment cohort and analytic validity data for older markers only

Sungur C,Akpolat T,Ozdemir O,Ozcebe O,Yasavul U,Turban C,Caglar S. Hematologic profile of dialysis patients receiving alpha-

Analytic validity data for older markers only

Analytic validity data for older markers only


Non CKD patients

Not biomarkers of interest—older markers only & Analytical validity data only (KQ2)

Treatment fixed and not based on biomarker

Treatment is not based on biomarker

Analytical validity data only (KQ2)

Analytic validity data only (for KQ2)

Not biomarkers of interest—older markers only & Analytical validity data only (KQ2)

Analytic validity data for older markers only

Treatment no dependent on markers and analytic validity data for older markers

New marker erythrocyte ferritin and analytic validity data for markers


Treatment not dependent on markers and analytic validity data for older markers only

Analytic validity data for older markers only

not test-directed Tx

Analytic validity data only (for KQ2)
No marker of interest studied

Treatment no dependent on markers and only one marker studied

Single arm Tx cohort (KQ3)

Treatment fixed and analytic validity data for older markers

Single arm treatment cohort and only older markers studied

Only one marker studied

Single arm Tx cohort (KQ3)

Not biomarkers of interest—older markers only & Analytical validity data only (KQ2)

Analytic validity data (KQ2). In terms of ROC analysis, newer marker to newer marker comparison made.

non CKD-patients

Not biomarkers of interest—older markers only (KQ2)

Analytical validity data only (KQ2)

Not biomarkers of interest—older markers only (KQ2)

Not biomarkers of interest (KQ2)

Not biomarkers of interest—older markers only (KQ2)

Guidelines

Not biomarkers of interest—older markers only & Analytical validity data only (KQ2)


Not biomarkers of interest—older markers only & Analytical validity data only (KQ2)


Analytic validity data only (KQ2)


Analytical validity data only (KQ2)


non-CKD population


Population not of interest


Sensitivity comparison only between two newer biomarkers, one of which is not listed in the protocol


Analytical validity data only (KQ2)


Analytic validity data only (for KQ2)


Single arm cohort study and analytic validity data only


Single arm Tx cohort (KQ3); No Dx information


estimation of body iron reserves in EPO therapy


Not biomarkers of interest—older markers only & Analytical validity data only (KQ2)


Not biomarkers of interest—older markers only & Analytical validity data only (KQ2)


Single arm Tx cohorts (KQ3)


Non-CKD population
Not test-directed Tx (KQ3); no Dx information

603) Wallerstedt SM, Ljungman S, Broms E, Andren L. [Iron deficiency is a common cause of bad response to epoetin treatment. It's important to follow iron status—not only the Hb value]. [Swedish]. Lakartidningen. 102(37):2550-1, 2553-5, 2005 Sep 12-18. PMID: 16200900
Single arm Tx cohort (KQ3); No Dx information

Not test-directed Tx

No biomarker comparison and treatment fixed

Single arm Tx cohort (KQ3)

Single arm tx cohort, outcome is mortality—if outcome were a diagnostic one, e.g. response to Tx then we would include for diagnostic odds ratios)

Analytic validity data only (KQ2)

non-CKD population; Not biomarkers of interest (KQ2);

While ROC information is available the it looks like most of the population is non-CKD

Single arm Tx cohort (KQ3); Not biomarkers of interest—older markers only (KQ2)

N<10; Single arm Tx cohort (KQ3); Not biomarkers of interest—older markers only & Analytical validity data only (KQ2)


The CDT biomarker discussed is not of interest

No comparison between biomarkers influencing treatments

Single arm Tx cohort (KQ3); Not biomarkers of interest—older markers only & Analytical validity data only (KQ2)


Yong K,Kairaitis L. Effects of proactive iron and erythropoiesis-stimulating agent protocol implementation on achieving clinical guideline targets for anaemia in a satellite haemodialysis patient cohort. Nephrology. 15(3):288-93, 2010 Apr. PMID: 20470296 Single arm cohort and no comparison between newer and older markers

Yorgin PD BA. Sodium ferric gluconate therapy in renal transplant and renal failure patients. No source-cochran. PMID: 11149105, Single arm Tx cohort (KQ3); No Dx information


Zanen AL,Adriaansen HJ,van Bommel EF,Posthuma R,Th de Jong GM. 'Oversaturation' of transferrin after intravenous ferric gluconate (Ferrlecit(R)) in haemodialysis patients. Nephrol Dial Transplant. 11(5):820-4, 1996 May. PMID: 8671901 not test-directed Tx


Zehnder C,Blumberg A. Human recombinant erythropoietin treatment in transfusion dependent anemic patients on maintenance hemodialysis. Clin Nephrol. 31(2):55-9, 1989 Feb. PMID: 2920469 Single arm Tx cohort (KQ3); Not biomarkers of interest—older markers only & Analytical validity data only (KQ2)

Zelichowski G,Lubas A,Wankowicz Z. [Effectiveness of subcutaneous and intravenous epoetin alpha in haemodialysed patients]. [Polish]. Pol Merkuriusz Lek. 17(98):143-7, 2004 Aug. PMID: 15603323 Single arm Tx cohort (KQ3); Not biomarkers of interest—older markers only & Analytical validity data only (KQ2)


Futenma A, Okura T, Kawahara H.[Clinical evaluation of red blood cell ferritin and serum ferritin in maintenance hemodialysis patients (author's transl)]. [Japanese] PMID: 6808206 Not relevant

Toriyama T, Matsu S, Fukatsu A, Takahashi H, Sato K, Mimuro N, Kawahara H.Effects of high-dose vitamin B6 therapy on microcytic and hypochromic anemia in hemodialysis patients. PMID: 8255009 Analytic validity data (KQ2)
van der Putten K, Jie KE, van den Broek D, Kraaijenhagen RJ, Laarakkers C, Swinkels DW, Braam B, Gaillard CA. Hepcidin-25 is a marker of the response rather than resistance to exogenous erythropoietin in chronic kidney disease/chronic heart failure patients. PMID: 20601671 Analytic validity data (KQ2)

Malovrh M, Hojs N, Premru V. The influence of need-based, continuous, low-dose iron replacement on hemoglobin levels in hemodialysis patients treated with erythropoiesis-stimulating agents. PMID: 20618233 Not relevant


Ford BA, Eby CS, Scott MG, Coyne DW. Intra-individual variability in serum hepcidin precludes its use as a marker of iron status in hemodialysis patients. PMID: 20668427 Analytic validity data (KQ2)

Barrios Y, Espinoza M, Baron MA. [Pro-hepcidin, its relation with indicators of iron metabolism and of inflammation in patients hemodialyzed treated or not with recombinant erythropoietin]. [Spanish] PMID: 20694291 Analytic validity data (KQ2)


Bratescu LO, Barsan L, Munteanu D, Stancu S, Mirescu G. Is hepcidin-25 a clinically relevant parameter for the iron status in hemodialysis patients?. PMID: 20797577 Analytic validity data (KQ2)

Ferrari P, Kulkarni H, Dheda S, Betti S, Harrison C, St Pierre TG, Olynuk JK. Serum iron markers are inadequate for guiding iron repletion in chronic kidney disease. PMID: 20876673 Analytic validity data (KQ2)

Qunibi WY, Martinez C, Smith M, Benjamin J, Mangione A, Roger SD. A randomized controlled trial comparing intravenous ferric carboxymaltose with oral iron for treatment of iron deficiency anaemia of non-dialysis-dependent chronic kidney disease patients. PMID: 20929915 Not relevant


Przybylowski P, Małyszko J, Małyszko JS, Koc-Zorawska E, Sadowski J, Mysliwiec M. Anemia in heart and kidney allotransplant recipients: is there a role for hepcidin?. PMID: 21168677 Analytic validity data (KQ2)


Mikhail A, Shrivastava R, Richardson D. Renal Association Clinical Practice Guideline on anaemia of chronic kidney disease. PMID: 21555890 Guideline document


653) Urrechaga E, Unceta M, Borque L, Escanero JF. Low hemoglobin density potential marker of iron availability. PMID: 21722324 Not relevant


657) Smorkalova EV, Aznabaeva LF, Nikulicheva VI, Safuanova GSh, Chepurnaia AN. [The characteristics of iron metabolism under iron-deficiency anemia and chronic disorders anemia]. [Russian] PMID: 21899115 Not relevant

658) Besarab A. Anemia and iron management. PMID: 21906169 Review article


661) Xu Y, Ding XQ, Zou JZ, Liu ZH, Jiang SH, Chen YM. Serum hepcidin in haemodialysis patients: associations with iron status and microinflammation. PMID: 22118000 Analytic validity data (KQ2)

662) Mahdavi MR, Makhlough A, Kosaryan M, Roshan P. Credibility of the measurement of serum ferritin and transferrin receptor as indicators of iron deficiency anemia in hemodialysis patients. PMID: 22165676 Analytic validity data (KQ2)
## Appendix C. Blank Extraction Form

### A. Source and Extractor

<table>
<thead>
<tr>
<th>Author, Year</th>
<th>Site, Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extractor</td>
<td>Funding source</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RefID</th>
<th>PMID</th>
<th>Is there &gt;1 form for this RefID (Y/N)?</th>
</tr>
</thead>
</table>

### Key Question

- **KQ1:** Management of iron deficiency anemia with biomarkers
- **KQ2:** Diagnostic accuracy; 2a: Reference standards; 2b: Diagnostic utility as an add-on test; 2c: Adverse events of testing;
- **KQ3:** Effect on Treatment—Intermediate outcomes & Adverse events;
- **KQ4:** Influencing factors

### B. Study description

<table>
<thead>
<tr>
<th>Study design</th>
<th>Recruitment method</th>
<th>Sampling population (can be more than one category)</th>
<th>Age Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inclusion criteria</td>
<td>Exclusion criteria</td>
<td>Multi-center (Y/N)?</td>
<td>Setting</td>
</tr>
<tr>
<td>Age Group</td>
<td>Setting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comments (e.g./o selection or other bias):</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. Cross Sectional; Case Control; Retrospective Cohort; Prospective Cohort; Non-Randomized Comparative Study; RCT; Other (mention)
2. Consecutive patients; Random sampling; Convenience sample; Selected sample; Other (mention)
3. CKD patients with Stage (specify) whether on HD or PD (specify); ESRD patients; Kidney transplant patients; Other (mention)
4. HD= Hemodialysis, PD= Peritoneal dialysis, ESRD= End Stage Renal Disease (not related to CKD staging)
5. Pediatric (incl. Adolescent); Adolescent (11-18 yrs); Adult (>18 yrs); Adult & Adolescent; Adult & Pediatric (incl Adolescent); Adult
6. Mean only; If estimated, please add “estimated” in brackets. If median, SE, range, IQR, or other, specify these.
7. Setting: Hospitals; Dialysis centers; Outpatient (health care setting); Emergency Dept; Community (non-health care setting); Other (mention)
### C. Participant characteristics *

<table>
<thead>
<tr>
<th>N enrolled</th>
<th>N analyzed</th>
<th>Cause of CKD (N, %)</th>
<th>Male, (n/N, %)</th>
<th>Age, y **</th>
<th>Race (n/N, %)</th>
<th>Blood Urea Nitrogen (mg/dL) **</th>
<th>Creatinine (mg/dL) **</th>
<th>Creatinine clearance (mL/min) **</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- ** N enrolled: Number of enrolled participants.
- ** N analyzed: Number of analyzed participants.
- ** Male, (n/N, %): Percentage of male participants.
- ** Age, y: Age in years.
- ** Race: Percentage of participants of each race.
- ** Blood Urea Nitrogen (mg/dL): Mean blood urea nitrogen.
- ** Creatinine (mg/dL): Mean creatinine.
- ** Creatinine clearance (mL/min): Mean creatinine clearance.

*Add headings for additional subgroups if presented.*

** Mean only; If estimated, please add “estimated” in brackets. If median, SE, range, IQR, or other, specify these.

*** MDRD, Cockcroft-Gault, CKD-EPI, Mayo, Schwartz (children), Cystatin C

**** Only if it is NOT a typical CKD, HD or PD or kidney transplant patient; if any of these four leave it blank. If not any of these 4, add description

### D. Diagnostic performances [assumes gold standard]

<table>
<thead>
<tr>
<th>Analysis #</th>
<th>Index Test 1</th>
<th>Outcome of interest 2</th>
<th>Markers used for Diagnosis (List Reference test(s)) 3</th>
<th>Definition of Response to Tx – variable 4</th>
<th>Dx trial treatment (iron/EPO, dose, duration etc.)</th>
<th># Participants recruited who did not receive the index test</th>
<th># Participants recruited who did not receive the reference test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 **Index Test:** CHr (Hb content of reticulocytes); %HYPO (% of hypochromic RBC); ZPP (erythrocyte zinc protoporphyrin); Soluble transferrin receptor; Hepcidin; Superconducting quantum interference device (SQUID); Other (mention)

2 **Reference Test:** Serum Ferritin; TSAT; Iron binding capacity; Serum iron; Bone marrow iron; Other (mention). The markers used can be a single reference test; a combination of reference tests; a combination of reference and index tests; Other (mention)

3 **Markers for Diagnosis; Response to Tx**

4 **Hb g/dL, Ht %, Other (mention)**

<table>
<thead>
<tr>
<th>Analysis #</th>
<th>Index test Cut-off</th>
<th>Outcome of interest</th>
<th>Sn data</th>
<th>Sp data</th>
<th>Sens (95% CI)</th>
<th>Spec (95% CI)</th>
<th>ROC (AUC, ± SD/SE, p value)</th>
<th>Other / Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>TP (I+R+)</td>
<td>FN (I-R+)</td>
<td>TN (I-R-)</td>
<td>FP (I+R-)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
E. Concordance [agreement between measurements – assume no gold standard]

<table>
<thead>
<tr>
<th>Comparison</th>
<th>N Enrolled</th>
<th>N Analyzed</th>
<th>Concordance Metric*</th>
<th>Value (95% CI or LOA)**</th>
<th>Other Text Description (eg, of bias)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Index Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Bland Altman plot; LOA, Limits of Agreement (±2SD); NOT correlation coefficients and OLS regression
** Delete or correct the incorrect value/item. If change, highlight yellow.

F. Predictive performances (KQ2)

<table>
<thead>
<tr>
<th>Comparison / Risk factor (Enter unit in text field if applicable) ¹</th>
<th>Predictor</th>
<th>Predictor Type (continuous/binary)</th>
<th>Predictor Cut-off (if binary)</th>
<th>Outcome of interest ²</th>
<th>Markers used for Diagnosis (List Reference test (/s)) ³</th>
<th>Definition of Response to Tx – variable ⁴</th>
<th>Statistical test used/estimate (HR/RR/OR/AdjOR)</th>
<th>Value (HR/RR/OR/AdjOR) (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Multivariate (Y/N)

If multivariate model, list covariates in model

¹ Tests: CHR (Hb content of reticulocytes); %HYPO (% of hypochromic RBC); ZPP (erythrocyte zinc protoporphyrin); Soluble transferrin receptor; Hepcidin; Superconducting quantum interference device (SQUID); Serum Ferritin; TSAT; Iron binding capacity; Serum iron; Bone marrow iron; Other (mention)

² The markers used can be a single reference test; a combination of reference tests; a combination of reference and index tests; Other (mention)

³ Hb g/dL, Ht %, Other (mention)

G. Subgroup Analysis (KQ4)
Use the tables in section D, E and F (as appropriate) and add a column in the beginning titled "subgroup" to indicate the subgroups for which results are being reported.

H. Adverse Effects or Harms Associated with Testing (KQ2b)
### I. Methodological Quality

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prospective or Retrospective?</td>
<td>(Y/N/Not mentioned)</td>
</tr>
<tr>
<td>Was patient selection consecutive?</td>
<td>(Y/N/Unclear)</td>
</tr>
<tr>
<td>Are subjects representative of the patients who will receive this test?</td>
<td>(Y/N/Unclear)</td>
</tr>
<tr>
<td>Were inclusion and exclusion criteria defined?</td>
<td>(Y/N/Unclear)</td>
</tr>
<tr>
<td>Clear description of population studied</td>
<td>(Y/N)</td>
</tr>
<tr>
<td>Was recruitment period defined?</td>
<td>(Y/N)</td>
</tr>
<tr>
<td>Adequate description of tests?</td>
<td>(Y/N)</td>
</tr>
<tr>
<td>Is the reference standard likely to correctly classify the target condition?</td>
<td>(Y/N/Unclear)</td>
</tr>
<tr>
<td>Is verification bias unlikely?</td>
<td>(Y/N/Unclear) (if no/unclear, describe)</td>
</tr>
<tr>
<td>Index Test Readers BLINDED to Reference Test Results?</td>
<td>(Y/N/No data)</td>
</tr>
<tr>
<td>Was time interval between index and reference test reported?</td>
<td>(Y/N/Unclear)</td>
</tr>
<tr>
<td>Analytic problem?</td>
<td>(Y/N) (if yes describe)</td>
</tr>
<tr>
<td>Were statistical tests to quantify uncertainty included eg 95%CI?</td>
<td>(Y/N)</td>
</tr>
<tr>
<td>Were uninterpretable/ intermediate test results reported?</td>
<td>(Y/N/Unclear)</td>
</tr>
<tr>
<td>Were withdrawals from the study explained?</td>
<td>(Y/N/Unclear)</td>
</tr>
<tr>
<td>Data loss / not analyzed (%)</td>
<td></td>
</tr>
<tr>
<td>Overall Quality (A/B/C)</td>
<td></td>
</tr>
</tbody>
</table>

**Briefly mention why you decided on the grade**

(a) Reported estimates of diagnostic accuracy may have limited clinical applicability (generalisability) if the spectrum of tested patients is not similar to the patients in whom the test will be used in practice. The spectrum of patients refers severity of the underlying target condition, demographic features and differential diagnosis and/or co-morbidity. The judgement should be based on both the method of recruitment and the characteristics of those recruited; (b) If the reference standard used is likely to correctly classify the target condition OR is the best method available; (c) Verification bias occurs when not all of the study group receive confirmation of the diagnosis by the reference standard; generally only occurs when patients are tested by the index test prior to the reference standard; (d) e.g., improper accounting for multiple measurements in same patient (should do clustered analysis); (e) Uninterpretable / indeterminate / intermediate results typically go unreported & are removed from the analysis, leading to biased assessments. Whether bias will arise depends on the possible correlation between uninterpretable test results and the true disease status; (f) Quality: Grade A (good) studies fulfill most commonly held concepts of high quality, including the following: blinding of assessors to results of the other test, blinding to clinical information, enrollment of consecutive patients, random order of measurements or simultaneous measurements with the compared methods, clear description of the evaluated population, setting, and measurement methods; appropriate measurement of outcomes; appropriate statistical and analytic methods and reporting; no reporting errors; not excessive data loss (<20%); and no obvious bias. Grade B (moderate) studies may be susceptible to some bias, but not sufficient to invalidate the results. Such studies do not meet the criteria described in category A. They have some deficiencies but none likely to cause major bias. Study may be missing information making assessment of the limitations and potential problems difficult. Grade C (poor) studies are subject to significant bias that may invalidate the results. Such studies may have serious errors in design, analysis or reporting. These studies may have large amounts of missing information or discrepancies in reporting.
Appendix D. Test Performance of Newer Markers for Assessing Iron Status as Defined by Classical Laboratory Markers

Methods for Statistical Analyses—Relationships Between Analytic Sensitivity and Specificity, and Test Agreement

There is a lack of a generally-accepted reference standard test for determining iron deficiency in the setting of CKD. Thus, we can expect that both newer and the classical laboratory biomarkers of iron status can err, i.e., the results of either test could be different than the “true” iron status. In such a case, assessing concordance may be the only meaningful option if none of the compared tests is an obvious choice for a reference standard, e.g., when both tests are alternative methodologies to measure the same quantity.

When the original studies used classical markers of iron status to define iron deficiency and used it as the reference standard in calculating the test accuracy (sensitivity and specificity) of newer markers of iron status, we considered these data are analytic sensitivity and specificity (or technical test performance) of newer markers. To facilitate the interpretation of analytic sensitivity and specificity, whenever possible, we calculated Cohen’s kappa statistic. The Cohen’s kappa statistic can be calculated based on the 2x2 contingency tables that were used to estimate the sensitivity and specificity comparing a newer marker of iron status with a “reference standard,” as defined by the classical markers of iron status in the original studies. We used the Landis and Koch interpretation of values of kappa to determine the level of agreement: <0, poor agreement; 0 to 0.20, slight agreement; 0.21 to 0.4, fair agreement; 0.41 to 0.60, moderate agreement; 0.61 to 0.80, substantial agreement; 0.81 to 1.0, almost perfect agreement.

When Cohen’s kappa statistic could not be calculated due to insufficient data in the original studies, we used a graph that visualized the relationships between sensitivity, specificity, and kappa to aid our interpretations of the test agreements, based on the reported sensitivities and specificities (Figure D1). This graph was produced based on the formulas describing the analytic relationship between sensitivity, specificity, kappa, and raw agreement for the 2x2 contingency tables. Four curves representing the ceiling of the four kappa coefficients were plotted on a squared space of sensitivity and specificity values. The curves provide all of the sensitivities and specificities pairs below which kappa cannot be achieved. For example, for k = 0.8 (which is considered to indicate substantial agreement), the line, sensitivity equal to specificity, intersects the elliptical curve for k = 0.8 exactly at (0.9, 0.9). Thus, these sensitivities and specificities, both of which are below 0.9, can never produce a kappa of 0.8 or greater.
Figure D1. Relationships between statistical measures of test performance: analytic sensitivity and specificity, and kappa (k) statistic

Results—Concordance Between Newer Markers and Classical Markers of Iron Status

In this section, we summarize the findings from 15 cross-sectional studies evaluating the test performance of newer markers for assessing iron status as defined by classical laboratory markers.5-19 Thirteen studies enrolled adult hemodialysis (HD CKD) patients. One study enrolled adult peritoneal dialysis (PD CKD) and nondialysis (ND CKD) patients,11 and another study enrolled pediatric HD and PD CKD patients.10 Many of these studies evaluated more than one newer marker, including content of hemoglobin in reticulocytes (CHr), percent hypochromic red blood cells (%HYPO), reticulocyte hemoglobin content (RetHe), soluble transferrin receptor (sTfR), and erythrocyte zinc protoporphyrin (ZPP). Studies used a variety of definitions for iron deficiency using classical laboratory markers, such as transferrin saturation (TSAT) or ferritin.

As described in the Methods, “reference standard” tests for determining iron deficiency in the context of CKD are lacking, and both newer and classical laboratory markers are subject to measurement errors. Thus, whenever it is possible, we either calculated Cohen’s kappa statistic based on the 2x2 contingency tables or estimated kappa based on a graph depicting the
relationships between sensitivity, specificity, and kappa to aid our interpretation of test agreements (Figure D1).4

Summary of Key Points

- Among adult HD CKD patients:
  - Content of hemoglobin in reticulocytes (CHr) and classical laboratory markers (TSAT or ferritin, alone or in combination) have poor to moderate test agreements.
  - Percent hypochromic red blood cells (%HYPO) and classical laboratory markers (TSAT or ferritin, alone or in combination) have poor to fair agreements.
  - Soluble transferring receptor (sTfR) and classical laboratory markers (TSAT or ferritin, alone or in combination) have poor to fair agreements.
  - Erythrocyte zinc protoporphyrin (ZPP) and classical laboratory marker (TSAT<20%) have poor to fair agreements for assessing iron status in HD CKD patients with normal iron store as indicated by serum ferritin concentrations (>100 ng/mL). Higher ZPP cutoffs were associated with better agreements with TSAT <20%.

Content of Hemoglobin in Reticulocytes

Key Points

Seven cross-sectional studies, enrolling a total of 911 adult HD CKD patients (individual counts ranged from 22 to 217 patients in each study),6,7,9,15,16,18,19 and one cross-sectional study with 19 PD CKD patients,11 evaluated the test performance of CHr for assessing iron status as defined by classical laboratory markers. Studies enrolled primarily older patients who received maintenance ESA treatment; however, maintenance ESA doses varied across studies. Baseline iron status (based on mean serum ferritin and TSTA concentrations) also varied across studies. CHr and classical laboratory markers (TSAT or ferritin) have poor to moderate agreements for assessing iron status in HD CKD patients.

Detailed Synthesis (Table D1 to D3)

Seven cross-sectional studies, enrolling a total of 833 adult HD CKD patients (individual counts from 22 to 217 patients in each study),6,7,9,15,16,18,19 and one cross-sectional study with 19 PD CKD patients,11 evaluated the test performance of CHr for assessing iron status as defined by classical laboratory markers. Four studies were conducted in Japan, two in the U.K., one in Germany, and one in South Korea. Studies enrolled primarily older men and women (mean age ranged from 49 to 62 years old). Baseline mean Hb concentrations ranged from 9.4 to 10.3 g/dL (reported in 5 studies). Four of the eight studies reported a mean ferritin concentration less than 100 ng/mL (indicating an insufficient iron store status),6,7,18,19 while the other four reported a mean ferritin concentration greater than 100 ng/mL at baseline (ranging from 198 to 427 ng/mL).9,11,15,16 Baseline mean TSAT ranged from 20 to 32 percent. Most patients received maintenance ESA treatment (one study enrolled 29 patients who had not received ESA treatment), but the maintenance ESA doses varied across studies (Table D1).

The test performance of CHr was assessed using four different reference standards of iron status as defined by classical laboratory markers: 1) TSAT ≤15%;9 2) TSAT ≤20%;6,7,15,19 3) ferritin ≤100 ng/mL;7 and 4) TSAT ≤20% or ferritin ≤100 ng/mL.11,15,16,18 Seven studies used
variable CHr cutoffs to define iron deficiency (ranging from <26 to <35 pg), of which only two both used cutoffs of CHr <31 and <32 pg. One study analyzed the test performance of CHr as a continuous measure.

The reported sensitivity and specificity pairs at different CHr cutoffs, as well as the estimated agreements between CHr and classical markers of iron status, are summarized in Tables D2 and D3. Seven studies (6 studies in HD CKD patients and 1 study in PD CKD patients) showed poor to moderate test agreements between CHr and classical markers of iron status. Most studies were small in sample size (N<100) and thus cannot provide precise test agreement estimates. Moreover, the heterogeneity in the test comparisons, iron status of the study populations, and background treatment may explain the inconsistencies in the test agreements across studies.

Another study of 149 HD CKD patients performed multivariate analyses, and showed that Chr was the most significant predictor among the markers examined (including hematocrit, %HYPO, log ferritin, sTfR, and epoetin user status). The multivariate analyses also showed that each 1 pg decrease in CHr and one standard deviation decrease (equivalent to 2.25 pg) in CHr were associated with a 1.88 (95 percent CI, 1.39 to 2.53) and 4.11 (95 percent CI, 2.10 to 8.05) fold increase, respectively, in the odds of iron deficiency (defined as TSAT < 20 percent).

Table D1. Characteristics of cross-sectional studies evaluating the test performance of reticulocyte hemoglobin content (CHr) for assessing iron status as defined by classical laboratory markers

<table>
<thead>
<tr>
<th>Study, Year [Pubmed ID]</th>
<th>Country</th>
<th>Sampling Population</th>
<th>N enrolled / N analyzed</th>
<th>Demographics</th>
<th>Kidney Function Indices</th>
<th>Anemia and Iron Status Indices</th>
<th>Background Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Adult HD CKD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Age (yr): 62</td>
<td>Hct (%): NR</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Race (%): NR</td>
<td>Ferritin (ng/mL): 33.4 μg/L [range, 4 to 56 μg/L]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TSAT (%): 25.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Age (yr): 62</td>
<td>Hct (%): NR</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Race (%): NR</td>
<td>Ferritin (ng/mL): ≤100 in 64%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TSAT (%): ≤20 in 51%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Age (yr): 64</td>
<td>Hct (%): NR</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Race (%): 100% Caucasian</td>
<td>Ferritin (ng/mL): 427.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TSAT (%): 19.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Age (yr): 59</td>
<td>Hct (%): NR</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Race (%): NR</td>
<td>Ferritin (ng/mL): 247.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TSAT (%): 25.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study, Year [Pubmed ID]</td>
<td>Country</td>
<td>Sampling Population</td>
<td>Nenrolled / Nanalyzed</td>
<td>Demographics</td>
<td>Kidney Function Indices</td>
<td>Anemia and Iron Status Indices</td>
<td>Background Treatment</td>
</tr>
<tr>
<td>-------------------------</td>
<td>---------</td>
<td>---------------------</td>
<td>-----------------------</td>
<td>--------------</td>
<td>------------------------</td>
<td>-------------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Kim, 2008[18190467]</td>
<td>South Korea</td>
<td>HD CKD</td>
<td>140 / 140</td>
<td>Male (%): 49 Age (yr): 56 Race (%): NR</td>
<td>NR</td>
<td>Hb (g/dL): 10 Hct (%): 32.6 ferritin (ng/mL): 224.5 TSAT (%): 27.6</td>
<td>ESA dose: NR (all patients received rHuEpo alpha intravenously)</td>
</tr>
<tr>
<td>Miwa, 2010[19624802]</td>
<td>Japan</td>
<td>HD CKD</td>
<td>217 / 752 samples (multiple samples from single subject)</td>
<td>Male (%): 59 Age (yr): 58 Race (%): NR</td>
<td>NR</td>
<td>Hb (g/dL): 10.3 Hct (%): 33 ferritin (ng/mL): 74.4 TSAT (%): 20.8</td>
<td>ESA dose: 5069 IU/wk</td>
</tr>
<tr>
<td>Tsuchiya, 2003[12608554]</td>
<td>Japan</td>
<td>HD CKD</td>
<td>149 / 149</td>
<td>Male (%): 38 Age (yr): 55 Race (%): NR</td>
<td>NR</td>
<td>Hb (g/dL): 10.2 Hct (%): 32.4 ferritin (ng/mL): 98.1 TSAT (%): 23</td>
<td>ESA dose: 93.1 IU/kg/wk (n=120, patients received EPO treatment)</td>
</tr>
<tr>
<td>Eguchi, 2010[20415234]</td>
<td>Japan</td>
<td>PD CKD</td>
<td>19 / 85 samples (multiple samples from single subject)</td>
<td>Male (%): 26 Age (yr): 48.6 Race (%): NR</td>
<td>BUN (mg/dL): 56 Creatinine (mg/dL): 11.6</td>
<td>Hb (g/dL): 9.8 Hct (%): 29.9 ferritin (ng/mL): 197.8 TSAT (%): 32.6</td>
<td>ESA dose: 17,833 IU/month (Epoetin-beta); 88.2 µg/ month (darbepoetin-alpha)</td>
</tr>
</tbody>
</table>

BUN=blood urea nitrogen; CHr=content of hemoglobin in reticulocytes; CKD=chronic kidney disease; Dx=diagnosis; ESA=erythropoiesis stimulating agents; ESRD=end stage renal disease; Hb=hemoglobin; HD=hemodialysis; Hct=hematocrit; IV=intravenous; IU=international units; NR=not reported; PD=peritoneal dialysis; TSAT=transferrin saturation; wk=week; yr=year

**Table D2. Summary of results for the test agreements between reticulocyte hemoglobin content (CHr) and classical markers of iron status**

<table>
<thead>
<tr>
<th>Study</th>
<th>Nanalyzed</th>
<th>CHr Cutoff (pg)</th>
<th>K (CI)</th>
<th>K (CI)</th>
<th>K (CI)</th>
<th>K (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adult HD CKD</td>
<td>25</td>
<td>&lt;26</td>
<td>0.74</td>
<td>0.29</td>
<td>0.74</td>
<td>(0.29, 0.74)</td>
</tr>
<tr>
<td>Study</td>
<td>N&lt;sub&gt;analyzed&lt;/sub&gt;</td>
<td>CHr Cutoff (pg)</td>
<td>K (CI)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>K (CI)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>K (CI)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>K (CI)&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>---------------</td>
<td>---------------------</td>
<td>-----------------</td>
<td>-------------------</td>
<td>-------------------</td>
<td>-------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Bhandari, 1997&lt;sup&gt;e&lt;/sup&gt;</td>
<td>22</td>
<td>&lt;26.5</td>
<td>0.47</td>
<td>(-0.03, 0.67)&lt;sup&gt;f&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kaneko, 2003&lt;sup&gt;i&lt;/sup&gt;</td>
<td>127</td>
<td>&lt;29.0</td>
<td>&lt;0.20&lt;sup&gt;i&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miwa, 2010&lt;sup&gt;g,s&lt;/sup&gt;</td>
<td>752 samples</td>
<td>≤30</td>
<td></td>
<td>0.20-0.40&lt;sup&gt;g&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kim, 2008&lt;sup&gt;g,s&lt;/sup&gt;</td>
<td>140</td>
<td>&lt;31</td>
<td></td>
<td>0.76</td>
<td>(0.64, 0.76)</td>
<td></td>
</tr>
<tr>
<td>Miwa, 2010&lt;sup&gt;g,s&lt;/sup&gt;</td>
<td>752 samples</td>
<td></td>
<td>0.40&lt;sup&gt;g&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kaneko, 2003&lt;sup&gt;i&lt;/sup&gt;</td>
<td>127</td>
<td>&lt;31.5</td>
<td>0.20&lt;sup&gt;i&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kaneko, 2003&lt;sup&gt;i&lt;/sup&gt;</td>
<td>127</td>
<td>&lt;32.0</td>
<td>0.20-0.40&lt;sup&gt;i&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miwa, 2010&lt;sup&gt;g,s&lt;/sup&gt;</td>
<td>752 samples</td>
<td></td>
<td>0.40&lt;sup&gt;g&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miwa, 2010&lt;sup&gt;g,s&lt;/sup&gt;</td>
<td>752 samples</td>
<td>&lt;32.2</td>
<td>0.20-0.40&lt;sup&gt;g&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kim, 2008&lt;sup&gt;g,s&lt;/sup&gt;</td>
<td>140</td>
<td>&lt;32.4</td>
<td>0.77</td>
<td>(0.64, 0.82)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kaneko, 2003&lt;sup&gt;i&lt;/sup&gt;</td>
<td>127</td>
<td>&lt;32.5</td>
<td>0.20-0.40&lt;sup&gt;i&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kaneko, 2003&lt;sup&gt;i&lt;/sup&gt;</td>
<td>127</td>
<td>&lt;33.0</td>
<td>0.62</td>
<td>(0.51, 0.62)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miwa, 2010&lt;sup&gt;g,s&lt;/sup&gt;</td>
<td>752 samples</td>
<td>&lt;33</td>
<td>0.20-0.40&lt;sup&gt;g&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miwa, 2010&lt;sup&gt;g,s&lt;/sup&gt;</td>
<td>752 samples</td>
<td>&lt;34</td>
<td>0.20-0.40&lt;sup&gt;g&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miwa, 2010&lt;sup&gt;g,s&lt;/sup&gt;</td>
<td>752 samples</td>
<td>&lt;35</td>
<td>0.20-0.40&lt;sup&gt;g&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tsuchiya, 2003&lt;sup&gt;i&lt;/sup&gt;</td>
<td>149</td>
<td>Per 1 pg decrease</td>
<td>Adj. OR=1.88 (CI 1.39, 2.53)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tsuchiya, 2003&lt;sup&gt;i&lt;/sup&gt;</td>
<td>149</td>
<td>Per 1 SD (2.25 pg) decrease</td>
<td>Adj. OR=4.11 (CI 2.10, 8.05)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bhandari, 1998&lt;sup&gt;g&lt;/sup&gt;</td>
<td>72</td>
<td>Reduced CHr (normal range: 25.9-30.6 )</td>
<td>0.21 (-0.01, 0.35)</td>
<td>0.06 (-0.14, 0.20)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Adult PD CKD**

| Eguchi, 2010<sup>g</sup> | 85                  | <33.8           | 0.20-0.40<sup>g</sup> |                   |                   |                   |

Adj OR=adjusted odds ratio; CHr=content of hemoglobin in reticulocytes; CI=95% confidence interval; CKD=chronic kidney disease; HD=hemodialysis; PD=peritoneal dialysis; SD=standard deviation

<sup>a</sup> Kappa was estimated based on the location of the reported sensitivity and specificity pair in Figure 3 (Chapter 2) depicting the relationship between statistical measures of test performance.

<sup>b</sup> Calculated Cohen kappa statistics based on the 2x2 contingency tables that were used to estimate the sensitivity and specificity.

<sup>c</sup> Estimated values: CHr was used as the reference standard and TSAT as the index test in the original study that a reported sensitivity of 85% and a specificity of 76%. 

---

D-6
Table D3. Study results for reticulocyte hemoglobin content (CHr)—assessing iron status as defined by classical laboratory markers

<table>
<thead>
<tr>
<th>Study, Year [Pubmed ID] Country</th>
<th>Lab Analysis or Assay</th>
<th>Index Test Cut-Off (pg)</th>
<th>Reference Standard (Iron Deficiency)</th>
<th>Iron Deficiency, %</th>
<th>Total N analyzed</th>
<th>Sensitivity, %</th>
<th>Specificity, %</th>
<th>Other Results</th>
<th>Risk of Biasa</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Adult HD CKD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kaneko, 2003** [12631092] Japan</td>
<td>Flow Cytometry (ADVIA)</td>
<td>&lt;29.0</td>
<td>TSAT ≤20%</td>
<td>34&lt;sup&gt;d&lt;/sup&gt;</td>
<td>197&lt;sup&gt;d&lt;/sup&gt;</td>
<td>9.9</td>
<td>87.5</td>
<td>NR</td>
<td>Medium</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;31.5</td>
<td>TSAT ≤20%</td>
<td>40.8</td>
<td>74.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;32.0</td>
<td>TSAT ≤20%</td>
<td>52.1</td>
<td>75.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;32.5</td>
<td>TSAT ≤20%</td>
<td>59.2</td>
<td>62.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;33.0</td>
<td>TSAT ≤20%</td>
<td>69</td>
<td>57</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kim, 2008&lt;sup&gt;6&lt;/sup&gt; [18190467] South Korea</td>
<td>ADVIA 120 haematologic analyser (Bayer Diagnostics, Tarrytown, NY, USA)</td>
<td>&lt;31</td>
<td>TSAT &lt;20% or ferritin &lt;100 μg/L</td>
<td>38</td>
<td>140</td>
<td>72</td>
<td>100</td>
<td>NR</td>
<td>Medium</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;32.4</td>
<td>TSAT &lt;20% or ferritin &lt;100 μg/L</td>
<td>96</td>
<td>84</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;33</td>
<td>TSAT &lt;20% or ferritin &lt;100 μg/L</td>
<td>100</td>
<td>68</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tsuchiya, 2003&lt;sup&gt;19&lt;/sup&gt; [12608554] Japan</td>
<td>Advia 120 Autoanalyser, Bayer Diagnostics</td>
<td>Per 1 pg decrease</td>
<td>TSAT &lt;20%</td>
<td>NR</td>
<td>149</td>
<td>NR</td>
<td>NR</td>
<td>OR: 2.26 (1.72,2.98), P&lt;0.001 AdjOR&lt;sup&gt;c&lt;/sup&gt;: 1.88 (1.39,2.53), P&lt;0.001</td>
<td>Medium</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Per 1 SD (2.25 pg) decrease</td>
<td>TSAT &lt;20%</td>
<td>149</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bhandari, 1997&lt;sup&gt;6&lt;/sup&gt; [9398126] UK</td>
<td>Bayer automated Technicon H3 analyser</td>
<td>&lt;26.5</td>
<td>TSAT &lt; 20%</td>
<td>36</td>
<td>22</td>
<td>47&lt;sup&gt;c&lt;/sup&gt;</td>
<td>95&lt;sup&gt;c&lt;/sup&gt;</td>
<td>NR</td>
<td>High</td>
</tr>
<tr>
<td>Bhandari, 1998&lt;sup&gt;6&lt;/sup&gt; [9589376] UK</td>
<td>Bayer automated Technicon H3 analyser</td>
<td>Reduced CHr (normal range-25.9-30.6)</td>
<td>TSAT ≤20%</td>
<td>64</td>
<td>72</td>
<td>33</td>
<td>89</td>
<td>From ROC analysis, CHr “failed to show as a strong performance”</td>
<td>High</td>
</tr>
<tr>
<td>Study, Year [Pubmed ID] Country</td>
<td>Lab Analysis or Assay</td>
<td>Index Test Cut-Off (pg)</td>
<td>Reference Standard (Iron Deficiency)</td>
<td>Iron Deficiency, %</td>
<td>Total Nanalyzed</td>
<td>Sensitivity, %</td>
<td>Specificity, %</td>
<td>Other Results</td>
<td>Risk of Bias</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>----------------------</td>
<td>--------------------------</td>
<td>--------------------------------------</td>
<td>------------------</td>
<td>---------------</td>
<td>----------------</td>
<td>----------------</td>
<td>---------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Cullen, 1999&lt;sup&gt;a&lt;/sup&gt; [10193816] Germany</td>
<td>Technicon H*3 hematology analyzer, Bayer Diagnostic</td>
<td>&lt;26</td>
<td>TSAT ≤15%</td>
<td>46</td>
<td>24</td>
<td>73</td>
<td>100</td>
<td>NR</td>
<td>High</td>
</tr>
<tr>
<td>Miwa, 2010&lt;sup&gt;b&lt;/sup&gt; [19624802] Japan</td>
<td>Advia 120, Siemens</td>
<td>&lt;30</td>
<td>TSAT ≤20% or ferritin ≤100 ng/mL</td>
<td>45</td>
<td>752 samples (from 153 patients)</td>
<td>42.8</td>
<td>91.8</td>
<td>NR</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;31</td>
<td>TSAT ≤20% or ferritin ≤100 ng/mL</td>
<td></td>
<td></td>
<td>51.9</td>
<td>85.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;32</td>
<td>TSAT ≤20% or ferritin ≤100 ng/mL</td>
<td></td>
<td></td>
<td>60.8</td>
<td>76.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;32.2</td>
<td>TSAT ≤20% or ferritin ≤100 ng/mL</td>
<td></td>
<td></td>
<td>74.3</td>
<td>68.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;33</td>
<td>TSAT ≤20% or ferritin ≤100 ng/mL</td>
<td></td>
<td></td>
<td>74.3</td>
<td>64.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;34</td>
<td>TSAT ≤20% or ferritin ≤100 ng/mL</td>
<td></td>
<td></td>
<td>83.5</td>
<td>53</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;35</td>
<td>TSAT ≤20% or ferritin ≤100 ng/mL</td>
<td></td>
<td></td>
<td>89.7</td>
<td>37.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adult PD CKD</td>
<td>Eguchi, 2010&lt;sup&gt;c&lt;/sup&gt; [20415234] Japan</td>
<td>NR</td>
<td>&lt;33.8</td>
<td>TSAT ≤20% or ferritin ≤100 ng/mL</td>
<td>NR</td>
<td>NR</td>
<td>60</td>
<td>64</td>
<td>NR</td>
</tr>
</tbody>
</table>

Adj OR=adjusted odds ratio; CHr=content of hemoglobin in reticulocytes; CI=confidence interval; Dx=diagnosis; HD=hemodialysis; NR=not reported; OR=odds ratio; SD=Standard deviation; Sens=sensitivity; Spec=specificity; PD=peritoneal dialysis; TSAT=transferrin saturation

<sup>a</sup> Based on QUADAS.

<sup>b</sup> Adjusted for age, gender, duration of hemodialysis, hematocrit, CHr, %HYPO, log<sub>10</sub> ferritin, sTfR, and epoetin user status.

<sup>c</sup> Estimated values: CHr was used as the reference standard and TSAT as the index test in the original study that reported sensitivity of 85% and a specificity of 76%.

<sup>d</sup> Total number of patients analyzed and prevalence of iron deficiency were based on the data reported in Table 1 of the original article. However, the estimated specificity based on the reported 2x2 table does not match the reported specificity in the text; therefore, we cannot use the 2x2 table to estimate kappa.
Percent Hypochromic Red Blood Cells

Key Points

Four cross-sectional studies, with a total of 495 adult HD CKD patients (per study enrollment ranged from 72 to 202 patients) evaluated the test performance of CHr for assessing iron status as defined by classical laboratory markers. Studies enrolled primarily older patients who received maintenance ESA treatment, although the maintenance ESA doses varied from study to study. Baseline iron status, based on mean serum ferritin and TSAT concentrations, also varied across studies.

Overall, %HYPO and classical laboratory markers (TSAT or ferritin) have poor to fair agreements for assessing iron status in HD CKD patients.

Detailed Synthesis (Table D4 to D6)

Four cross-sectional studies, with a total of 495 adult HD CKD patients, evaluated the test performance of %HYPO for assessing iron status as defined by classical laboratory markers. Studies were conducted in several different countries: the U.K., Germany, Poland, and Japan. All studies enrolled primarily older men and women (mean age ranged from 55 to 64 years old). Baseline mean Hb concentrations ranged from 8.8 to 10.2 g/dl (reported in 3 studies). Of the four included studies, two reported a mean ferritin concentration less than 100 ng/mL (indicating an insufficient iron store status), and the other two reported a mean ferritin concentration of 274 and 427 ng/mL at baseline. Baseline mean TSAT ranged from 20 to 38 percent across studies. Most patients received maintenance ESA treatment (a total of 74 patients did not received ESA treatment), although the maintenance ESA doses varied across studies (Table D4).

The test performance of %HYPO was assessed using three different reference standards of iron status (as defined by classical laboratory markers): 1) TSAT ≤ 15%; 2) TSAT ≤ 20%; and 3) ferritin ≤100 ng/mL. Three studies used variable %HYPO cutoffs to define iron deficiency ranging from >1.5% to >10%, with two studies using the same cutoffs of >5% and >10%. One study analyzed the test performance of %HYPO as a continuous measure.

The reported sensitivity and specificity pairs at different %HYPO cutoffs, as well as the estimated agreements between %HYPO and classical markers of iron status, are summarized in Tables D5 and D6. Three studies showed poor to fair test agreements between %HYPO and classical markers of iron status in HD CKD patients, one of which also reported that %HYPO “failed to show as a strong performance as measure of iron deficiency” based on ROC analyses. The heterogeneity in the test comparisons, iron status of the study populations, and background treatment may explain the inconsistencies in the test agreements across studies.

Another study, enrolling 149 HD CKD patients, showed that, when compared to a %HYPO ≤1.5 percent, a %HYPO ≥4.1% and a %HYPO between 1.6 and 4.0% were associated with a respective 8.53 (95 percent CI, 3.42 to 21.26) and 2.20 (95 percent CI, 0.88 to 5.48) fold increase in the odds of iron deficiency (defined as a TSAT < 20 percent). However, %HYPO was not a significant predictor of iron deficiency in a multivariate model including other markers (including hematocrit, CHr, log ferritin, sTfR, and epoetin user status).
Table D4. Characteristics of cross-sectional studies evaluating the test performance of percent hypochromic red blood cell (%HYPO) for assessing iron status as defined by classical laboratory markers

<table>
<thead>
<tr>
<th>Study, Year</th>
<th>Sampling Population</th>
<th>Nenrolled / NaNalyzed</th>
<th>Demographics</th>
<th>Kidney Function Indices</th>
<th>Anemia and Iron Status Indices</th>
<th>Background Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Adult HD CKD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bhandari, 1998'</td>
<td>HD CKD</td>
<td>72 / 72</td>
<td>Male (%): 72</td>
<td></td>
<td>Hb (g/dL): NR</td>
<td>ESA dose: 2000</td>
</tr>
<tr>
<td>[9589378] UK</td>
<td>Age (yr): 62</td>
<td></td>
<td></td>
<td></td>
<td>Hct (%): NR</td>
<td>Iron washout: NC</td>
</tr>
<tr>
<td></td>
<td>Race (%): NR</td>
<td></td>
<td></td>
<td></td>
<td>ferritin (ng/mL): NR</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>≤100 in 64%</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TSAT (%): NR</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>≤20 in 51%</td>
<td></td>
</tr>
<tr>
<td>Cullen, 1999''</td>
<td>HD CKD</td>
<td>36 / 25</td>
<td>Male (%): 47</td>
<td></td>
<td>Hb (g/dL): 9.4</td>
<td>ESA dose: NR</td>
</tr>
<tr>
<td>[10193816] Germany</td>
<td>Age (yr): 64</td>
<td></td>
<td></td>
<td></td>
<td>Hct (%): NR</td>
<td>(All patients received ESA treatment intermittently before the start of the study)</td>
</tr>
<tr>
<td></td>
<td>Race (%): 100% Caucasian</td>
<td></td>
<td></td>
<td></td>
<td>ferritin (ng/mL): 427.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TSAT (%): 19.5</td>
<td></td>
</tr>
<tr>
<td>Matuszkiewicz-Rowinska 2003'\textsuperscript{17}</td>
<td>HD CKD</td>
<td>186 / 186</td>
<td>Male (%): 55</td>
<td></td>
<td>Hb (g/dL): 8.8</td>
<td>ESA dose: 71 IU/kg/wk (141 patients)</td>
</tr>
<tr>
<td>[14682204] Poland</td>
<td>Age (yr): 18 to 75</td>
<td></td>
<td></td>
<td></td>
<td>Hct (%): 26.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Race (%): NR</td>
<td></td>
<td></td>
<td></td>
<td>ferritin (ng/mL): 274</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TSAT (%): 38</td>
<td>Iron washout: No washout</td>
</tr>
<tr>
<td>Tsuchiya, 2003'' \textsuperscript{19}</td>
<td>HD CKD</td>
<td>202 / 149</td>
<td>Male (%): 38</td>
<td></td>
<td>Hb (g/dL): 10.2</td>
<td>ESA dose: 93.1 IU/kg/wk (n=120, patients received EPO treatment)</td>
</tr>
<tr>
<td>[12608554] Japan</td>
<td>Age (yr): 55</td>
<td></td>
<td></td>
<td></td>
<td>Hct (%): 32.4</td>
<td>Iron washout: No washout</td>
</tr>
<tr>
<td></td>
<td>Race (%): NR</td>
<td></td>
<td></td>
<td></td>
<td>ferritin (ng/mL): 98</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TSAT (%): 23</td>
<td></td>
</tr>
</tbody>
</table>

%HYPO=percent of hypochronic red blood cell; CKD=chronic kidney disease; CI=confidence interval; Dx=dagnosis; ESA=erythropoiesis stimulating agents; ESRD=end stage renal disease; Hb=hemoglobin; HD=hemodialysis; Hct=hematocrit; IV=intravenous; IU=international units; NR=not reported; TSAT=transferrin saturation; wk=week; yr=year
Table D5. Summary of results for the test agreements between percent hypochromic red blood cell (%HYPO) and classical markers of iron status

<table>
<thead>
<tr>
<th>Study</th>
<th>N&lt;sub&gt;analyzed&lt;/sub&gt;</th>
<th>%HYPO cutoff</th>
<th>TSAT ≤15% K (CI)&lt;sup&gt;a&lt;/sup&gt;</th>
<th>TSAT ≤20% K (CI)&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Ferritin ≤100 ng/mL K (CI)&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Adult HD CKD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cullen, 1999&lt;sup&gt;2&lt;/sup&gt;</td>
<td>25</td>
<td>&gt;2.5</td>
<td>0.36</td>
<td>(0.08, 0.51)</td>
<td></td>
</tr>
<tr>
<td>Cullen, 1999&lt;sup&gt;2&lt;/sup&gt;</td>
<td>25</td>
<td>&gt;5</td>
<td>0.51</td>
<td>(0.05, 0.67)</td>
<td></td>
</tr>
<tr>
<td>Matuszkiewicz-Rowinska 2003&lt;sup&gt;17&lt;/sup&gt;</td>
<td>186</td>
<td>&gt;10</td>
<td>0.41</td>
<td>(-0.07, 0.75)</td>
<td></td>
</tr>
<tr>
<td>Tsuchiya, 2003&lt;sup&gt;19&lt;/sup&gt;</td>
<td>48 vs. 54</td>
<td>1.6-4.0 vs. ≤1.5</td>
<td>OR=2.20 (CI 0.88, 5.48)&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tsuchiya, 2003&lt;sup&gt;19&lt;/sup&gt;</td>
<td>47 vs. 54</td>
<td>≥4.0 vs. ≤1.5</td>
<td>OR=8.53 (CI 3.42, 21.3)&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bhandari, 1998&lt;sup&gt;3&lt;/sup&gt;</td>
<td>72</td>
<td>“Each % increase”</td>
<td>-0.04 (CI 0.17, -0.28)</td>
<td></td>
<td>0.11</td>
</tr>
</tbody>
</table>

%HYPO=percent of hypochronic red blood cell; CI=95% confidence interval; CKD=chronic kidney disease; HD=hemodialysis; OR=odds ratio

<sup>a</sup> %HYPO was not a significant predictor in the multivariate model.

<sup>b</sup> Calculated Cohen kappa statistics based on the 2x2 contingency tables that were used to estimate the sensitivity and specificity.
### Table D6. Study results for percent hypochromic red blood cell (%HYPO)—iron status as defined by classical laboratory markers

<table>
<thead>
<tr>
<th>Study, Year [Pubmed ID]</th>
<th>Lab Analysis or Assay</th>
<th>Index Test Cut-Off, %</th>
<th>Reference Standard (Iron Deficiency)</th>
<th>Iron Deficiency, %</th>
<th>Total Nanalyzed</th>
<th>Sens, %</th>
<th>Spec, %</th>
<th>Other Results</th>
<th>Risk of Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adult HD CKD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Matuszkiewicz-Rowinska 2003 [14682204] Poland</td>
<td>Blue dye conventional smears and calculated on the basis of light microscope</td>
<td>&gt;5</td>
<td>TSAT &lt;20%</td>
<td>13</td>
<td>186</td>
<td>84</td>
<td>34</td>
<td>NR</td>
<td>Medium</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&gt;10</td>
<td>TSAT &lt;20%</td>
<td>13</td>
<td>186</td>
<td>68</td>
<td>58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tsuchiya, 2003 [12608554] Japan</td>
<td>ADVIA 120 autoanalyzer (Bayer Diagnostic)</td>
<td>≤1.5</td>
<td>TSAT &lt;20%</td>
<td>38</td>
<td>149</td>
<td>NR</td>
<td>NR</td>
<td>OR: 1.00 [cutoff ≤1.5%] OR: 2.20 [cutoff 1.6%-4.0%], P=0.090 OR: 8.53 [cutoff ≥4.1%], P&lt;0.001</td>
<td>Medium</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.6-4.0</td>
<td>TSAT &lt;20%</td>
<td>38</td>
<td>149</td>
<td>NR</td>
<td>NR</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>≥4.</td>
<td>TSAT &lt;20%</td>
<td>38</td>
<td>149</td>
<td>NR</td>
<td>NR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bhandari, 1998 [9589378] UK</td>
<td>Technicon H3 analyser</td>
<td>“Each % increase”</td>
<td>TSAT ≤20%</td>
<td>51</td>
<td>72</td>
<td>76</td>
<td>21</td>
<td>From ROC analysis, %HYPO “failed to show as a strong performance as measure of iron deficiency”</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ferritin ≤100 µg/L</td>
<td>58</td>
<td>72</td>
<td>18</td>
<td>70</td>
<td>From ROC analysis, %HYPO “failed to show as a strong performance as measure of iron deficiency”</td>
<td></td>
</tr>
<tr>
<td>Cullen, 1999 [10193816] Germany</td>
<td>Technicon H*3 hematology analyzer, Bayer Diagnostic</td>
<td>&gt;2.5</td>
<td>TSAT &lt;15%</td>
<td>44</td>
<td>25</td>
<td>91</td>
<td>54</td>
<td>NR</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&gt;5</td>
<td>TSAT &lt;15%</td>
<td>44</td>
<td>25</td>
<td>91</td>
<td>62</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>&gt;10</td>
<td>TSAT &lt;15%</td>
<td>44</td>
<td>25</td>
<td>64</td>
<td>77</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

%HYPO=percent of hypochronic red blood cell; HD=hemodialysis; NR=not reported; OR=odds ratio Sens=sensitivity; Spec=specificity; TSAT=transferrin saturation

a Based on QUADAS.
Reticulocyte Hemoglobin Equivalent

Key Points
Three cross-sectional studies, enrolling more than 270 adult HD CKD patients (one study included 1500 samples from an unclear number of patients),8,14,18 and one cross-sectional study with 19 PD CKD patients and an unclear number of ND CKD patients,11 evaluated the test performance of RetHe for assessing iron status as defined by classical laboratory markers. Studies enrolled primarily older patients who received maintenance ESA treatment, although the maintenance ESA doses varied across studies. Baseline iron status, based on mean serum ferritin and TSTA concentrations, also varied across studies.

One study examined the test performance of RetHe using two different reference standards, and showed that the test performance of RetHe was less favorable for assessing functional iron deficiency than for assessing traditional parameters for iron deficiency in HD CKD patients.8

Detailed Synthesis (Table D7 to D9)
Three studies, enrolling more than 270 adult HD CKD patients (one study included 1500 samples from an unclear number of patients),8,14,18 and one cross-sectional study with 19 PD CKD patients and an unclear number of ND CKD patients11 evaluated the test performance of RetHe for assessing iron status as defined by classical laboratory markers. Two studies were conducted in Japan, one in the U.S., and one in Bosnia and Herzegovia. Studies enrolled primarily older men and women (mean age ranged from 49 to 65 years old). One study did not report any information on patients’ background treatment or anemia or iron status.8 In the other three studies, baseline mean Hb concentrations ranged from 9.8 to 11.3 g/dL. One study reported a mean ferritin concentration of less than 100 ng/mL (indicating an insufficient iron store status),18 while the other two reported a mean ferritin concentration greater than 100 ng/mL at baseline (139 and 559 ng/mL).11,14 Baseline mean TSAT ranged from 20 to 39 percent across studies. All patients received maintenance ESA treatment, although the maintenance ESA doses varied across studies (Table D7).

The test performance of RetHe was assessed using four different reference standards of iron status (as defined by classical laboratory markers): 1) TSAT \(\leq 20\%\) and ferritin \(\leq 100\ ng/mL\);11,14 2) TSAT \(\leq 20\%\) or ferritin \(\leq 100\ ng/mL\);18 3) serum iron < 40 \(\mu g/dL\), TSAT<20%, ferritin <100 ng/mL, and Hb <11 g/dL;8 and 4) TSAT<20%, ferritin 100-800 ng/mL, and Hb <11 g/dL.8 No two studies used the same reference standard in the same patient population (i.e., HD CKD, PD CKD, or ND CKD patients). The RetHe cutoffs that were used to define iron deficiency ranged from <27.2 to < 35 pg, and no studies used the same cutoffs of RetHe. Thus, the consistencies of study findings cannot be assessed.

The reported sensitivity and specificity pairs at different RetHe cutoffs, as well as the estimated agreements between RetHe and classical markers of iron status are summarized in Tables D8 and D9. The three studies in HD CKD patients showed poor to moderate test agreements between RetHe and classical markers of iron status.8,14,18 One study examined the test performance of RetHe using two different reference standards: “traditional parameters for iron deficiency” (serum iron < 40 \(\mu g/dL\), TSAT<20%, ferritin <100 ng/mL, and Hb <11 g/dL) and “functional iron deficiency” (TSAT<20%, ferritin 100-800 ng/mL, and Hb <11 g/dL). This study showed that the test performance of RetHe was less favorable for assessing functional iron deficiency than for assessing traditional parameters for iron deficiency in HD CKD patients.
(areas under the curve were 0.657 vs. 0.913, respectively). However, this study had an unclear descriptions of the study population and incorrect statistical analyses due to nonadjustment for within-patient correlation.

One cross-sectional study evaluated the test performance of RetHe for assessing iron status in 19 PD CKD and 84 ND CKD patients, separately. The test agreement between RetHe and classical markers of iron status was poor. However, had a potential for selection bias, an unclear description of the study population (including how 85 samples were drawn from 19 patients), and a bias in results due to nonadjustment for within-patient correlation.

Table D7. Characteristics of cross-sectional studies evaluating the test performance of reticulocyte hemoglobin equivalent (RetHe) for assessing iron status as defined by classical laboratory markers

<table>
<thead>
<tr>
<th>Study, Year [Pubmed ID]</th>
<th>Sampling Population</th>
<th>N_enrolled / N_analyzed</th>
<th>Demographics</th>
<th>Kidney Function Indices</th>
<th>Anemia and Iron Status Indices</th>
<th>Background Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Adult HD CKD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brugnara, 2006 [16999719] USA</td>
<td>HD CKD 1500 samples</td>
<td>Male (%): 56 Age (yr): 63.5</td>
<td>Hb (g/dL): NR Hct (%): NR ferritin (ng/mL): NR TSAT (%): NR</td>
<td>ESA dose: NR</td>
<td>Iron washout: NR</td>
<td></td>
</tr>
<tr>
<td>Hukic, 2010 [21246919] Bosnia and Herzegovina</td>
<td>HD CKD continuing hospital peritoneal dialysis 53 / 53 Male (%): NR Age (yr): 53 Race (%): NR</td>
<td>Hb (g/dL): 11.0 Hct (%): 0.36 ferritin (ng/mL): 559.22 TSAT (%): 38.35</td>
<td>ESA dose: NR (all patients were on ESA therapy)</td>
<td>Iron washout: No. Additional iron therapy was given to all patients to maintain ferritin between 300 and 500 ng/mL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miwa, 2010 [19624802] Japan</td>
<td>HD CKD 217 / 752 samples (multiple samples from single subject) Male (%): 59 Age (yr): 58.4 Race (%): NR</td>
<td>Hb (g/dL): 10.3 Hct (%): 33 ferritin (ng/mL): 74.4 TSAT (%): 20.8</td>
<td>ESA dose: 5069 IU/week</td>
<td>Iron washout: No washout</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Adult PD CKD and ND CKD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eguchi, 2010 [20415234] Japan</td>
<td>PD CKD 19 / 85 samples (multiple samples from single subject) Male (%): 26 Age (yr): 48.6 Race (%): NR</td>
<td>BUN (mg/dL): 56 Creatinine (mg/dL): 11.6 eGFR (mL/min): NR</td>
<td>Hb (g/dL): 9.8 Hct (%): 29.9 ferritin (ng/mL): 197.8 TSAT (%): 32.6</td>
<td>ESA dose: 17,833 IU/ month (Epoetin-beta); 88.2 µg/ month (darbepoetin-alpha)</td>
<td>Iron washout: NR</td>
<td></td>
</tr>
<tr>
<td>Study, Year [Pubmed ID]</td>
<td>Sampling Population</td>
<td>N&lt;sub&gt;enrolled&lt;/sub&gt; / N&lt;sub&gt;analyzed&lt;/sub&gt;</td>
<td>Demographics</td>
<td>Kidney Function Indices</td>
<td>Anemia and Iron Status Indices</td>
<td>Background Treatment</td>
</tr>
<tr>
<td>-------------------------</td>
<td>---------------------</td>
<td>---------------------------------</td>
<td>--------------</td>
<td>-------------------------</td>
<td>--------------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>ND CKD</td>
<td>84/ NR</td>
<td>Male (%): 44</td>
<td>Age (yr): 65</td>
<td>BUN (mg/dL): NR</td>
<td>Hb (g/dL): 11.3</td>
<td>ESA dose: 6393 IU/ month</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Race (%): NR</td>
<td></td>
<td>Creatinine (mg/dL): NR</td>
<td>Hct (%): 34.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BUN (mg/dL): NR</td>
<td>ferritin (ng/mL): 138.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Creatinine (mg/dL): 3.0</td>
<td>TSAT (%): 28.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>eGFR (mL/min): 24</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CKD=chronic kidney disease; Dx=diagnosis; ESA=erythropoiesis stimulating agents; Hb=hemoglobin; HD=hemodialysis; Hct=hematocrit; IU=international units; NR=not reported; PD=peritoneal dialysis; TSAT=transferrin saturation; yr=year
Table D8. Summary of results for the test agreements between reticulocyte hemoglobin equivalent (RetHe) and classical markers

<table>
<thead>
<tr>
<th>Study</th>
<th>N_{analyzed}</th>
<th>RetHe Cutoff (pg)</th>
<th>Serum Iron &lt;40 µg/dL, TSAT &lt;20%, Ferritin &lt;100 mg/mL, and Hb &lt; 11 g/dL</th>
<th>TSAT &lt;20%, Ferritin 100-800 mg/mL, and Hb &lt; 11 g/dL</th>
<th>TSAT ≤20% and Ferritin ≤100 ng/mL</th>
<th>TSAT ≤20% or Ferritin ≤100 ng/mL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Adult HD CKD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brugnara, 2006^{a}</td>
<td>1500</td>
<td>&lt;27.2</td>
<td></td>
<td>K (CI)^{a} 0.60-0.80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brugnara, 2006^{a}</td>
<td>1500</td>
<td>&lt;27.9</td>
<td></td>
<td>K (CI)^{a} 0.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miwa, 2010^{a}</td>
<td>752 samples</td>
<td>&lt;30</td>
<td></td>
<td>K (CI)^{a} 0.20-0.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miwa, 2010^{a}</td>
<td>752 samples</td>
<td>&lt;31</td>
<td></td>
<td>K (CI)^{a} 0.20-0.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hukic, 2010^{a}</td>
<td>53 samples</td>
<td>&lt;31.1</td>
<td></td>
<td>K (CI)^{a} 0.47 (0.20, 0.54)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miwa, 2010^{a}</td>
<td>752 samples</td>
<td>&lt;32</td>
<td></td>
<td>K (CI)^{a} 0.20-0.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miwa, 2010^{a}</td>
<td>752 samples</td>
<td>&lt;33</td>
<td></td>
<td>K (CI)^{a} 0.20-0.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miwa, 2010^{a}</td>
<td>752 samples</td>
<td>&lt;34</td>
<td></td>
<td>K (CI)^{a} 0.20-0.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miwa, 2010^{a}</td>
<td>752 samples</td>
<td>&lt;35</td>
<td></td>
<td>K (CI)^{a} 0.20-0.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Adult PD CKD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eguchi, 2010^{a}</td>
<td>85 samples</td>
<td>&lt;32.7</td>
<td></td>
<td>K (CI)^{a} 0.20-0.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Adult ND CKD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eguchi, 2010^{a}</td>
<td>NR</td>
<td>&lt;31</td>
<td></td>
<td>K (CI)^{a} 0.20</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CI=95% confidence interval; CKD=chronic kidney disease; HD=hemodialysis; ND=nondialysis; PD=peritoneal dialysis; RetHe=reticulocyte hemoglobin equivalent; SD=standard deviation

^{a} Kappa was estimated based on the location of the reported sensitivity and specificity pair in Figure 3 (Chapter 2) depicting the relationship between statistical measures of test performance.

^{b} Calculated Cohen kappa statistics based on the 2x2 contingency tables that were used to estimate the sensitivity and specificity.
<table>
<thead>
<tr>
<th>Study, Year [Pubmed ID]</th>
<th>Lab Analysis or Assay</th>
<th>Index Test Cut-Off (pg)</th>
<th>Reference Standard (Iron Deficiency)</th>
<th>Iron Deficiency, %</th>
<th>Total N&lt;sub&gt;analyzed&lt;/sub&gt;</th>
<th>Sens, %</th>
<th>Spec, %</th>
<th>Other Results</th>
<th>Risk of Bias&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Adult HD CKD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hukic, 2010&lt;sup&gt;14&lt;/sup&gt; [21246919]</td>
<td>Sysmex XE 2100</td>
<td>&lt;31.1</td>
<td>TSAT &lt;20% and ferritin &lt;100 µg/L</td>
<td>49</td>
<td>53</td>
<td>50</td>
<td>96</td>
<td>AUC (CI): 0.73 (0.59, 0.84)</td>
<td>Medium</td>
</tr>
<tr>
<td>Miwa, 2010&lt;sup&gt;18&lt;/sup&gt; [19624802]</td>
<td>XE-2100 with upgraded software (XE RET MASTER, Sysmex)</td>
<td>&lt;30</td>
<td>TSAT ≤ 20% or ferritin ≤ 100 ng/mL</td>
<td>45</td>
<td>752 samples (from 153 patients)</td>
<td>51.9</td>
<td>85.7</td>
<td>NR</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;31</td>
<td>TSAT ≤ 20% or ferritin ≤ 100 ng/mL</td>
<td>51.9</td>
<td>85.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;32</td>
<td>TSAT ≤ 20% or ferritin ≤ 100 ng/mL</td>
<td>60.8</td>
<td>76.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;33</td>
<td>TSAT ≤ 20% or ferritin ≤ 100 ng/mL</td>
<td>74.3</td>
<td>64.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;34</td>
<td>TSAT ≤ 20% or ferritin ≤ 100 ng/mL</td>
<td>83.5</td>
<td>53</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;35</td>
<td>TSAT ≤ 20% or ferritin ≤ 100 ng/mL</td>
<td>89.7</td>
<td>37.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brugnara, 2006&lt;sup&gt;16&lt;/sup&gt; [16999719]</td>
<td>XE 2100</td>
<td>&lt;27.2</td>
<td>serum iron &lt; 40 µg/dL, TSAT&lt;20%, ferritin &lt;100 ng/mL and Hb &lt;11 g/dL (“traditional iron deficiency”)</td>
<td>NR</td>
<td>1500</td>
<td>93.3</td>
<td>83.2</td>
<td>AUC (CI): 0.913 (NR)</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;27.9</td>
<td>TSAT&lt;20%, ferritin 100-800 ng/mL and Hb &lt;11 g/dL (“functional iron deficiency anemia”)</td>
<td>40.2</td>
<td>80.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Adult PD CKD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eguchi, 2010&lt;sup&gt;11&lt;/sup&gt; [20415234]</td>
<td>Sysmex XE 2100 automated blood cell counter</td>
<td>&lt;32.7</td>
<td>TSAT &lt;20% and ferritin &lt;100ng/mL</td>
<td>15.5</td>
<td>85</td>
<td>80</td>
<td>56</td>
<td>NR</td>
<td>High</td>
</tr>
<tr>
<td>Study, Year [Pubmed ID]</td>
<td>Lab Analysis or Assay</td>
<td>Index Test Cut-Off (pg)</td>
<td>Reference Standard (Iron Deficiency)</td>
<td>Iron Deficiency, %</td>
<td>Total N&lt;sub&gt;analyzed&lt;/sub&gt;</td>
<td>Sens, %</td>
<td>Spec, %</td>
<td>Other Results</td>
<td>Risk of Bias&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>-------------------------</td>
<td>-----------------------</td>
<td>-------------------------</td>
<td>--------------------------------------</td>
<td>-------------------</td>
<td>-----------------------</td>
<td>--------</td>
<td>--------</td>
<td>---------------</td>
<td>---------------------</td>
</tr>
<tr>
<td><strong>Adult ND CKD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eguchi, 2010&lt;sup&gt;11&lt;/sup&gt; [20415234]</td>
<td>Sysmex XE 2100 automated blood cell counter</td>
<td>&lt;31</td>
<td>TSAT &lt;20% and ferritin &lt;100ng/mL</td>
<td>15.5</td>
<td>NR</td>
<td>54</td>
<td>70</td>
<td>NR</td>
<td>High</td>
</tr>
</tbody>
</table>

AUC=area under the curve; CI=95% confidence interval; Dx=diagnosis; HD=hemodialysis; NR=not reported; RetHe=reticulocyte hemoglobin equivalent; Sens=sensitivity; Spec=specificity; TSAT=transferrin saturation

<sup>a</sup> Based on QUADAS.
Soluble Transferrin Receptor

Key Points

Four cross-sectional studies, enrolling a total of 325 adult HD CKD patients, and 1 cross-sectional study with 27 pediatric HD and PD CKD patients evaluated the test performance of sTfR for assessing iron status as defined by classical laboratory markers. Of these studies, two were rated as being at a medium risk of bias, and three at a high risk of bias. Four studies enrolled primarily older HD CKD patients (mean age ranged from 43 to 62 years old), and one study enrolled 11 pediatric HD CKD patients (mean age 16 years old) and 16 pediatric PD CKD patients (mean age 13 years old). Baseline iron status, based on mean serum ferritin and TSTA concentrations, varied across studies.

Overall, sTfR and classical laboratory markers (TSAT or ferritin) have poor to fair agreements for assessing iron status in adult HD CKD patients.

Detailed Synthesis (Table D10 to D12)

Four cross-sectional studies, enrolling a total of 325 adult HD CKD patients, and 1 cross-sectional study with 27 pediatric HD and PD CKD patients evaluated the test performance of sTfR for assessing iron status as defined by classical laboratory markers. Studies were conducted in different countries: Belgium, Italy, India, Japan, and Germany. Four studies enrolled primarily older HD CKD patients (mean age ranged from 43 to 62 years old), and one study enrolled 11 pediatric HD CKD patients (mean age 16 years old) and 16 pediatric PD CKD patients (mean age 13 years old). One study did not report any information on background treatment or patients’ anemia or iron status. In the other four studies, baseline mean Hb concentrations ranged from 10.2 to 11.5 g/dL. One of the four studies reported a mean ferritin concentration less than 100 ng/mL (indicating an insufficient iron store status), while the other three reported a mean ferritin concentration greater than 100 ng/mL at baseline (ranging from 124 to 353 ng/mL). Patients in two studies received maintenance ESA treatment, while the other three did not report information on background ESA treatment.

The test performance of sTfR was assessed using three different reference standards of iron status (as defined by classical laboratory markers): 1) TSAT <20% and ferritin <100 ng/mL; 2) TSAT <20%; and 3) TSAT <16% and ferritin <12 ng/mL. Studies used variable CHr cutoffs to define iron deficiency: >1.5 mg/L, >3.05 nmol/L, >8.5 mg/L, and >28 nmol/L (pediatric CKD). One study analyzed the test performance of sTfR as a continuous measure.

The reported sensitivity and specificity pairs at different sTfR cutoffs, as well as the estimated agreements between sTfR and classical markers of iron status, are summarized in Tables D11 and D12. Among adult HD CKD patients, three studies showed poor to fair test agreements between sTfR and classical markers of iron status. Another study performed multivariate analyses and showed that sTfR remained a significant predictor among other markers (including hematocrit, CHr, %HYPO, log ferritin, and epoetin user status). The multivariate analyses also showed that each 0.1 mg/mL increase in sTfR and one standard deviation increase (equivalent to 4.28 mg/L) in sTfR were associated with a respective 1.88 (95 percent CI, 1.29 to 2.53) and 1.69 (95 percent CI, 1.03 to 2.76) fold increase in the odds of iron deficiency (defined as TSAT < 20 percent). However, CHr was the strongest predictor in this multivariate model.
Among pediatric CKD patients, one cross-sectional study with a total of 16 pediatric PD CKD patients and 11 pediatric HD CKD patients evaluated the test performance of sTfR for assessing iron status. The test agreement between sTfR (a cutoff of > 28 nM/L to define iron deficiency) and classical laboratory markers of iron deficiency (TSAT < 20% and ferritin < 100 ng/mL) was poor and of large uncertainty (wide confidence interval) due to small sample size.

Table D10. Characteristics of studies evaluating the test performance of soluble transferring receptor (sTfR) for assessing iron status as defined by classical laboratory markers

<table>
<thead>
<tr>
<th>Study, Year</th>
<th>Country</th>
<th>Sampling Population</th>
<th>N enrolled / N analyzed</th>
<th>Demographics</th>
<th>Kidney Function Indices</th>
<th>Anemia and Iron Status Indices</th>
<th>Background Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adult HD CKD</td>
<td></td>
<td>HD CKD</td>
<td>95 / 87</td>
<td>Male (%): NR</td>
<td>Hb (g/dL): 10.2</td>
<td>ESA dose: NR</td>
<td>Iron washout: 3 months</td>
</tr>
<tr>
<td></td>
<td>Belgium</td>
<td></td>
<td></td>
<td>Age (yr): 61</td>
<td>Hct (%): NR</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Race (%): NR</td>
<td>ferritin (ng/mL): 353.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>HD CKD</td>
<td>39 / 39</td>
<td>Male (%): 54</td>
<td>Hb (g/dL): 11.1</td>
<td>ESA dose: 122 IU/kg/wk (among</td>
<td>Iron washout: 2-7 days</td>
</tr>
<tr>
<td></td>
<td>Italy</td>
<td></td>
<td></td>
<td>Age (yr): 62</td>
<td>Hct (%): 34.4</td>
<td>29 patients treated with EPO)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>HD CKD</td>
<td>40 / 40</td>
<td>Male (%): 53</td>
<td>ferritin (ng/mL): 204</td>
<td>ESA dose: NR</td>
<td>Iron washout: NR</td>
</tr>
<tr>
<td></td>
<td>India</td>
<td></td>
<td></td>
<td>Age (yr): 43</td>
<td>TSAT (%): 30</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>HD CKD</td>
<td>149 / 149</td>
<td>Male (%): 38</td>
<td>Hb (g/dL): 10.2</td>
<td>ESA dose: 93.1 IU/kg/wk (n=120,</td>
<td>Iron washout: No washout and no change allowed in dose</td>
</tr>
<tr>
<td></td>
<td>Japan</td>
<td></td>
<td></td>
<td>Age (yr): 55</td>
<td>Hct (%): 32.4</td>
<td>patients received EPO treatment)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Race (%): NR</td>
<td>ferritin (ng/mL): 98.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TSAT (%): 23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pediatric CKD</td>
<td></td>
<td>HD CKD</td>
<td>11 / 11</td>
<td>Male (%): 64</td>
<td>Hb (g/dL): 11.3</td>
<td>ESA dose: 166.4 IU/kg</td>
<td>Iron washout: no washout (HD CKD patients received intravenous iron; PD CKD patients received oral iron)</td>
</tr>
<tr>
<td></td>
<td>Germany</td>
<td></td>
<td></td>
<td>Age (yr): 16.1</td>
<td>Hct (%): NR</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>HD CKD</td>
<td>16 / 16</td>
<td>Male (%): 75</td>
<td>ferritin (ng/mL): 280</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Age (yr): 12.7</td>
<td>TSAT (%): 37.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PD CKD</td>
<td></td>
<td></td>
<td>Race (%): NR</td>
<td>Hb (g/dL): 11.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>HD CKD</td>
<td></td>
<td></td>
<td>Hct (%): 12.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ferritin (ng/mL): 124.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TSAT (%): 28.4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CKD=chronic kidney disease; CI=95% confidence interval; Dx=diagnosis; ESA=erythropoiesis stimulating agents; Hb=hemoglobin; HD=hemodialysis; Hct=hematocrit; IU=international units; NR=not reported; sTfR=soluble transferrin receptor; TSAT=transferrin saturation; wk=week; yr=year
Table D11. Summary of results for the test agreements between soluble transferring receptor (sTfR) and classical markers

<table>
<thead>
<tr>
<th>Study</th>
<th>N&lt;sub&gt;analyzed&lt;/sub&gt;</th>
<th>sTfR Cutoff</th>
<th>TSAT&lt;16% and Ferritin &lt;12 ng/mL</th>
<th>TSAT &lt;20%</th>
<th>TSAT ≤20% and Ferritin ≤100 ng/mL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Adult HD CKD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fusaro, 2005&lt;sup&gt;12&lt;/sup&gt;</td>
<td>39</td>
<td>&gt;1.5 mg/L</td>
<td>0.35 (0.06, 0.35)</td>
<td>0.06 (-0.28, 0.36)</td>
<td></td>
</tr>
<tr>
<td>Gupta, 2003&lt;sup&gt;13&lt;/sup&gt;</td>
<td>40</td>
<td>&gt;8.5 mg/L</td>
<td>0.50 (0.16, 0.66)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baldus, 1998&lt;sup&gt;16&lt;/sup&gt;</td>
<td>87</td>
<td>&gt;3.05 nmol/L</td>
<td>0.16 (-0.03, 0.31)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tsuchiya, 2003&lt;sup&gt;19&lt;/sup&gt;</td>
<td>149</td>
<td>Each 0.1 mg/L increase</td>
<td>Adj OR: 1.88 (1.29, 2.53)&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tsuchiya, 2003&lt;sup&gt;19&lt;/sup&gt;</td>
<td>149</td>
<td>Each SD (4.28 mg/L) increase</td>
<td>Adj OR: 1.69 (1.03, 2.76)&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Pediatric HD and PD CKD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daschner, 1999&lt;sup&gt;10&lt;/sup&gt;</td>
<td>27</td>
<td>&gt;28 nmol/L</td>
<td>0.10 (-0.30, 0.46)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Adj OR=adjusted odds ratio; CI=95% confidence interval; CKD=chronic kidney disease; HD=hemodialysis; PD=peritoneal dialysis; SD=standard deviation; sTfR=soluble transferring receptor

<sup>a</sup> Multivariate model included the following covariates: age, gender, duration of hemodialysis, hematocrit, Chr, %HYPO, log10 ferritin, sTfR, and erythropoietin user status.

<sup>b</sup> Calculated Cohen kappa statistics based on the 2x2 contingency tables that were used to estimate the sensitivity and specificity.
Table D12. Study results for soluble transferring receptor (sTfR)—iron status as defined by classical laboratory markers

<table>
<thead>
<tr>
<th>Study, Year [Pubmed ID]</th>
<th>Lab Analysis or Assay</th>
<th>Index Test Cut-Off</th>
<th>Reference Standard (Iron Deficiency)</th>
<th>Iron Deficiency, % Nanalyzed</th>
<th>Sensitivity, %</th>
<th>Specificity, %</th>
<th>Other Results</th>
<th>Risk of Bias a</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Adult HD CKD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fusaro, 2005 [15772926]</td>
<td>Enzyme-linked immunosorbent assay (ELISA)- R&amp;D systems (Minneapolis, MN, USA)</td>
<td>&gt;1.5 mg/L</td>
<td>TSAT &lt;20% and ferritin &lt;100 ng/mL</td>
<td>41</td>
<td>39</td>
<td>63</td>
<td>43</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&gt;1.5 mg/L</td>
<td>TSAT &lt;20%</td>
<td>23</td>
<td>39</td>
<td>100</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>Tsuchiya, 2003 [12608554]</td>
<td>ELISA kit (Eiken Kagaku, Tokyo)</td>
<td>Per 10 g/L increase</td>
<td>TSAT &lt;20% and ferritin &lt;100 ng/mL</td>
<td>NR</td>
<td>149</td>
<td>NR</td>
<td>NR</td>
<td>OR: 2.26 (1.72, 2.98) P&lt;0.001 AdjOR b: 1.88 (1.29, 2.53) P&lt;0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Per 1 SD (424.8 g/L) increase</td>
<td>TSAT &lt;20% and ferritin &lt;100 ng/mL</td>
<td>NR</td>
<td>149</td>
<td>NR</td>
<td>NR</td>
<td>OR: 2.33 (1.55, 3.50) AdjOR b: 1.69 (1.03, 2.76)</td>
</tr>
<tr>
<td>Baldus, 1998 [9543601]</td>
<td>ELISA—Amgen Diagnostics.</td>
<td>&gt;3.05 µmol/L</td>
<td>TSAT &lt;20%</td>
<td>45</td>
<td>87</td>
<td>26</td>
<td>90</td>
<td>NR</td>
</tr>
<tr>
<td>Gupta, 2003 [15025343]</td>
<td>ELISA kit from Bio Plus</td>
<td>&gt;8.5 µg/mL</td>
<td>ferritin &lt;12 ng/mL and TSAT&lt;16%</td>
<td>35</td>
<td>40</td>
<td>86</td>
<td>69</td>
<td>NR</td>
</tr>
<tr>
<td><strong>Pediatric HD CKD and PD CKD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daschner, 1999 [10543327]</td>
<td>Enzyme-linked immunosorbent assay -- R&amp;D systems (Wiesbaden, Germany)</td>
<td>&gt; 28 nmol/L</td>
<td>ferritin &lt; 100 ng/mL and TSAT &lt; 20%</td>
<td>48</td>
<td>27</td>
<td>38</td>
<td>71</td>
<td>NR</td>
</tr>
</tbody>
</table>

AUC=area under the curve; AdjOR=adjusted odds ratio; CI=confidence interval; HD=hemodialysis; NKF-DOQI=National Kidney Foundation’s kidney disease outcomes quality initiative; NR=not reported; OR=odds ratio; Sens=sensitivity; sTfR=soluble transferrin receptor; Spec=specificity; TSAT=transferrin saturation

a Based on QUADAS.

b Adjusted for age, gender, duration of hemodialysis, hematocrit, CHr, %HYPO, log10 ferritin, sTfR, and epoetin user status.
Erythrocyte Zinc Protoporphyrin

Key Points

Two cross-sectional studies, enrolling a total of 281 adult HD CKD patients\(^5,17\) evaluated the test performance of CHr for assessing iron status as defined by classical laboratory markers. Of these studies, one was rated as being at a medium risk of bias, and the other at a high risk of bias. Studies enrolled primarily older patients with sufficient iron store (based on mean serum ferritin concentration).

Overall, ZPP and classical laboratory markers (TSAT<20%) have poor to fair agreements for assessing iron status in HD CKD patients with normal iron store as indicated by mean serum ferritin concentrations >100 ng/mL.

Detailed Synthesis (Table D13 to D15)

Two cross-sectional studies, enrolling a total of 281 adult HD CKD patients, evaluated the test performance of CHR for assessing iron status as defined by classical laboratory markers.\(^5,17\) These studies used a TSAT less than 20 percent to define ID. One study was conducted in Poland,\(^17\) and the other conducted in Belgium.\(^5\)

Both studies enrolled primarily older men and women (mean age 61 years old; age range 18 to 75 years old). Baseline mean Hb concentrations were 10.2 and 8.8 g/dL. Both studies reported a mean ferritin concentration greater than 100 ng/mL at baseline (353 and 274 ng/mL, respectively). Reported baseline mean TSATs were 24 to 38 percent, respectively. Both studies reported different background treatments (Table D13).

Both studies assessed the test performance of ZPP using TSAT <20% as the reference standard of iron status. The reported sensitivity and specificity pairs at different ZPP cutoffs, as well as the estimated agreements between CHr and classical markers of iron status are summarized in Tables D14 and D15. The studies showed poor to fair test agreements between ZPP and TSAT, and there were consistent threshold effects across studies. Higher ZPP cutoffs were associated with better agreements with TSAT <20%. One study also assessed the test performance of ZPP to log ferritin index ratio for assessing iron deficiency as defined by TSAT <20%. The reported sensitivity and specificity pairs for ZPP to log ferritin index ≥40 were 76% and 83%, respectively, and the sensitivity and specificity pairs for ZPP to log ferritin index ≥45 were 72% and 86%, respectively.\(^17\)

Table D13. Characteristics of cross-sectional studies evaluating the test performance of erythrocyte zinc protoporphyrin (ZPP) for assessing iron status as defined by classical laboratory markers

<table>
<thead>
<tr>
<th>Study, Year [Pubmed ID]</th>
<th>Sampling Population</th>
<th>Nenrolled / Nanalyzed</th>
<th>Demographics</th>
<th>Kidney Function Indices</th>
<th>Anemia and Iron Status Indices</th>
<th>Background Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adult HD CKD</td>
<td>HD CKD</td>
<td>95 / 87</td>
<td>Male (%): NR</td>
<td>Hb (g/dL): 10.2</td>
<td>ESA dose: NR</td>
<td>Iron washout: 3 months</td>
</tr>
<tr>
<td>Baldus, 1998(^7)</td>
<td>[9543601] Belgium</td>
<td></td>
<td>Age (yr): 61</td>
<td>Hct (%): NR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study, Year [Pubmed ID]</td>
<td>Sampling Population</td>
<td>Nenrolled / Nanalyzed</td>
<td>Demographics</td>
<td>Kidney Function Indices</td>
<td>Anemia and Iron Status Indices</td>
<td>Background Treatment</td>
</tr>
<tr>
<td>-------------------------</td>
<td>---------------------</td>
<td>----------------------</td>
<td>--------------</td>
<td>------------------------</td>
<td>-------------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Matuszkiewicz-Rowinska 2003 [14682204] Poland</td>
<td>HD CKD</td>
<td>186 / 186</td>
<td>Male (%): 55</td>
<td>NR</td>
<td>Hb (g/dL): 8.8</td>
<td>ESA dose: 71 IU/kg/wk (141 patients)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Age (yr): 18 to 75</td>
<td></td>
<td>Hct (%): 26.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Race (%): NR</td>
<td></td>
<td>Ferritin (ng/mL): 274</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TSAT (%): 38.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ESA dose: 71 IU/kg/wk (141 patients)</td>
<td></td>
</tr>
</tbody>
</table>

Table D14. Summary of results for the test agreements between erythrocyte zinc protoporphyrin (ZPP) and classical markers

<table>
<thead>
<tr>
<th>Study</th>
<th>Nanalyzed</th>
<th>ZPP Cutoff (µmol/mol)</th>
<th>TSAT &lt;20%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Adult HD CKD</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baldus, 1998</td>
<td>87</td>
<td>&gt;40</td>
<td>0.14 (0.08, 0.36)</td>
</tr>
<tr>
<td>Matuszkiewicz-Rowinska 2003</td>
<td>186</td>
<td>&gt;65</td>
<td>0.25 (0.11, 0.35)</td>
</tr>
<tr>
<td>Matuszkiewicz-Rowinska 2003</td>
<td>186</td>
<td>&gt;80</td>
<td>0.34 (0.16, 0.5)</td>
</tr>
<tr>
<td>Matuszkiewicz-Rowinska 2003</td>
<td>186</td>
<td>&gt;90</td>
<td>0.40 (0.21, 0.57)</td>
</tr>
</tbody>
</table>

CI=95% confidence interval; CKD=chronic kidney disease; HD=hemodialysis

Calculated Cohen kappa statistics based on the 2x2 contingency tables that were used to estimate the sensitivity and specificity.
Table D15. Summary results for erythrocyte zinc protoporphyrin (ZPP)—iron status as defined by classical laboratory markers

<table>
<thead>
<tr>
<th>Study, Year [Pubmed ID]</th>
<th>Lab Analysis or Assay</th>
<th>Index Test Cut-Off (µmol/Mol Heme)</th>
<th>Reference Standard (Iron Deficiency)</th>
<th>Iron Deficiency, %</th>
<th>Nanalyzed</th>
<th>Sens</th>
<th>Spec</th>
<th>Other Results</th>
<th>Risk of Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adult HD CKD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Matuszkiewicz-Rowinska</td>
<td>Hematofluorometer</td>
<td>&gt;65</td>
<td>TSAT &lt; 20%</td>
<td>13</td>
<td>186</td>
<td>72</td>
<td>70</td>
<td>NR</td>
<td>Medium</td>
</tr>
<tr>
<td>2003 [1714682204]</td>
<td>206 D, AVIV Biomedicals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>&gt;80</td>
<td>TSAT &lt; 20%</td>
<td>56</td>
<td>85</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>&gt;90</td>
<td>TSAT &lt; 20%</td>
<td>56</td>
<td>89</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ZPP/log ferritin index ≥40</td>
<td>TSAT &lt; 20%</td>
<td>76</td>
<td>83</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ZPP/log ferritin index ≥45</td>
<td>TSAT &lt; 20%</td>
<td>72</td>
<td>86</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baldus, 1998 [9543601]</td>
<td>Hematofluorometer</td>
<td>&gt;40</td>
<td>TSAT &lt; 20%</td>
<td>44</td>
<td>87</td>
<td>41</td>
<td>73</td>
<td>NR</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>206 D, AVIV Biomedicals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AUC=area under the curve; CI=95 % confidence interval; Dx=diagnosis; NR= not reported; SE=standard error; Sens=sensitivity; Spec=specificity; TSAT=transferrin saturation

a Based on QUADAS.

References:


# Appendix E. Quality Criteria and Individual Study Grades

## Table D-1. Quality criteria and individual study grades for Key Question 2

<table>
<thead>
<tr>
<th>Author Year</th>
<th>PMID</th>
<th>Prosp (y/n/NA)</th>
<th>Consecutive (y/n/nd)</th>
<th>Eligibility Defined (y/n/nd)</th>
<th>Pop Defined (y/n/nd)</th>
<th>Recruit Defined (y/n/nd)</th>
<th>Verification Bias Unlikely (y/n/nd)</th>
<th>Test Readers Blinded (y/n/nd)</th>
<th>Time Interval (y/n/nd)</th>
<th>No Analytic Problem (y/n)</th>
<th>Uncertainty (y/n)</th>
<th>Withdrawals Explained (y/n/nd/na)</th>
<th>% Data Loss</th>
<th>Overall Risk of Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baldus, 1998 [9543601]</td>
<td>N (cross sectional)</td>
<td>ND</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>ND</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>NA</td>
<td>0%</td>
<td>HIGH</td>
<td></td>
</tr>
<tr>
<td>Bhandari, 1997 [9398126]</td>
<td>N (cross sectional)</td>
<td>ND</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>ND</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>NA</td>
<td>0%</td>
<td>HIGH</td>
<td></td>
</tr>
<tr>
<td>Bhandari, 1998 [9589378]</td>
<td>N (cross sectional)</td>
<td>ND</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>ND</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>NA</td>
<td>0%</td>
<td>HIGH</td>
<td></td>
</tr>
<tr>
<td>Bovy, 2007 [17237481]</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>ND</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>NA</td>
<td>0%</td>
<td>MEDIUM</td>
<td></td>
</tr>
<tr>
<td>Brugnara, 2006 [16999719]</td>
<td>N (Retrospective)</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>ND</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>NA</td>
<td>0%</td>
<td>HIGH</td>
<td></td>
</tr>
<tr>
<td>Buttarello, 2010 [20472854]</td>
<td>Y</td>
<td>ND</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>ND</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>14%</td>
<td>MEDIUM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chuang, 2003 [12543894]</td>
<td>Y</td>
<td>ND</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>ND</td>
<td>N</td>
<td>Unclear</td>
<td>Y</td>
<td>NA</td>
<td>0%</td>
<td>HIGH</td>
<td></td>
</tr>
<tr>
<td>Cullen, 1999 [10193816]</td>
<td>N (cross sectional)</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>ND</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>30%</td>
<td>HIGH</td>
<td></td>
</tr>
<tr>
<td>Daschner, 1999 [10543327]</td>
<td>N (cross sectional)</td>
<td>ND</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>ND</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>NA</td>
<td>0%</td>
<td>HIGH</td>
<td></td>
</tr>
<tr>
<td>Domrongkitchaiporn, 1999 [10401012]</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>ND</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>8.7%</td>
<td>MEDIUM</td>
<td></td>
</tr>
<tr>
<td>Eguchi, 2010 [20415234]</td>
<td>Y</td>
<td>ND</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Unclear (n analyzed is not specified)</td>
<td>ND</td>
<td>ND</td>
<td>N (multiple sampling from same patient)</td>
<td>N</td>
<td>N</td>
<td>ND</td>
<td>HIGH</td>
</tr>
<tr>
<td>Fishbane, 1995 [7872320]</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>ND</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>19%</td>
<td>HIGH</td>
<td></td>
</tr>
<tr>
<td>Fishbane, 1997 [9211366]</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>ND</td>
<td>ND</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>36%</td>
<td>HIGH</td>
<td></td>
</tr>
<tr>
<td>Fusaro, 2005 [15772926]</td>
<td>N (cross sectional)</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>ND</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>NA</td>
<td>0%</td>
<td>MEDIUM</td>
<td></td>
</tr>
<tr>
<td>Gupta, 2003 [15025343]</td>
<td>ND</td>
<td>ND</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>ND</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>NA</td>
<td>0%</td>
<td>HIGH</td>
<td></td>
</tr>
<tr>
<td>Hukic, 2010 [21246919]</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>ND</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>0%</td>
<td>MEDIUM</td>
<td></td>
</tr>
<tr>
<td>Kaneko, 2003 [12631092]</td>
<td>Y</td>
<td>NA</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>8%</td>
<td>MEDIUM</td>
<td></td>
</tr>
<tr>
<td>Author Year</td>
<td>PMID</td>
<td>Prosp (y/n/NA)</td>
<td>Consecutive (y/n/nd)</td>
<td>Eligibility Defined (y/n/nd)</td>
<td>Pop Defined (y/n/nd)</td>
<td>Recruit Defined (y/n/nd)</td>
<td>Verification Bias Unlikely (y/n/nd)</td>
<td>Test Readers Blinded (y/n/nd)</td>
<td>Time Interval (y/n/nd)</td>
<td>No Analytic Problem (y/n)</td>
<td>Uncertainty (y/n)</td>
<td>Withdrawals Explained (y/n/nd/na)</td>
<td>% Data Loss</td>
<td>Overall Risk of Biasa</td>
</tr>
<tr>
<td>-------------</td>
<td>------------</td>
<td>----------------</td>
<td>---------------------</td>
<td>-----------------------------</td>
<td>---------------------</td>
<td>-------------------------</td>
<td>------------------------------------</td>
<td>-------------------------------</td>
<td>----------------------</td>
<td>-------------------------</td>
<td>-----------------</td>
<td>--------------------------------</td>
<td>-------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Kim, 2008</td>
<td>[18190467]</td>
<td>N (cross sectional)</td>
<td>ND</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>ND</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>NA</td>
<td>0%</td>
<td>MEDIUM</td>
</tr>
<tr>
<td>Matuszkiewicz-Rowinska, 2003</td>
<td>[14682204]</td>
<td>N (cross sectional)</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>ND</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>NA</td>
<td>0%</td>
<td>MEDIUM</td>
</tr>
<tr>
<td>Mitsuiki, 2003</td>
<td>[14586744]</td>
<td>N (Retrospective)</td>
<td>ND</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>ND</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>NA</td>
<td>0%</td>
<td>MEDIUM</td>
</tr>
<tr>
<td>Mittman, 1997</td>
<td>[9398141]</td>
<td>Y</td>
<td>ND</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>ND</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>0%</td>
<td>MEDIUM</td>
</tr>
<tr>
<td>Miwa, 2010</td>
<td>[19624802]</td>
<td>N (cross sectional)</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>ND</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>29%</td>
<td>HIGH</td>
</tr>
<tr>
<td>Silva, 1998</td>
<td>[9794562]</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>ND</td>
<td>Y</td>
<td>N</td>
<td>(incorporation bias)</td>
<td>N</td>
<td>Y</td>
<td>40%</td>
</tr>
<tr>
<td>Tessitore, 2001</td>
<td>[11427634]</td>
<td>ND</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>ND</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>NA</td>
<td>0%</td>
<td>HIGH</td>
</tr>
<tr>
<td>Tessitore, 2010</td>
<td>[20538788]</td>
<td>Y</td>
<td>ND</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>ND</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>0%</td>
<td>LOW</td>
</tr>
<tr>
<td>Tsuchiya, 2003</td>
<td>[12608554]</td>
<td>N (cross sectional)</td>
<td>ND</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>ND</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>NA</td>
<td>0%</td>
<td>MEDIUM</td>
</tr>
<tr>
<td>Van Wyck, 2005</td>
<td>[16316362]</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>ND</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>14%</td>
<td>MEDIUM</td>
</tr>
</tbody>
</table>

N = No; NA = not applicable; ND = not described; Y= Yes

a Criteria derived from QUADAS 36
Table D-2. Quality criteria and individual study grades for Key Question 3

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Fishbane, 2001</td>
<td>[11737617]</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>MEDIUM</td>
</tr>
<tr>
<td>Kaneko, 2003</td>
<td>[12631092]</td>
<td>Unclear</td>
<td>Unclear</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>MEDIUM</td>
</tr>
</tbody>
</table>

KQ = Key Question; N = No; Y= Yes

a Criteria derived from Cochrane risk of bias tool for intervention studies

Table D-3. Quality criteria and individual study grades for Key Question 4

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>PMID</th>
<th>Prosp (y/n/NA)</th>
<th>Consecutive (y/n/nd)</th>
<th>Eligibility Defined (y/n/nd)</th>
<th>Pop Defined (y/n/nd)</th>
<th>Recruit Defined (y/n/nd)</th>
<th>Verification Bias unlikely (y/n/nd)</th>
<th>Test Readers Blinded (y/n/nd)</th>
<th>Time Interval (y/n/nd)</th>
<th>No Analytic Problem (y/n)</th>
<th>Uncertainty (y/n)</th>
<th>Withdrawals Explained (y/n/nd/na)</th>
<th>%Data Loss</th>
<th>Overall Risk of Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ahluwalia, 1997</td>
<td>[9328369]</td>
<td>Y</td>
<td>ND</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>ND</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>10.8%</td>
<td>MEDIUM</td>
</tr>
<tr>
<td>Singh, 2007</td>
<td>[17396118]</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>ND</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>28%</td>
<td>MEDIUM</td>
</tr>
</tbody>
</table>

N = No; NA = not applicable; ND = not described; Y= Yes

a Criteria derived from QUADAS