

*Draft Comparative Effectiveness Review*

---

Number XX

**Laboratory Biomarkers for Assessing Iron Status and Managing Iron Deficiency in Late Stage Chronic Kidney Disease Patients with Anemia**

**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](http://www.ahrq.gov)

**Contract No.** [redacted]

**Prepared by:**

[redacted]

**Investigators:**

[redacted]

**AHRQ Publication No. xx-EHCxxx**  
**<Month Year>**

This report is based on research conducted by an Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. xxx-xxxx-xxxxx). The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

The information in this report is intended to help health care decisionmakers—patients and clinicians, health system leaders, and policymakers, among others—make well-informed decisions and thereby improve the quality of health care services. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information, i.e., in the context of available resources and circumstances presented by individual patients.

This report may be used, in whole or in part, as the basis for development of clinical practice guidelines and other quality enhancement tools, or as a basis for reimbursement and coverage policies. AHRQ or U.S. Department of Health and Human Services endorsement of such derivative products may not be stated or implied.

This document is in the public domain and may be used and reprinted without special permission. Citation of the source is requested.

Persons using assistive technology may not be able to fully access information in this report. For assistance contact [EffectiveHealthCare@ahrq.hhs.gov](mailto:EffectiveHealthCare@ahrq.hhs.gov).

None of the investigators has any affiliations or financial involvement that conflicts with the material presented in this report.

**Suggested citation:** [Author List]. Laboratory Biomarkers for Iron Status and Managing Iron Deficiency in Late Stage Chronic Kidney Disease in Patients with Anemia. No. XX. (Prepared by the XX Evidence-based Practice Center under Contract No. XXX-XXXX-XXXXX.) AHRQ Publication No. XX. Rockville, MD: Agency for Healthcare Research and Quality. 2012. [www.effectivehealthcare.ahrq.gov/reports/final.cfm](http://www.effectivehealthcare.ahrq.gov/reports/final.cfm).

## 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](http://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](http://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](mailto: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**

**Key Informants**

**Technical Expert Panel**

**Peer Reviewers**

# Laboratory Biomarkers for Assessing Iron Status and Managing Iron Deficiency in Late Stage Chronic Kidney Disease Patients with Anemia

## Structured Abstract

**Background:** Iron management (iron status assessment and iron treatment) is an essential part of the treatment of anemia associated with chronic kidney disease (CKD), as there are concerns regarding the adverse effects associated with both elevated doses of ESAs and supplemental iron. For this reason, assessing iron status is integral to both iron and anemia managements in CKD patients. 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 July 2011.

**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.

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 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 from a few studies 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 iron deficiency than classical markers (TSAT <20 or ferritin <100 ng/mL). There is also 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 limited our confidence in evaluating the consistency of findings across studies.

For Key Question 3 (impact on intermediate outcomes of newer markers compared to 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 CHr compared to 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 three 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 a good predictive ability when used singly to determine iron deficiency as

defined by a response to iron challenge test. 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 to those guided by TSAT or ferritin. These results suggest that CHr may be a suitable alternative marker of iron status for guiding iron treatment, and could potentially reduce the frequency of iron testing and potential harms from IV iron treatment. 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

<b>Executive Summary</b> .....	<b>ES-1</b>
Background.....	ES-1
Objectives .....	ES-2
Key Questions and Analytic Framework (Figure ES-1).....	ES-2
Key Question 1 .....	ES-3
Key Question 2 .....	ES-3
Key Question 3 .....	ES-3
Key Question 4 .....	ES-3
Methods.....	ES-4
Data Sources and Selection.....	ES-4
Quality (Risk of Bias) Assessment of Individual Studies.....	ES-5
Data Synthesis .....	ES-5
Test Performance Terms and Definitions .....	ES-6
Grading the Body of Evidence .....	ES-7
Results .....	ES-7
Literature Search.....	ES-7
Discussion.....	ES-12
Key Findings and Strength of Evidence .....	ES-12
Findings in Relationship to What is Already Known .....	ES-14
Applicability and Implications for Clinical and Policy Decisionmaking .....	ES-15
Limitations of the Evidence Base.....	ES-15
Research Gaps .....	ES-15
Conclusions.....	ES-17
References.....	ES-17
<b>Introduction</b> .....	<b>1</b>
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 .....	7
<b>Methods</b> .....	<b>8</b>
AHRQ Task Order Officer.....	8
Topic Refinement and Review Protocol .....	8
Literature Search Strategy.....	8
Search Strategy .....	8
Inclusion and Exclusion Criteria .....	9
Study Selection .....	10
Data Extraction .....	10
Risk of Bias – Assessment of Individual Studies.....	10
Data Synthesis .....	11
Summary Tables.....	11
Graphical Presentation of Study Results.....	12
Test Performance Terms and Definitions .....	12

Strength of the Body of Evidence .....	13
Applicability .....	13
Peer Review and Public Commentary.....	13
<b>Results.....</b>	<b>14</b>
Introduction.....	14
Literature Searches .....	14
Description of Included Studies .....	14
Key Question 1. Comparative Effectiveness of Newer versus Older Markers of Iron Status for the Diagnosis and Management of Iron Deficiency Anemia.....	15
Key Question 2. Test Performance of Newer Markers Compared to the Older Markers of Iron Status .....	16
2a. Reference Standards for the Diagnosis of Iron Deficiency in Studies Evaluating Test Performance .....	16
2b. Adverse Effects or Harms Associated with Testing .....	44
Key Question 3. Intermediate Outcomes Comparing Iron Management Guided by Newer Laboratory Markers with Those of Iron Management Guided by Older Laboratory Markers ..	44
Key Points (Table 3.1) .....	44
Detailed Synthesis .....	45
Key Question 4: Factors affecting test performance and clinical utility.....	51
Key Points .....	51
Detailed Synthesis (Tables 4.1 and 4.2).....	51
<b>Discussion.....</b>	<b>55</b>
Key Findings and Strength of Evidence .....	55
Findings in Relationship to What is Already Known .....	58
Applicability and Implications for Clinical and Policy Decisionmaking .....	58
Limitations of the Evidence Base.....	59
Research Gaps .....	59
Conclusions.....	61
<b>References.....</b>	<b>63</b>
<b>Abbreviations and Acronyms .....</b>	<b>68</b>

## Tables

Table ES-1. Summary of the strength of evidence addressing Key Questions .....	ES-12
Table ES-2. Research gaps and suggestions for future research .....	ES-15
Table A. Study eligibility criteria .....	9
Table 2.1. 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 .....	19
Table 2.2. Characteristics of studies evaluating the ability of newer or classical markers of iron status to predict the response to IV iron treatment .....	21
Table 2.3. 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 <sup>a</sup> in seven cohort studies among adult HD CKD patients.....	23
Table 2.4. 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 .....	27
Table 2.5. Summary results of the ability of CHr to predict the response to IV iron treatment....	31

Table 2.6. 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.....	34
Table 2.7. Summary results of the ability of %HYPO to predict the response to IV iron treatment in HD CKD patients .....	36
Table 2.8. 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.....	37
Table 2.9. Summary results of the ability of sTfR to predict the response to IV iron treatment in HD CKD patients.....	38
Table 2.10. Overall strength of evidence for the test performance of Erythrocyte Zinc Protoporphyrin (ZPP) comparing with that of classical markers of iron status to predict a response to IV iron treatment.....	39
Table 2.11. Summary results of the test performance of erythrocyte zinc protoporphyrin (ZPP) in predicting a response to IV iron treatment in adult HD CKD patients .....	41
Table 2.12. Summary results of the ability of serum hepcidin to predict the response to IV iron treatment in HD CKD patients.....	43
Table 2.13. Adverse effects or harms reported in the 27 studies included in Key Question 2.....	44
Table 3.1. 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 .....	45
Table 3.2. 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.....	48
Table 3.3. Dose of epoetin required to maintain hematocrit targets.....	49
Table 3.4. Changes in iron status markers .....	49
Table 4.1. Characteristics of studies evaluating factors affecting test performance and clinical utility.....	53
Table 4.2. Test accuracy of TSAT, ferritin, CHr and sTfR for predicting change in hemoglobin in subgroups of IV iron and no IV iron treatment.....	54
Table B. Summary of the strength of evidence addressing Key Questions.....	55
Table C. Research gaps and suggestions for future research.....	60

## Figures

Figure ES-1. Analytic framework.....	ES-4
Figure 1. Roles of current and newly proposed markers of iron status.....	4
Figure 2. Analytic framework.....	6
Figure 3: Literature flow .....	15
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 .....	25
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.....	26

## 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 which 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.<sup>1</sup> 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,<sup>1</sup> as there remain outstanding concerns regarding the adverse effects associated with elevated doses of ESAs<sup>2</sup> and supplemental iron.<sup>3</sup>

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);<sup>1</sup> however, taking a bone marrow sample is invasive and involves risks of infection or bleeding at the biopsy site.<sup>4</sup> 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).<sup>5,6</sup> 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.<sup>7-9</sup> 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,<sup>10</sup> 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);<sup>11</sup> 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.<sup>12</sup> 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.<sup>13</sup>
  - 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 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.
  - 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.<sup>14</sup>
  - Superconducting QUantum Interference Device (SQUID) is a non-invasive method for the detection and quantification of liver iron content,<sup>15</sup> 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 of 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 ES-1)

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 reviews primary, overarching question.

## Key Question 1

What is the impact on patient centered outcomes of using newer laboratory biomarkers<sup>1</sup> as a replacement for or an add-on to the older laboratory biomarkers of iron status<sup>2</sup> 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<sup>a</sup> as a replacement for or an add-on to the older markers<sup>b</sup> 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)

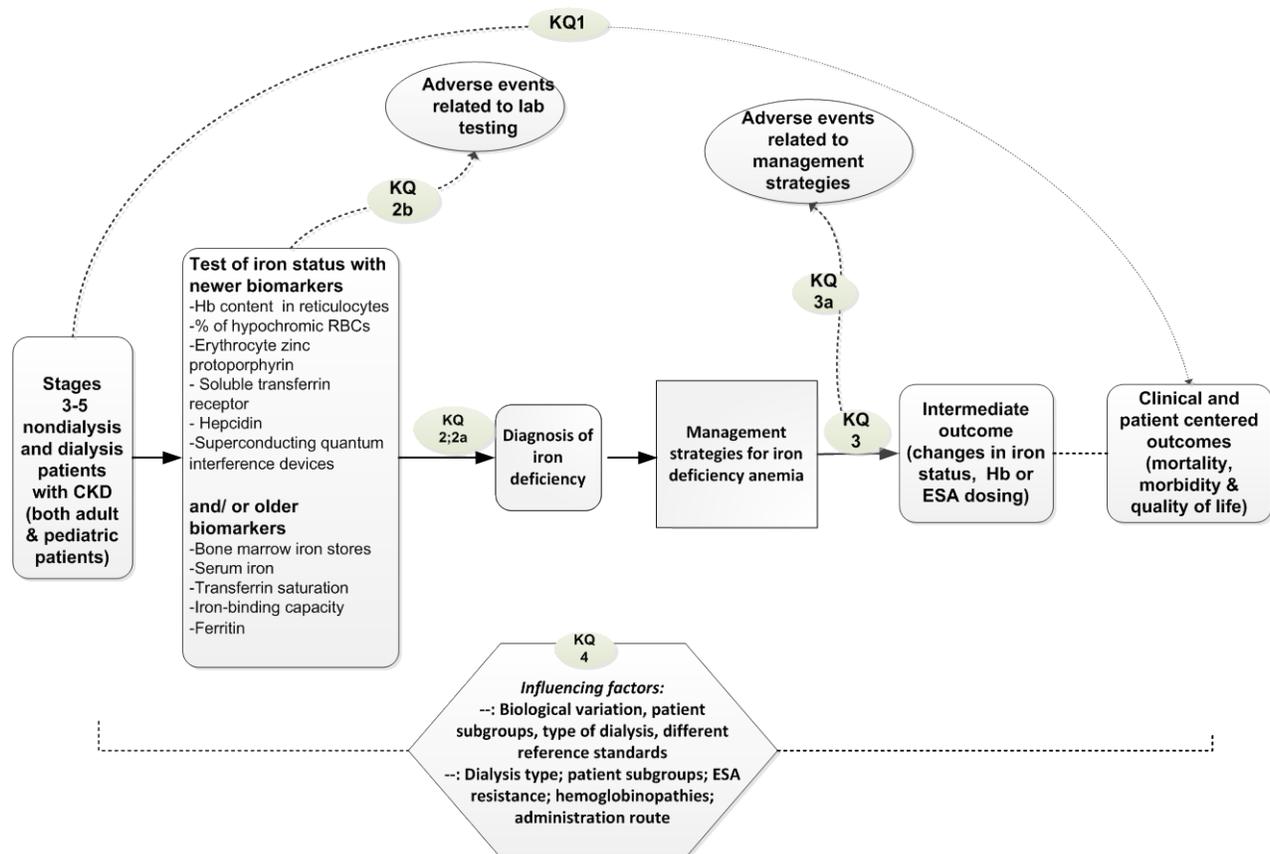
---

<sup>1</sup> Content of hemoglobin [Hb] in reticulocytes, percentage of hypochromic red blood cells, erythrocyte zinc protoporphyrin, soluble transferrin receptor, hepcidin, and superconducting quantum interference devices

<sup>2</sup> Bone marrow iron stores, serum iron, transferrin saturation, iron-binding capacity, and ferritin

- Treatment regimen (i.e., repletion or continuous treatment)
- Interactions between treatments (i.e., patients treated with versus without erythropoiesis-stimulating agents, patients treated with versus without iron-replacement therapy)
- Other factors (based on additional information in the reviewed papers).

**Figure ES-1. Analytic framework**



CKD=chronic kidney disease; ESA=erythropoiesis-stimulating agents; Hb=hemoglobin level

## Methods

### Data Sources and Selection

We conducted literature searches of studies in MEDLINE® (from inception to July 2011) and the Cochrane Central Register of Controlled Trials (through the third quarter of 2011). Studies published in any language with adult human subjects were screened to identify articles relevant to each Key Question. We also consulted 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 peritoneal dialysis [PD]); and patients with a kidney transplant. For interventions, eligible studies were those involving the newer laboratory biomarkers (see 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 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) and 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 *Methods Guide for Effectiveness and Comparative Effectiveness Review* (from here on referred to as the Methods Guide).<sup>16</sup> 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.<sup>1</sup> 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 “function 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 versus 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 versus 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 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.<sup>17</sup> 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.<sup>16</sup> 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. 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 by order of 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 5753 citations. From these, 661 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 631 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 one Polish- and one 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),<sup>10,18-43</sup> two reported intermediate outcomes comparing iron management guided by newer laboratory markers with iron management guided by classical markers (Key Question 3),<sup>42,44</sup> and three (in two articles) reported data on factors affecting test performance comparing newer with classical laboratory markers of iron status (Key Question 4).<sup>45,46</sup> 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 to 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;<sup>10,18-20,24,25,27,29-33,36,39,42</sup> These studies used the following definitions for iron deficiency: 1) TSAT  $\leq$  15%;<sup>24</sup> 2) TSAT  $\leq$  20%;<sup>18-20,29,33,39,42</sup> 3) ferritin  $\leq$  100 ng/mL;<sup>20</sup> 4) TSAT  $\leq$  20% and ferritin  $\leq$  100 ng/mL;<sup>25,27,29-31,39</sup> 5) TSAT  $\leq$  20% or ferritin  $\leq$  100 ng/mL;<sup>27,32,36,42</sup> 6) serum iron  $<$  40  $\mu$ g/dL, TSAT  $<$  20%, ferritin  $<$  100 ng/mL, and Hb  $<$  11 g/dL;<sup>10</sup> 7) TSAT  $<$  20%, ferritin 100-800 ng/mL, and Hb  $<$  11 g/dL;<sup>10</sup> and 8) TSAT  $<$  16% and ferritin  $<$  12 ng/mL.<sup>30</sup> 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.<sup>21-23,26,28,34,35,37,38,40,41,43</sup> 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.<sup>21,22,38,40,43</sup> Other reference standards included a  $\geq$  15 percent increase in Hb,<sup>37</sup> an increase in Hct of  $\geq$  3 percent and/or a  $\geq$  30 percent reduction in EPO dose,<sup>23</sup>  $>$  1 point increase in corrected reticulocyte index,<sup>28</sup> and 5% increase in Hct or a decrease in EPO dose of  $>$  2000 units per treatment.<sup>41</sup> 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, one retrospective cohort, and one 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.<sup>21-23,26,28,34,35,37,38,40,41,43</sup> Of these, 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 the eight enrolled adult hemodialysis (HD CKD) patients,<sup>21,22,28,34,35,37,38</sup> and one study enrolled adult nondialysis (ND CKD) patients.<sup>40</sup> The remaining four studies investigated the test performance of newer laboratory markers alone. Of these four, three enrolled adult HD CKD patients,<sup>23,28,43</sup> and one enrolled adult peritoneal dialysis (PD CKD) patients.<sup>26</sup> 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,<sup>21-23,28,34,35,37,38</sup> one cohort study enrolling 23 PD CKD patients,<sup>26</sup> and one cohort study enrolling 95 ND CKD patients<sup>40</sup> 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 TSTA 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 making 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).<sup>21,35</sup> Only one study performed multivariate analyses to predict a response to IV iron treatment (defined as an increase in Hct of  $\geq 3$  percent and/or a  $\geq 30$  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).<sup>23</sup>

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 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.<sup>21,22,28,37,38,43</sup> 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 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.<sup>37,38</sup>

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.<sup>21,37</sup> 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,<sup>37</sup> and one at a medium risk of bias.<sup>21</sup> The response to IV iron treatment was defined differently in the two studies, either as an increase in Hb concentration  $\geq 1$  g/dL after intravenous iron treatment,<sup>21</sup> or as an increase in Hb >15 percent from baseline.<sup>37</sup>

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 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.<sup>37,41</sup> 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 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 seven of the 27 identified studies reported information on harms.<sup>23,26,35,40-43</sup> 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 of 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 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 to 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 to 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).<sup>45</sup> 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 ( $\leq$

25%), 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% epoetin dose increase alone), the sensitivity and specificity pairs for a TSAT cutoff of  $\geq 19$  percent and a ferritin cutoff of  $\geq 726$  ng/mL were 29 and 70 percent, and 27 and 69 percent, respectively. The sensitivity and specificity pairs for a CHr cutoff of  $\geq 31.2$  pg and a sTfR cutoff of  $\geq 5.9$  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  $\geq 31.2$  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  $< 20\%$ , ferritin 100-800 ng/mL, and Hb  $< 11$  g/dL) than for assessing “traditional parameters for iron deficiency” (serum iron  $< 40$   $\mu$ g/dL, TSAT  $< 20\%$ , ferritin  $< 100$  ng/mL, and Hb  $< 11$  g/dL) in HD CKD patients.<sup>10</sup> 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 ES-1**.

**Table ES-1. Summary of the strength of evidence addressing Key Questions**

Key Questions	Strength of Evidence	Summary, Comments, and Conclusions
---------------	----------------------	------------------------------------

Key Questions	Strength of Evidence	Summary, Comments, and Conclusions
<b>Key Question 2.</b> <b>What is the diagnostic test accuracy of newer markers of iron status as a replacement for or an add-on to classical laboratory markers?</b>	Low / Insufficient (depending on the test comparisons, study populations, or test performance outcomes)	<ul style="list-style-type: none"> <li>• Among adult HD CKD patients, there is a low level of evidence that:               <ul style="list-style-type: none"> <li>○ 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 two studies suggest that CHr (with cutoff values of &lt;27 or &lt;28 pg) has a better sensitivity and specificity in predicting iron deficiency than classical markers (TSAT &lt;20 or ferritin &lt;100 ng/mL).</li> <li>○ 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 &gt;6% or &gt;10%) has a better sensitivity and specificity to predict iron deficiency (as defined by a response to IV iron treatment) than classical markers (TSAT &lt;20% or ferritin &lt;100 ng/mL).</li> <li>○ Soluble transferrin receptor (sTfR) has a similar test performance compared with classical markers (TSAT or ferritin) to predict a response to IV iron treatment.</li> </ul> </li> <li>• There is insufficient evidence regarding:               <ul style="list-style-type: none"> <li>○ Test performance of newer markers of iron status as an add-on to older markers.</li> <li>○ Test performance comparing erythrocyte zinc protoporphyrin (ZPP) and hepcidin to predict a response to IV iron treatment in adult HD CKD patients.</li> <li>○ 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.</li> </ul> </li> </ul>
<b>2a. What reference standards are used for the diagnosis of iron status in studies evaluating test accuracy?</b>	Not rated (descriptive data)	<ul style="list-style-type: none"> <li>• There is a lack of generally accepted reference standard tests for determining iron deficiency in the context of CKD.<sup>1</sup> 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 “function 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.</li> </ul>
<b>2b. What are the adverse effects or harms associated with testing using newer and/or older markers of iron status?</b>	Insufficient	<ul style="list-style-type: none"> <li>• Only 7 of the 27 studies reported information:               <ul style="list-style-type: none"> <li>○ 3 studies reported no adverse events associated with iron therapy during the study periods</li> <li>○ A total of 5 deaths reported. Studies did not attribute these deaths to either testing or any treatment.</li> <li>○ Most of the reported harms were attributed to iron therapy.</li> </ul> </li> </ul>
<b>Key Question 3.</b> <b>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?</b>	Low	<ul style="list-style-type: none"> <li>• 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 to those guided by TSAT or ferritin.</li> <li>• 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.</li> <li>• 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.</li> <li>• No study compared iron management guided by classical markers with that of newer markers (%HYPO, sTfR, Ret-He, ZPP, or hepcidin).</li> </ul>

Key Questions	Strength of Evidence	Summary, Comments, and Conclusions
<b>3a. What are the adverse effects or harms associated with the treatments guided by tests of iron status?</b>	Insufficient	<ul style="list-style-type: none"> <li>• Only 1 RCT explicitly monitored the adverse events:               <ul style="list-style-type: none"> <li>○ 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).</li> <li>○ One patient in the TSAT group dropped out because of massive bleeding due to a femoral bone fracture and need for blood transfusion.</li> <li>○ There were no significant differences in the hospitalization or infection rates of the two iron management groups.</li> </ul> </li> </ul>
<b>Key Question 4. What factors affect the test performance and clinical utility of newer markers of iron status?</b>	Insufficient	<ul style="list-style-type: none"> <li>• 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:               <ul style="list-style-type: none"> <li>○ 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.</li> <li>○ 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.</li> <li>○ 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.</li> </ul> </li> <li>• No study performed analyses by patient subgroups.</li> <li>• 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.</li> </ul>

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

## 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.<sup>1,6</sup> 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%, the combination of %HYPO >6% with CHr ≤29 pg, and the combination of % HYPO >6 with ZPP >52 µmol/mol.<sup>37</sup> 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).<sup>42,44</sup> These two trials (one conducted in the U.S. 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 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,<sup>47</sup> 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 ES-2**.

**Table ES-2. Research gaps and suggestions for future research**

Key Questions	Research Gaps	Suggestions for Future Research
---------------	---------------	---------------------------------

Key Questions	Research Gaps	Suggestions for Future Research
<b>Key Question 2.</b> <b>What is the diagnostic test accuracy of newer markers of iron status as a replacement for or an add-on to classical laboratory markers?</b>	Insufficient evidence for the test performance of newer markers of iron status as an add-on to older markers	<ul style="list-style-type: none"> <li>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.</li> </ul>
	Many existing studies are at a high risk of bias, limiting their utility in informing clinical practice	<ul style="list-style-type: none"> <li>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.<sup>48</sup></li> <li>Studies assessing diagnostic accuracy should instead aim to enroll patients representative of the spectrum of disease typically seen in clinical practice.</li> <li>Future studies should provide details about the study base and sampling methods.</li> </ul>
<b>Key Question 3.</b> <b>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?</b>	There is no uniform iron management algorithms across studies	<ul style="list-style-type: none"> <li>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</li> <li>Assessing impact of the most promising iron management algorithms on both intermediate and patient outcomes through prospective observational studies or RCTs.</li> </ul>
<b>Key Question 4.</b> <b>What factors affect the test performance and clinical utility of newer markers of iron status?</b>	Insufficient evidence to draw conclusions regarding factors that may affect the test performance or clinical utility of laboratory markers of iron status	<ul style="list-style-type: none"> <li>Future studies are need to evaluated the following factors, suggested by the experts: <ul style="list-style-type: none"> <li>Biological variation in diagnostic indices</li> <li>Use of different diagnostic reference standards</li> <li>Type of dialysis (i.e., peritoneal or hemodialysis)</li> <li>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])</li> <li>Route of iron administration (i.e., oral or intravenous)</li> <li>Treatment regimen (i.e., repletion or continuous treatment)</li> <li>Interactions between treatments (i.e., patients treated with versus without ESA, patients treated with versus without iron-replacement therapy)</li> </ul> </li> </ul>
	Whether test performance and clinical utility of newer or classical markers of iron status vary by different CKD populations are not known	<ul style="list-style-type: none"> <li>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.</li> <li>Power calculations should be performed to take into account for the planed subgroup analyses.</li> </ul>

Key Questions	Research Gaps	Suggestions for Future Research
<b>Cross-cutting issues (for Key Question 2, 3, and 4)</b>	There is no reference standard for determining iron deficiency in CKD patients	<ul style="list-style-type: none"> <li>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.</li> </ul>
	Existing studies were underpowered leading to imprecise estimates	<ul style="list-style-type: none"> <li>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.<sup>49</sup></li> </ul>
	There is no decision analysis to assess how using combinations of different markers to guide iron management strategies might influence clinical outcomes	<ul style="list-style-type: none"> <li>Patient outcomes of interest are               <ul style="list-style-type: none"> <li>Mortality</li> <li>Morbidity (e.g., cardiac or liver toxicity and infection)</li> <li>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).</li> <li>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)</li> </ul> </li> <li>For studies assessing clinical outcomes, blinding to test results to the outcome assessors is essential to avoid bias.<sup>48,50</sup></li> </ul>

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 have good predictive ability when they were used singly to determine iron deficiency as defined by a response to iron challenge test. 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 to those guided by TSAT or ferritin. These results suggest that CHr may be a suitable alternative marker of iron status for guiding iron treatment, and could potentially reduce the frequency of iron testing and potential harms from IV iron treatment.

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

1. Anaemia Management in Chronic Kidney Disease - Rapid Update 2011. National Clinical Guideline Centre; 2011.

2. Eckardt KU. Chronic kidney disease: Are elevated doses of ESAs associated with adverse outcomes? *Nature Reviews Nephrology* 2010 Oct;6(10):566-68.
3. Hayat A. Safety issues with intravenous iron products in the management of anemia in chronic kidney disease. *Clin Med Res* 2008 Dec;6(3-4):93-102.
4. Bain BJ. Bone marrow biopsy morbidity: review of 2003. *J Clin Pathol* 2005 Apr;58(4):406-08.
5. IV. NKF-K/DOQI Clinical Practice Guidelines for Anemia of Chronic Kidney Disease: update 2000. *Am J Kidney Dis* 2001 Jan;37(1 Suppl 1):S182-S238.
6. KDOQI Clinical Practice Guidelines and Clinical Practice Recommendations for Anemia in Chronic Kidney Disease. 2006.
7. Winkel P, Statland BE, Bokelund H. Factors contributing to intra-individual variation of serum constituents: 5. Short-term day-to-day and within-hour variation of serum constituents in healthy subjects. *Clin Chem* 1974 Dec;20(12):1520-27.
8. Winkel P, Statland BE, Bokelund H. The effects of time of venipuncture on variation of serum constituents. Consideration of within-day and day-to-day changes in a group of healthy young men. *Am J Clin Pathol* 1975 Oct;64(4):433-47.
9. Wish JB. Assessing iron status: beyond serum ferritin and transferrin saturation. *Clin J Am Soc Nephrol* 2006 Sep;1 Suppl 1:S4-S8.
10. Brugnara C, Schiller B, Moran J. Reticulocyte hemoglobin equivalent (Ret He) and assessment of iron-deficient states. *Clinical & Laboratory Haematology* 2006 Oct;28(5):303-08.
11. Braun J, Lindner K, Schreiber M, et al. Percentage of hypochromic red blood cells as predictor of erythropoietic and iron response after i.v. iron supplementation in maintenance haemodialysis patients. *Nephrol Dial Transplant* 1997 Jun;12(6):1173-81.
12. Macdougall IC, Cavill I, Hulme B, et al. Detection of functional iron deficiency during erythropoietin treatment: a new approach. *BMJ* 1992 Jan 25;304(6821):225-26.
13. Braun J. Erythrocyte zinc protoporphyrin. *Kidney Int Suppl* 1999 Mar;69:S57-S60.
14. Babitt JL, Lin HY. Molecular mechanisms of hepcidin regulation: implications for the anemia of CKD. *Am J Kidney Dis* 2010 Apr;55(4):726-41.
15. Gandon Y, Olivie D, Guyader D, et al. Non-invasive assessment of hepatic iron stores by MRI. *Lancet* 2004 Jan 31;363(9406):357-62.
16. Agency for Healthcare Research and Quality. *Methods Reference Guide for Effectiveness and Comparative Effectiveness Reviews, Version 1.0.* 2010.
17. Glas AS, Lijmer JG, Prins MH, et al. The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol* 2003 Nov;56(11):1129-35.
18. Baldus M, Walter H, Thies K, et al. Transferrin receptor assay and zinc protoporphyrin as markers of iron-deficient erythropoiesis in end-stage renal disease patients. *Clinical Nephrology* 1998 Mar;49(3):186-92.
19. Bhandari S, Norfolk D, Brownjohn A, et al. Evaluation of RBC ferritin and reticulocyte measurements in monitoring response to intravenous iron therapy. *American Journal of Kidney Diseases* 1997 Dec;30(6):814-21.

20. Bhandari S, Turney JH, Brownjohn AM, et al. Reticulocyte indices in patients with end stage renal disease on hemodialysis. *Journal of Nephrology* 1998 Mar;11(2):78-82.
21. Bovy C, Gothot A, Delanaye P, et al. Mature erythrocyte parameters as new markers of functional iron deficiency in haemodialysis: sensitivity and specificity. *Nephrology Dialysis Transplantation* 2007 Apr;22(4):1156-62.
22. Buttarello M, Pajola R, Novello E, et al. Diagnosis of iron deficiency in patients undergoing hemodialysis. *American Journal of Clinical Pathology* 2010 Jun;133(6):949-54.
23. Chuang CL, Liu RS, Wei YH, et al. Early prediction of response to intravenous iron supplementation by reticulocyte haemoglobin content and high-fluorescence reticulocyte count in haemodialysis patients. *Nephrology Dialysis Transplantation* 2003 Feb;18(2):370-77.
24. Cullen P, Soffker J, Hopfl M, et al. Hypochromic red cells and reticulocyte haemoglobin content as markers of iron-deficient erythropoiesis in patients undergoing chronic haemodialysis. *Nephrology Dialysis Transplantation* 1999 Mar;14(3):659-65.
25. Daschner M, Mehls O, Schaefer F. Soluble transferrin receptor is correlated with erythropoietin sensitivity in dialysis patients. *Clinical Nephrology* 1999 Oct;52(4):246-52.
26. Domrongkitchaiporn S, Jirakranont B, Atamasrikul K, et al. Indices of iron status in continuous ambulatory peritoneal dialysis patients. *American Journal of Kidney Diseases* 1999 Jul;34(1):29-35.
27. Eguchi A, Tsuchiya K, Tsukada M, et al. [Clinical usefulness of reticulocyte hemoglobin equivalent (RET-He) in patients at the pre-dialysis stage and in patients on peritoneal dialysis]. [Japanese]. *Nippon Jinzo Gakkai Shi* 2010;Japanese(2):132-40.
28. Fishbane S, Galgano C, Langley RC, Jr., et al. Reticulocyte hemoglobin content in the evaluation of iron status of hemodialysis patients. *Kidney International* 1997 Jul;52(1):217-22.
29. Fusaro M, Munaretto G, Spinello M, et al. Soluble transferrin receptors and reticulocyte hemoglobin concentration in the assessment of iron deficiency in hemodialysis patients. *Journal of Nephrology* 2005 Jan;18(1):72-79.
30. Gupta M, Kannan M, Gupta S, et al. Contribution of iron deficiency to anemia in chronic renal failure. *Indian Journal of Pathology & Microbiology* 2003 Oct;46(4):563-64.
31. Hukic F, Nuhbegovic S, Brkic S, et al. Biochemical markers of iron status in hemodialysis patients. *Medicinski Arhiv* 2010;64(4):219-22.
32. Kim JM, Ihm CH, Kim HJ. Evaluation of reticulocyte haemoglobin content as marker of iron deficiency and predictor of response to intravenous iron in haemodialysis patients. *International Journal of Laboratory Hematology* 2008 Feb;30(1):46-52.
33. Matuszkiewicz-Rowinska J, Ostrowski G, Niemczyk S, et al. [Red cell zinc protoporphyrin and its ratio to serum ferritin (ZPP/logSF index) in the detection of iron deficiency in patients with end-stage renal failure on hemodialysis]. [Polish]. *Polskie Archiwum Medycyny Wewnetrznej* 2003 Jul;110(1):703-10.
34. Mitsuiki K, Harada A, Miyata Y. Assessment of iron deficiency in chronic hemodialysis patients: investigation of cutoff values for reticulocyte hemoglobin content. *Clinical & Experimental Nephrology* 2003 Mar;7(1):52-57.

35. Mittman N, Sreedhara R, Mushnick R, et al. Reticulocyte hemoglobin content predicts functional iron deficiency in hemodialysis patients receiving rHuEPO. *American Journal of Kidney Diseases* 1997 Dec;30(6):912-22.
36. Miwa N, Akiba T, Kimata N, et al. Usefulness of measuring reticulocyte hemoglobin equivalent in the management of haemodialysis patients with iron deficiency. *International Journal of Laboratory Hematology* 2010 Apr;32(2):248-55.
37. Tessitore N, Solero GP, Lippi G, et al. The role of iron status markers in predicting response to intravenous iron in haemodialysis patients on maintenance erythropoietin. *Nephrology Dialysis Transplantation* 2001 Jul;16(7):1416-23.
38. Tessitore N, Girelli D, Campostrini N, et al. Hepcidin is not useful as a biomarker for iron needs in haemodialysis patients on maintenance erythropoiesis-stimulating agents. *Nephrology Dialysis Transplantation* 2010 Dec;25(12):3996-4002.
39. Tsuchiya K, Okano H, Teramura M, et al. Content of reticulocyte hemoglobin is a reliable tool for determining iron deficiency in dialysis patients. *Clinical Nephrology* 2003 Feb;59(2):115-23.
40. Van Wyck DB, Roppolo M, Martinez CO, et al. A randomized, controlled trial comparing IV iron sucrose to oral iron in anemic patients with nondialysis-dependent CKD. *Kidney International* 2005 Dec;68(6):2846-56.
41. Fishbane S, Lynn RI. The utility of zinc protoporphyrin for predicting the need for intravenous iron therapy in hemodialysis patients. *American Journal of Kidney Diseases* 1995 Mar;25(3):426-32.
42. Kaneko Y, Miyazaki S, Hirasawa Y, et al. Transferrin saturation versus reticulocyte hemoglobin content for iron deficiency in Japanese hemodialysis patients. *Kidney International* 2003 Mar;63(3):1086-93.
43. Silva J, Andrade S, Ventura H, et al. Iron supplementation in haemodialysis--practical clinical guidelines. *Nephrology Dialysis Transplantation* 1998 Oct;13(10):2572-77.
44. Fishbane S, Shapiro W, Dutka P, et al. A randomized trial of iron deficiency testing strategies in hemodialysis patients. *Kidney International* 2001 Dec;60(6):2406-11.
45. Singh AK, Coyne DW, Shapiro W, et al. Predictors of the response to treatment in anemic hemodialysis patients with high serum ferritin and low transferrin saturation. *Kidney International* 2007 Jun;71(11):1163-71.
46. Ahluwalia N, Skikne BS, Savin V, et al. Markers of masked iron deficiency and effectiveness of EPO therapy in chronic renal failure. *American Journal of Kidney Diseases* 1997 Oct;30(4):532-41.
47. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making* 1991 Apr;11(2):88-94.
48. Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. *Ann Intern Med* 2003 Jan 7;138(1):40-44.
49. Rothman K, Greenland S. *Modern Epidemiology*. Lippincott Williams and Wilkins; 1998.
50. Whiting P, Rutjes A, Reitsma J, et al. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Medical Research Methodology* 2003;3(1):25. PMID: 10.1186/1471-2288-3-25.

## 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.<sup>1</sup> CKD patients are classified as having progressed to one of five stages, depending on the severity of their condition (CKD stage 1-5).<sup>2</sup> 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.<sup>3</sup> 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.<sup>4</sup>

## 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.<sup>5</sup> Anemia remains common among patients presenting for renal transplantation, and persists in the post-transplant period.<sup>6,7</sup> 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.<sup>8</sup> 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.<sup>9-15</sup> However, 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.<sup>16</sup> Before the development of erythropoietic stimulating agents (ESAs), blood transfusion was the primary treatment option for anemia associated with CKD. 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.<sup>17</sup> 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,<sup>16</sup> as there are concerns regarding the adverse effects associated with elevated doses of ESAs<sup>18</sup> and supplemental iron.<sup>19</sup>

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.<sup>5,20</sup> 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  $\times$  100), and serum ferritin (referred to as “ferritin”) concentrations every 3 months. 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).<sup>5</sup> 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);<sup>16</sup> however, taking a bone marrow sample is invasive and carries the risks of infection or bleeding at the biopsy site.<sup>21</sup> 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 a predictors of iron deficiency or iron overload when it is either extremely low (TSAT) or extremely high (ferritin).<sup>20</sup>

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.<sup>22-24</sup> 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.<sup>25,26</sup>

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.<sup>21</sup> 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 blockage (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.<sup>24</sup>

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 not influenced by the underlying state of inflammation in CKD, and their measurement more accurately reflects the state of iron supply and demand, as compared to older markers.<sup>24</sup>

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)<sup>27</sup>—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.<sup>28</sup> 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.<sup>29</sup>

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



## 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.

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 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.

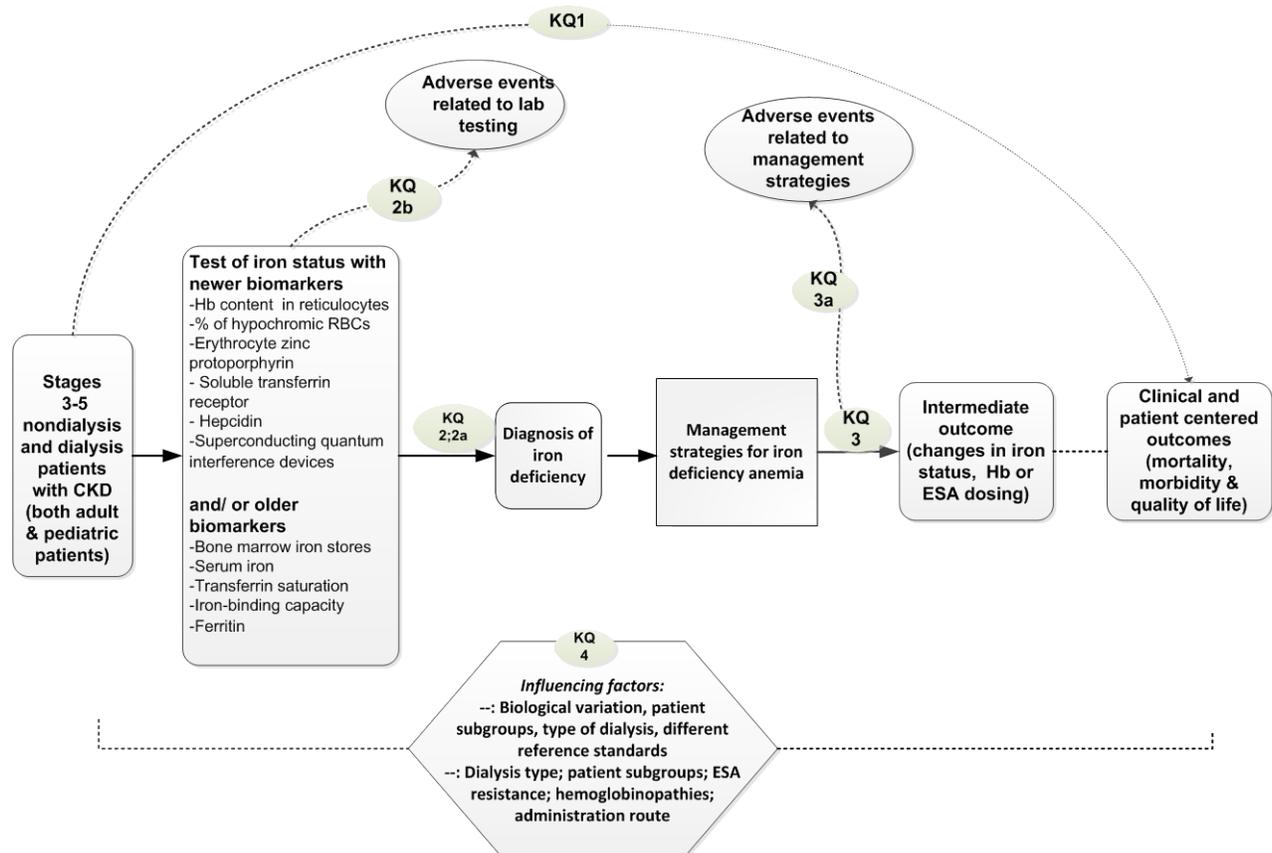
Key Question 1 subsumes Key Questions 2, 3 and 4, which collectively address the impact on patient centered outcomes of using the newer laboratory biomarkers as a replacement for or in addition to classical laboratory biomarkers of iron status for assessing and management of iron deficiency. Specifically, 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<sup>3</sup> as a replacement for or an add-on to the older laboratory biomarkers of iron status<sup>4</sup> 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<sup>a</sup> as a replacement for or an add-on to the older markers<sup>b</sup> in stages 3-5 CKD patients nondialysis and dialysis, and in patients with a kidney transplant?

<sup>3</sup> Content of hemoglobin [Hb] in reticulocytes, percentage of hypochromic red blood cells, erythrocyte zinc protoporphyrin, soluble transferrin receptor, hepcidin, and superconducting quantum interference devices

<sup>4</sup> Bone marrow iron stores, serum iron, transferrin saturation, iron-binding capacity, and ferritin

- c. What reference standards are used for the diagnosis of iron deficiency in studies evaluating test performance?
- d. 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)

## **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 Review*, hereafter referred to as “the Methods Guide” (available at <http://www.effectivehealthcare.ahrq.gov/methodsguide.cfm>).<sup>33</sup> The main sections in this chapter reflect the elements of the protocol established for the CER; certain methods map to the PRISMA checklist.<sup>34</sup> 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 July 2011) and the Cochrane Central Register of Controlled Trials (through the third quarter of 2011). 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 grey literature or unpublished studies.

We also screened the reference lists of related guidelines and selected narrative reviews and primary articles for additional articles.

## Inclusion and Exclusion Criteria

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

**Table A. Study eligibility criteria**

Key Question/PICO	Inclusion Criteria
<b>Key Question 1 (overarching question)</b>	
<b>Populations</b>	<ul style="list-style-type: none"> <li>• Pediatric and adult nondialysis patients with stage 3, 4, or 5 CKD</li> <li>• Patients with CKD undergoing dialysis (hemo- or peritoneal dialysis)</li> <li>• Patients with a kidney transplant</li> </ul>
<b>Interventions</b>	<ul style="list-style-type: none"> <li>• Newer laboratory biomarkers<sup>c</sup> to assess iron status and manage iron deficiency either as a replacement for or in addition to older laboratory biomarkers</li> </ul>
<b>Comparators</b>	<ul style="list-style-type: none"> <li>• Older laboratory biomarkers<sup>d</sup> to assess iron status and manage iron deficiency</li> </ul>
<b>Outcomes</b>	<ul style="list-style-type: none"> <li>• Mortality</li> <li>• Morbidity (e.g., cardiac or liver toxicity and infection)</li> <li>• 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).</li> <li>• 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)</li> </ul>
<b>Study designs</b>	<ul style="list-style-type: none"> <li>• Randomized controlled trials</li> <li>• Nonrandomized controlled trials</li> <li>• Observational studies with concurrent comparison groups</li> </ul>
<b>Key Question 2, 3 and 4</b>	
<b>Populations</b>	<ul style="list-style-type: none"> <li>• Pediatric and adult nondialysis patients with stage 3, 4, or 5 CKD</li> <li>• Patients with CKD undergoing dialysis (hemo- or peritoneal dialysis)</li> <li>• Patients with a kidney transplant</li> </ul>
<b>Interventions</b>	<ul style="list-style-type: none"> <li>• Newer laboratory biomarker alone<sup>a</sup> or in combination with older laboratory biomarkers of iron status<sup>b</sup>.</li> </ul>
<b>Comparators</b>	<ul style="list-style-type: none"> <li>• Older laboratory biomarkers of iron status, which include bone marrow iron stores, serum iron, transferrin saturation, iron-binding capacity, and ferritin.</li> </ul>
<b>Outcomes</b>	<p>Key Question 2 and 4:</p> <ul style="list-style-type: none"> <li>• 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 non-response to treatment.</li> <li>• Adverse effects or harms associated with laboratory testing</li> </ul> <p>Key question 3 and 4:</p> <ul style="list-style-type: none"> <li>• Intermediate outcomes</li> <li>• Increase in Hb or hematocrit, or more consistent maintenance of Hb or hematocrit within the desired range</li> <li>• Use of erythropoiesis-stimulating agent (ESA) for maintenance of Hb within the desired range (stable dose in contrast to escalating dose resulting in net decreased ESA dose in hyporesponsive patients or actual decreased ESA dose in relatively responsive patients)</li> <li>• Adverse effects or harms associated with different management strategies</li> </ul>
<b>Study designs</b>	<p>Key Question 2:</p> <ul style="list-style-type: none"> <li>• Any design</li> </ul>

Key Question/PICO	Inclusion Criteria
	Key Question 3 and 4: <ul style="list-style-type: none"> <li>• Randomized controlled trials</li> <li>• Nonrandomized controlled trials</li> <li>• Observational studies with concurrent comparison groups</li> </ul>
<b>Study settings</b>	<ul style="list-style-type: none"> <li>• Any setting: primary or specialty care, in-facility or home, and inpatient or outpatient.</li> </ul>

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

<sup>b</sup>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 (<http://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.<sup>33</sup> 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]<sup>35</sup> tool for studies of diagnostic performance and the Cochrane risk of bias tool for intervention studies<sup>36</sup>), 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).<sup>37</sup>

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.<sup>16</sup> 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 “function 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 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 versus 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.<sup>38</sup>

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 (<http://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.<sup>39</sup> Below, we list the test performance terms and definitions used in the current report:

- **Receiver operating characteristic curve:** ROC curves compare sensitivity versus 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.<sup>40</sup> 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.<sup>33,33</sup> 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 of our level of confidence that the evidence reflected the true effect for the major comparisons of interest.<sup>33</sup> 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,<sup>33</sup> 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

[To be added]

# 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.

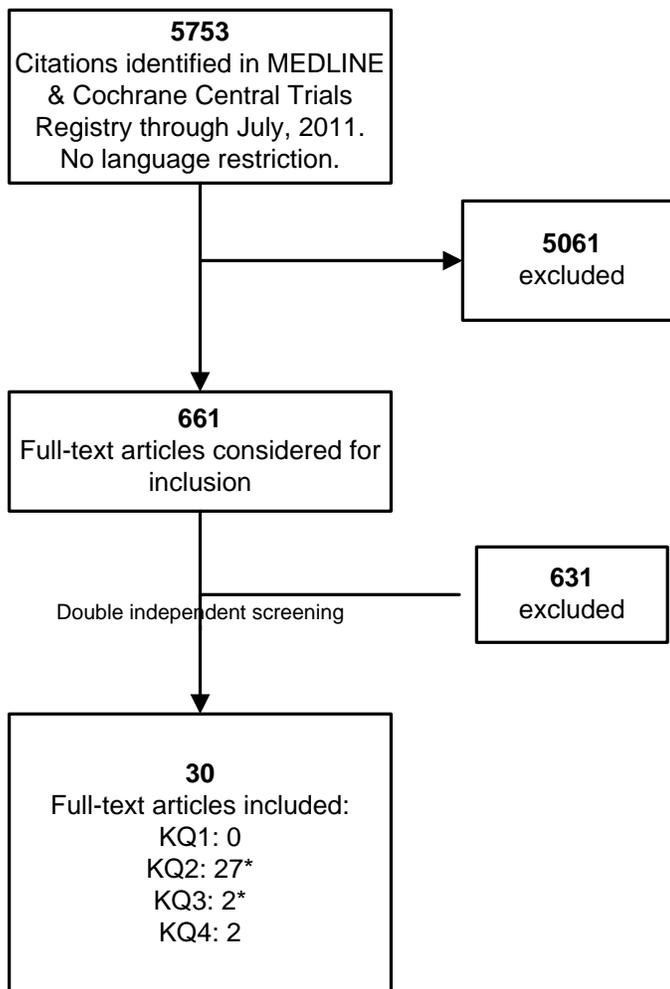
## Literature Searches

The literature search yielded 5753 citations. From these, 661 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 631 articles were rejected on double, independent full-text screening because they did not meet one or more of the PICO criteria for a particular KQ (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,<sup>41-70</sup> 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),<sup>41-67</sup> 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),<sup>66,70</sup> 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).<sup>68,69</sup> 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.<sup>41,64,68,70</sup> Eight studies received funding from non-profit sources, such as national kidney training fellowships,<sup>43,59</sup> internal university hospital grant,<sup>52</sup> academic foundation grant,<sup>56</sup> or government funding.<sup>42,47,62,66</sup>

Detailed characteristics of included studies are presented later with results for each Key Question.

**Figure 3: Literature flow**

\* Total for articles included in the key questions do not add up to 30 because one study<sup>66</sup> contributed to both Key question 2 and Key question 3.

## 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 to 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.<sup>41-43,45,48,49,51,53-57,60,63,66</sup> These studies used the following definitions: 1) TSAT  $\leq$  15%;<sup>48</sup> 2) TSAT  $\leq$  20%;<sup>41-43,53,57,63,66</sup> 3) ferritin  $\leq$  100 ng/mL;<sup>43</sup> 4) TSAT  $\leq$  20% and ferritin  $\leq$  100 ng/mL;<sup>49,51,53-55,63</sup> 5) TSAT  $\leq$  20% or ferritin  $\leq$  100 ng/mL;<sup>51,56,60,66</sup> 6) serum iron  $<$  40  $\mu$ g/dL, TSAT  $<$  20%, ferritin  $<$  100 ng/mL, and Hb  $<$  11 g/dL;<sup>45</sup> 7) TSAT  $<$  20%, ferritin 100-800 ng/mL, and Hb  $<$  11 g/dL;<sup>45</sup> and 8) TSAT  $<$  16% and ferritin  $<$  12 ng/mL.<sup>54</sup> 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.<sup>44,46,47,50,52,58,59,61,62,64,65,67</sup> However, there exists a high degree of heterogeneity in the reference standards used across studies as well (details are described later in **Table 2.1**). The most commonly used definition for a response to IV iron treatment was an increase in hemoglobin (Hb) concentration  $\geq$  1 g/dL after a (variable) period of IV iron treatment.<sup>44,46,62,64,67</sup> Other reference standards include a  $\geq$  15 percent increase in Hb,<sup>61</sup> an increase in Hct of  $\geq$  3 percent and/or a  $\geq$  30 percent reduction in EPO dose,<sup>47</sup>  $>$  1 point increase in corrected reticulocyte index,<sup>52</sup> and 5% increase in Hct or a decrease in EPO dose of  $>$  2000 units per treatment.<sup>65</sup> 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 versus newer biomarkers in predicting a response. Thus, the results from these studies are 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,<sup>44,46,52,58,59,61,62</sup> and one study enrolled adult nondialysis (ND CKD) patients.<sup>64</sup> The remaining four studies investigated the test performance of newer laboratory markers alone. Of these four, three enrolled adult HD CKD patients,<sup>47,52,67</sup> and one enrolled adult peritoneal dialysis (PD CKD) patients.<sup>50</sup> None of the reviewed studies enrolled pediatric CKD patients, and we did not include studies evaluating the test performance of classical markers alone.

**Table 2.1** 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  $\geq 1$  g/dL after a period of IV iron treatment (**Table 2.1**). 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 2.2**). 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 2.3**. 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,<sup>44,46,58,59,61,62,65</sup> and three additional studies that investigated the test performance of newer laboratory markers alone.<sup>47,52,67</sup> The results from each of the single studies examining adult ND CKD patients<sup>64</sup> and adult PD CKD patients<sup>50</sup> were not plotted in the ROC space.

### Summary of key points (Table 2.1 to 2.3; 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 2.1. 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**

Study, Year [UI]	Population	Total N enrolled	IV Iron Treatment	Reference Standard (Response to IV Iron Therapy)	Ferritin	TSAT	CHr	RetHe	%HYPO	ZPP	sTfR	Hepcidin
<b>Studies investigating both newer and classical markers</b>												
Bovy, 2007 <sup>44</sup> [17237481]	HD CKD	32	IV iron sucrose (1200 mg total) - 100 mg at the end of dialysis session over 4 wks	≥1 g/dL increase in Hb during the 4-week IV iron Tx	√	√	√		√		√	
Buttarello, 2010 <sup>46</sup> [20472854]	HD CKD	69	IV iron gluconate and α-darbepoetin to maintain Hb between 11.0 & 12.0 g/dL	≥1 g/dL increase in Hb at any time after the third wk of IV iron Tx	√	√	√	√	√			
Fishbane, 1995 <sup>65</sup> [7872320]	HD CKD	62	1,000 mg IV iron dextran in 100 mg doses over 10 sequential HD Tx	5% increase in Hct or a decrease in EPO dose of >2000 units/ treatment over 3-6 mths	√	√				√		
Mitsuiki, 2003 <sup>58</sup> [14586744]	HD CKD	27	40 mg of chondroitin sulfate-iron colloid IV once a wk after the regular dialysis session	Change in Hct ≥3% (or change in Hb ≥1 g/dL) within 8 wks after IV iron Tx	√	√	√					
Mittman, 1997 <sup>59</sup> [9398141] US	HD CKD	79	Single bolus of 500 mg IV iron dextran over 2 hours during a regular hemodialysis session	>1 point increase in corrected reticulocyte index at any point during the 2 wks after IV iron Tx	√	√	√					
Tessitore, 2001 <sup>61</sup> [11427634]	HD CKD	125	IV sodium ferric gluconate complex in sucrose as a slow (2 min) IV bolus at end of dialysis	≥15% increase in Hb at any 2 consecutive measurements (evaluated every 2 wks)	√	√	√		√*	√	√	

Study, Year [UI]	Population	Total N <sub>enrolled</sub>	IV Iron Treatment	Reference Standard (Response to IV Iron Therapy)	Ferritin	TSAT	CHr	RetHe	%HYPO	ZPP	sTfR	Hepcidin
			with 31 or 62 mg iron as per predialysis serum transferrin (< or > 170 mg/dL, respectively)									
Tessitore, 2010 <sup>62</sup> [20538788]	HD CKD	56	1 g intravenous iron (62.5 mg ferric gluconate at 16 consecutive dialysis sessions)	≥1 g/dL increase in Hb after 6 wks IV iron treatment	√	√	√		√*			√
Van Wyck, 2005 <sup>64</sup> [16316362]	ND CKD	95	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.	≥ 1 g/dL increase in Hb after 8 wks IV iron Tx	√	√	√					
<b>Studies investigating newer markers alone</b>												
Chuang, 2003 <sup>47</sup> [12543894]	HD CKD	95	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	Rise in Hct of ≥3% or a reduction in rHuEpo dose of ≥30% over the baseline values at the end of the study			√					
Fishbane, 1997 <sup>52</sup> [9211366]	HD CKD	50	1,000 mg of IV iron dextran infused over two hours as a single-dose infusion	1 point increase in the corrected reticulocyte index within two wks of IV iron Tx			√		√			
Silva, 1998 <sup>67</sup> [9794562]	HD CKD	33	IV Iron saccharate 20 mg diluted in 10 mL saline, and	≥ 1 g/dL increase in Hb during the 6 mths of IV iron Tx					√			

Study, Year [UI]	Population	Total N <sub>enrolled</sub>	IV Iron Treatment	Reference Standard (Response to IV Iron Therapy)	Ferritin	TSAT	Chr	RetHe	%HYPO	ZPP	sTfR	Hepcidin
			given in last 10 minutes of dialysis									
Domrongkitchaiporn, 1999 <sup>50</sup> [10401012]	PD CKD	23	IV iron - 1000 mg ferric saccharate-infused over 2 hours in two divided doses 1 wk apart	Sustained >1 g/dL increase in Hb within 3 mths of .IV iron T			√					

%HYPO=percent of hypochromic 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; PD=peritoneal Dialysis; RetHe=reticulocyte hemoglobin equivalent; rHuEpo=recombinant human erythropoietin; sTfR=soluble transferrin receptor; ZPP=erythrocyte zinc protoporphyrin; TSAT=transferrin saturation; Tx=treatment; UI=universal identifier/Pubmed ID; wk=week

√ = marker was investigated

√\* = best predictors of iron deficiency among all other markers

**Table 2.2. Characteristics of studies evaluating the ability of newer or classical markers of iron status to predict the response to IV iron treatment**

Study, Year [UI] Country	Study Design Recruitment Method	Sampling Population	N <sub>enrolled</sub> / N <sub>analyzed</sub>	Demographics	Anemia and Iron Status Indices	Background Treatment	Risk of Bias
<b>Studies investigated both newer and classical markers</b>							
Bovy, 2007 <sup>44</sup> [17237481] 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 <sup>46</sup> [20472854] 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

Study, Year [UI] Country	Study Design Recruitment Method	Sampling Population	N <sub>enrolled</sub> / N <sub>analyzed</sub>	Demographics	Anemia and Iron Status Indices	Background Treatment	Risk of Bias
Fishbane, 1995 <sup>65</sup> [7872320] US	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 <sup>58</sup> [14586744] 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 <sup>59</sup> [9398141] US	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 <sup>61</sup> [11427634] 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
Tessitore, 2010 <sup>62</sup> [20538788] Italy	Prospective cohort Selected sample	HD CKD	56/56	Male (%): 57 Age (yr): 67 Race (%): NR	Hb (g/dL): 11.6 Hct (%): NR ferritin (ng/mL):146 TSAT (%): 20	ESA dose: 8000 IU/wk  Iron washout: 10 wks	Low
<b>Studies investigated newer markers alone</b>							
Chuang, 2003 <sup>47</sup> [12543894] Taiwan	Prospective Cohort Selected sample	HD CKD	95/65	Male (%): 51 Age (yr): 60 Race (%): NR	Hb (g/dL): 9.8 Hct (%): 30.1 ferritin (ng/mL): 244 TSAT (%): 38.5	ESA dose: 90 IU/wk/kg  Iron washout: 12 wks	High
Fishbane, 1997 <sup>62</sup> [9211366] US	Prospective cohort Random sampling	HD CKD	50/32	Male (%): NR Age (yr): NR Race (%): NR	Hb (g/dL): NR Hct (%): 32.7 ferritin (ng/mL): 231 TSAT (%): NR	ESA dose: NR  Iron washout: 4 wks	High

Study, Year [UI] Country	Study Design Recruitment Method	Sampling Population	N <sub>enrolled</sub> / N <sub>analyzed</sub>	Demographics	Anemia and Iron Status Indices	Background Treatment	Risk of Bias
Silva, 1998 <sup>67</sup> [9794562] Portugal	Prospective cohort Selected sample	HD CKD	33/33	Male (%): 61 Age (yr): 58 Race (%): NR	Hb (g/dL):10.8 Hct (%): NR ferritin (ng/mL):137 TSAT (%): 27	ESA dose: 118.2 IU/kg/wk  Iron washout: NR (61% patients received oral iron)	High
Van Wyck, 2005 <sup>64</sup> [16316362] US	Prospective cohort	ND CKD (stage 3-5)	95/79	Male (%): 33 Age (yr): 62 Race (%): Caucasian-56 Black-38 Other-6	Hb (g/dL): 10.2 Hct (%): NR ferritin (ng/mL): 92.6 TSAT (%): 16.4	ESA dose: NR  Iron washout: 24 wks	Medium
Domrongkitchaiporn, 1999 <sup>50</sup> [10401012] Thailand	Prospective Cohort Selected sample	PD CKD	23/21	Male (%): 67 Age (yr): 51 Race (%): NR	Hb (g/dL): 8.4 Hct (%): NR ferritin (ng/mL): 643 TSAT (%): 33.9	ESA dose: 71 IU/wk/kg  Iron washout: 4 wks	Medium

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; NR=not reported; TSAT=transferrin saturation; UI=universal identifier/Pubmed ID; wk=week; yr=year

**Table 2.3. 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<sup>a</sup> in seven cohort studies among adult HD CKD patients**

Iron Status Marker	Total Number of Studies (Total N) [Risk of Bias]	Overall Test Accuracy When Compared with TSAT	Sensitivity and Specificity When Compared with TSAT <20%	Overall Test Accuracy When Compared with Ferritin	Sensitivity and Specificity When Compared with Ferritin <100 ng/mL	Sensitivity and Specificity when Compared with TSAT <20% or Ferritin <100 ng/mL	Other Comparative Results
<b>CHr /RetHe</b>	6 CHr studies <sup>44,46,58,59,61,62</sup> (388) [1 low, <sup>62</sup> 4 medium, <sup>44,46,58,59</sup> 1 high risk <sup>61</sup> ]	NS difference (4 studies) <sup>44,61,62</sup> CHr better (2 study) <sup>46,58</sup>	CHr <30 or <29 pg worse (1 study) <sup>44</sup> CHr <27 or <28 pg better (1 study) <sup>59</sup>	NS difference (2 studies) <sup>44,62</sup> CHr better (3 study) <sup>46,58,61</sup>	CHr <29 pg worse (1 study) <sup>44</sup> CHr <30 pg better (1 study) <sup>44</sup> CHr <27 or <28 pg better (1 study) <sup>59</sup>	CHr <30 or <29 pg worse (1 study) <sup>44</sup> CHr <27 or <28 pg better (1 study) <sup>59</sup>	Combination of %HYPO >6% with CHr ≤29 pg produced minor improvement in sensitivity and specificity (1 study) <sup>61</sup>
	1 RetHe study <sup>46</sup>	RetHe better (1					NS difference between RetHe

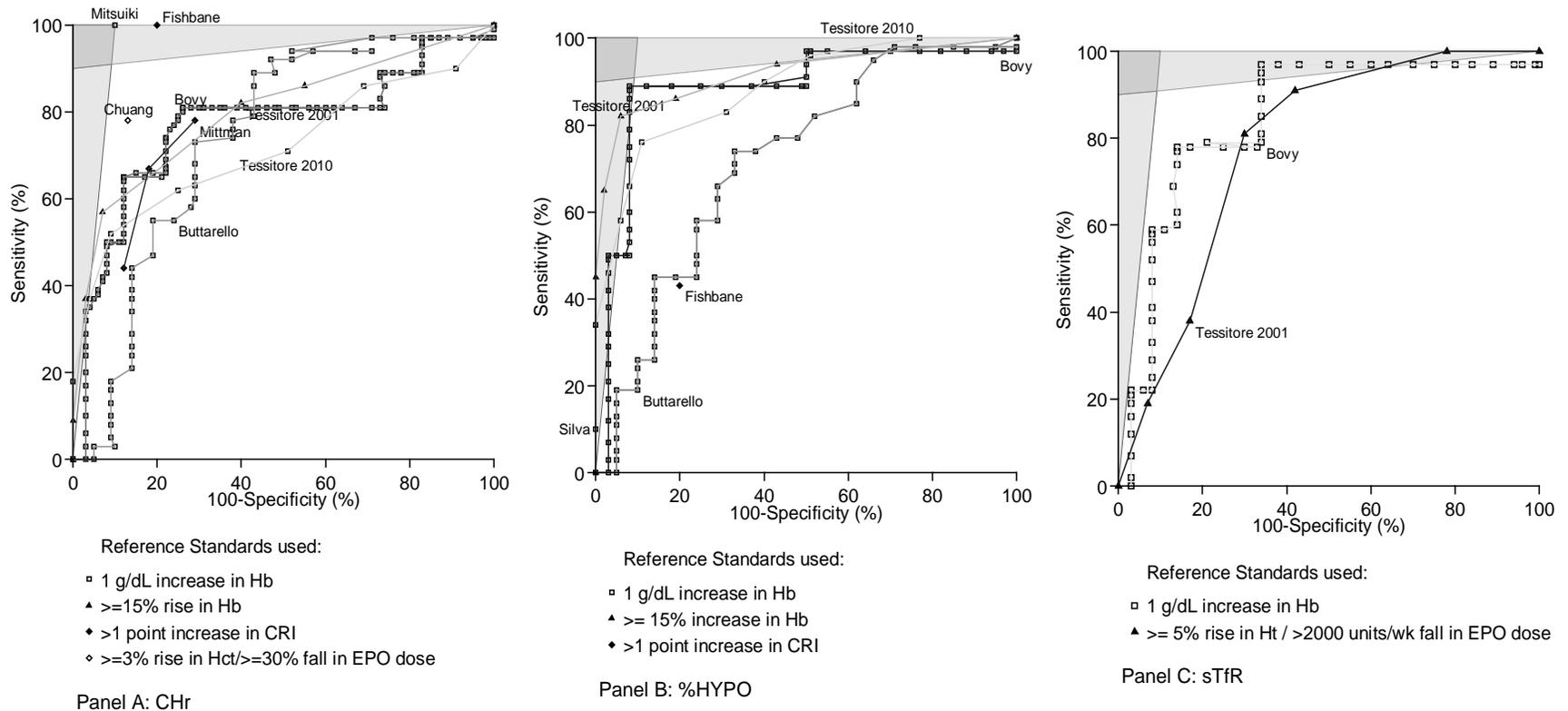
Iron Status Marker	Total Number of Studies (Total N) [Risk of Bias]	Overall Test Accuracy When Compared with TSAT	Sensitivity and Specificity When Compared with TSAT <20%	Overall Test Accuracy When Compared with Ferritin	Sensitivity and Specificity When Compared with Ferritin <100 ng/mL	Sensitivity and Specificity when Compared with TSAT <20% or Ferritin <100 ng/mL	Other Comparative Results
	(69) [1 medium risk <sup>46</sup> ]	study) <sup>46</sup>					and CHr (1 study) <sup>46</sup>
<b>%HYPO</b>	4 studies <sup>44,46,61,62</sup> (282) [1 low, <sup>62</sup> 2 medium, <sup>44,46</sup> 1 high risk <sup>61</sup> ]	NS difference (1 study) <sup>44</sup> %HYPO better (3 studies) <sup>46,61,62</sup>	%HYPO >10% better (1 study) <sup>44</sup> %HYPO >6% better (1 study) <sup>61</sup>	NS difference (1 study) <sup>44</sup> %HYPO better (3 studies) <sup>46,61,62</sup>	%HYPO >10% better (1 study) <sup>44</sup> %HYPO >6% better (1 study) <sup>61</sup>	%HYPO >10% better (1 study) <sup>44</sup> %HYPO >6% better (1 study) <sup>61</sup>	%HYPO was the only significant predictor of a response to IV iron treatment among all other markers <sup>b</sup> (2 study) <sup>61,62</sup>  Combination of % HYPO >6 with TSAT≤20% produced a substantial increase in sensitivity but reduce in specificity (1 study) <sup>61</sup>
<b>ZPP</b>	2 studies <sup>61,65</sup> (187) [2 high risk <sup>61,65</sup> ]	NS difference (1 study) <sup>61</sup> ZPP better (1 study) <sup>65</sup>	ZPP >90 μmol/mol better (1 study) <sup>65</sup> ZPP >52 μmol/mol better (1 study) <sup>61</sup>	NS difference (1 study) <sup>61</sup> ZPP better (1 study) <sup>65</sup>	ZPP >90 μmol/mol better (1 study) <sup>65</sup> ZPP >52 μmol/mol better (1 study) <sup>61</sup>	ZPP >52 μmol/mol better (1 study) <sup>61</sup>	Combination of % HYPO >6 with ZPP >52 μmol/mol produced a substantial increase in sensitivity but reduce in specificity (1 study) <sup>61</sup>
<b>sTfR</b>	2 studies <sup>44,61</sup> (157) [1 medium, <sup>44</sup> 1 high risk <sup>61</sup> ]	NS difference (2 studies) <sup>44,61</sup>	sTfR >1.5 pg better (1 study) <sup>61</sup>	NS difference (2 studies) <sup>44,61</sup>	sTfR >1.5 pg better (1 study) <sup>61</sup>	sTfR >1.5 pg better (1 study) <sup>61</sup>	
<b>Hepcidin</b>	1 study <sup>62</sup> (56) [1 low risk <sup>62</sup> ]	NS difference (1 study) <sup>62</sup>		NS difference (1 study) <sup>62</sup>			

AUC=area under the curve; IV=intravenous; NS=not significant

<sup>a</sup> Response to IV iron treatment (the reference standard) was defined variably across studies (see also **Table 2.21**)

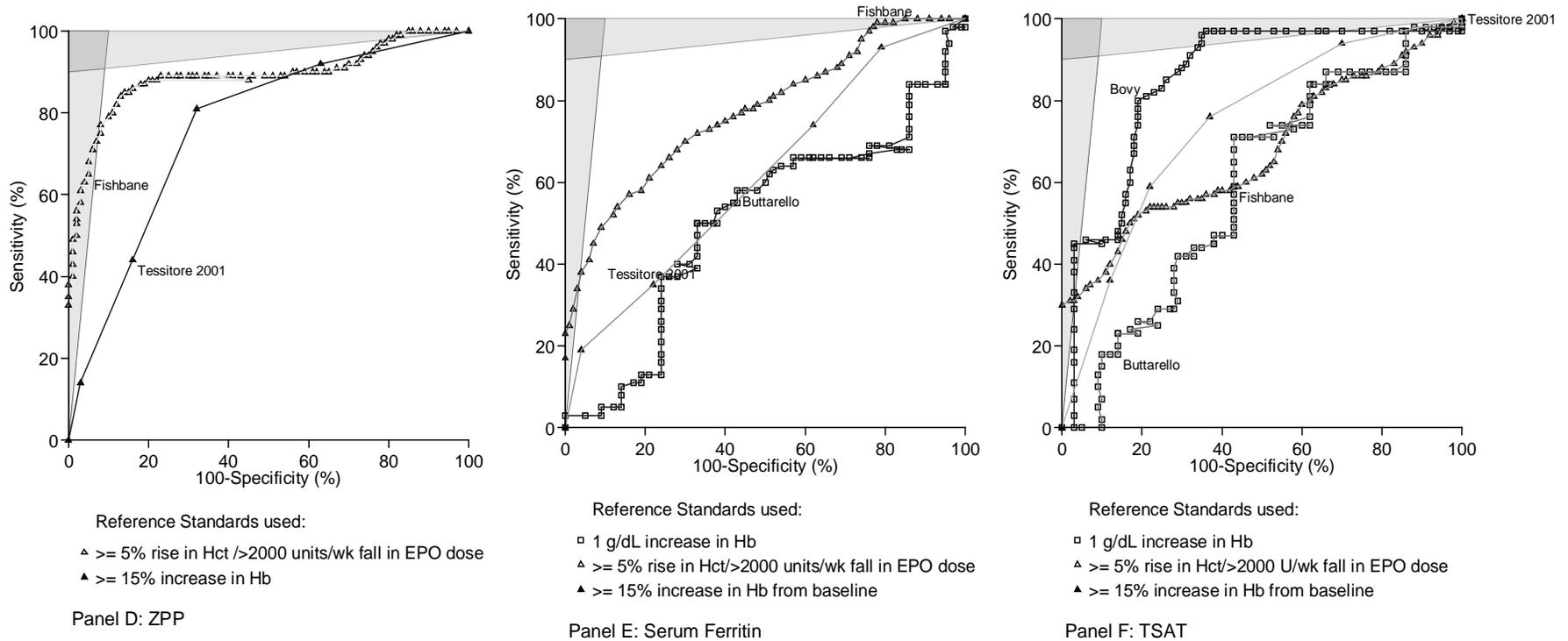
<sup>b</sup> The multivariate logistic regression analysis included HFE genotype, ferritin, TSAT, %Hypo, CHr, Hep-25 and Hep-20 in the same model.

**Figure 4. Indirect comparisons of the overall test accuracy of newer versus classical markers of iron status (at baseline) to predict response to IV iron among adult HD CKD patients – CHR, %HYPO, sTfR**



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.

**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**



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

### Key Points (Table 2.4)

Eight cohort studies enrolling 533 adult HD CKD patients,<sup>44,46,47,52,58,59,61,62</sup> one cohort study enrolling 23 PD CKD patients,<sup>50</sup> and one cohort study enrolling 95 ND CKD patients<sup>64</sup> 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. 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).<sup>44,59</sup> Only one study performed multivariate analyses to predict a response to IV iron treatment (defined as an increase in Hct of  $\geq 3$  percent and/or a  $\geq 30$  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).<sup>47</sup>

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 2.4. 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**

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

CKD=chronic kidney disease; HD=hemodialysis; ND=nondialysis; PD=peritoneal dialysis

## Detailed Synthesis (Tables 2.2 and 2.5)

### HD CKD patients

Eight studies evaluated the test performance of CHr to predict a IV response in 533 adult CKD patients.<sup>44,46,47,52,58,59,61,62</sup> 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.<sup>46</sup> 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, chondritin-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.<sup>47,59</sup>

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  $\geq 1$  g/dL;<sup>44,46,58,62</sup> 2) a  $\geq 15$  percent increase in Hb;<sup>61</sup> 3) an increase in Hct of  $\geq 3$  percent and/or a  $\geq 30$  percent reduction in EPO dose;<sup>47</sup> and 4) >1 point increase in corrected reticulocyte index.<sup>52,59</sup> 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,<sup>62</sup> six at a medium risk of bias,<sup>44,46,58,59</sup> and three at a high risk of bias.<sup>47,52,61</sup> 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 to TSAT and ferritin based on the AUC values (**Table 2.2**). 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  $\geq 3$  percent and/or a  $\geq 30$  percent reduction in EPO dose).<sup>47</sup> 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]).<sup>47</sup> 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.<sup>61</sup>

Only two studies reported the sensitivities and specificities of CHr (at different cutoff values) in comparison to classical markers (TSAT <20 or ferritin <100 ng/mL) to predict iron deficiency (as defined by a response to IV iron treatment).<sup>44,59</sup> 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 that 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  $\geq 1.0$  g/dL of Hb within three months of starting treatment.<sup>50</sup> 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+  $\geq 10$  and LR-  $\leq 0.1$ ) 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  $\geq 1.0$  g/dL.<sup>64</sup> 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 2.5. Summary results of the ability of CHr to predict the response to IV iron treatment**

Study, Year [UI]	Lab Analysis or Assay	N <sub>analyzed</sub> (% Responders)	CHr Cut-off (pg)	Sens, %	Spec, %	AUC (CI)	Ferritin Comparison AUC (CI)	CHr vs. Ferritin P value	TSAT Comparison AUC (CI)	CHr vs. TSAT P value	Other Results
<b>Adult HD CKD</b>											
Tessitore, 2010 <sup>62</sup> [20538788]	PBSCIIc mass spectrometer and copperloaded immobilized metal-affinity capture ProteinChip arrays (IMAC30-Cu2+)	56 (38)	<32	57	75	0.697 (0.537,0.855)	0.552 (0.391, 0.713)	NS	0.593 (0.431, 0.754)	NS	CHr AUC not significantly different from AUC of hepcidin isoforms (P >0.12)
Bovy, 2007 <sup>44</sup> [17237481]	ADVIA 120 cell counter system, Bayer	32 (38)	<29	25	100	0.752 (0.583, 0.921)	0.834 (0.685, 0.983)	NS	0.896 (0.778, 1.0)	NS	NR
Buttarelo, 2010 <sup>46</sup> [20472854]	ADVIA 120 hematology system, Bayer (CHr)	59 (NR)	<31.2	47	83	0.74 (0.60, 0.89)	0.53 (0.38, 0.69)	NR	0.56 (0.40, 0.72)	NR	NR
	XE 5000 (RetHe)	59 (NR)	<30.6	45	83	0.72(0.58,0.86) P<0.003	0.53 (0.38, 0.69)	NR	0.56 (0.40, 0.72)	NR	NR
Mitsuiki, 2003 <sup>58</sup> [14586744]	ADVIA 120 hematology system, Bayer	27 (63)	<32	100	90	0.95 (0.89,1.00)	0.591 (0.415, 0.767)	NR	0.676 (0.474, 0.878)	NR	NR

Study, Year [UI]	Lab Analysis or Assay	N <sub>analyzed</sub> (% Responders)	CHr Cut-off (pg)	Sens, %	Spec, %	AUC (CI)	Ferritin Comparison AUC (CI)	CHr vs. Ferritin P value	TSAT Comparison AUC (CI)	CHr vs. TSAT P value	Other Results
Mittman, 1997 <sup>59</sup> [9398141]	Technicon H3RTC Hematology Analyzer, Bayer Diagnostic	79 (59)	<26	44	88	NR	NA	NA	NA	NA	NR
			<27	67	82						
			<28	78	71						
			Change in CHr from baseline to 2 wks >2 pg	100	31						
			Change in CHr from baseline to 2 wks >2.5 pg	89	40						
Chuang, 2003 <sup>47</sup> [12543894]	Technicon H*3 automated cell counter, Bayer Laboratory	65 (65)	<28	78	87	NR	NA	NA	NA	NA	OR=29.04 (5.36,157.33) with the best cutoff <28 pg
			Change in CHr from baseline to 2 wks >1.2 pg	80	83	NR	NA	NA	NA	NA	OR=27.85 (5.37,144.3) with the best cutoff >1.2 pg
			Change in CHr from baseline to 4 wks >1.2 pg	87	83	NR	NR	NR	NR	NR	OR=6.2 (1.94,19.8, P=0.002) with a cut off >1.2 pg
Fishbane, 1997 <sup>52</sup> [9211366]	Technicon* H3, Bayer Laboratory	32	<26	100	80	NR	NR	NR	NR	NR	NR

Study, Year [UI]	Lab Analysis or Assay	N <sub>analyzed</sub> (% Responders)	CHr Cut-off (pg)	Sens, %	Spec, %	AUC (CI)	Ferritin Comparison AUC (CI)	CHr vs. Ferritin P value	TSAT Comparison AUC (CI)	CHr vs. TSAT P value	Other Results
Tessitore, 2001 <sup>61</sup> [11427634]	Advia 120 Hematology Analyser, Bayer Diagnostics	125 (41)	≤29	57	93	0.798 (0.714, 0.880)	0.633 (0.514, 0.752)	P<0.05	0.753 (0.669, 0.837)	NS	NR
<b>Adult PD CKD</b>											
Domrongkitchaiporn, 1999 <sup>50</sup> [10401012]	Technicon * H3, Bayer Laboratory	21 (71)	<28	20	83	NR	NR	NR	NR	NR	NR
			<29	47	83						
			<31	53	66						
<b>Adult ND CKD</b>											
Van Wyck, 2005 <sup>64</sup> [16316362] US	NR	35 (44)	< 25	0	97	NR	NR	NR	NR	NR	“Baseline TSAT, ferritin, and CHr are poor predictors of response to IV iron” as shown in the ROC curves
			< 27	3	92						
			< 29	12	86						
			< 31	33	84						
			< 33	83	57						
			< 35	95	24						

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; UI=universal identifier/Pubmed ID

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

<sup>b</sup>Reticulocyte count multiplied by the hematocrit divided by 40.

## Percent Hypochromic Red Blood Cells

### Key Points (Table 2.6)

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.<sup>44,46,52,61,62,67</sup> 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.<sup>61,62</sup>

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 2.6. 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**

Number of Studies (Total N Analyzed)	Risk of Bias	Consistency	Directness	Precision	Overall Strength of Evidence
6 (356 CKD patients on HD)	1 low risk 2 medium risk 3 high risk studies	Consistent	indirect	Imprecise	Low

CKD=chronic kidney disease; HD=hemodialysis

### Detailed Synthesis (Tables 2.2 and 2.7)

Six cohort studies, enrolling a total of 356 HD CKD patients, evaluated the ability of %HYPO to predict a response to IV iron treatment.<sup>44,46,52,61,62,67</sup> 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,<sup>46</sup> and two studies excluded patients with high normal serum ferritin values.<sup>52,67</sup> 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  $\geq 1$  g/dL after treatment,<sup>44,46,62,67</sup> one study defined response as  $\geq 15$  percent increase in Hb at any two consecutive measurements,<sup>61</sup> and one study defined response as >1 point increase in corrected reticulocyte index within 2 weeks.<sup>52</sup> 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,<sup>62</sup> two at a medium risk,<sup>44,46</sup> and three at a high risk of bias.<sup>52,61,67</sup> 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 to TSAT and ferritin, based on the AUC values.<sup>46,61,62</sup> These studies defined a response to IV iron treatment as either an increase in Hb concentration  $\geq 1$  g/dL after treatment,<sup>44,46,62</sup> or  $\geq 15$  percent increase in Hb at any two consecutive measurements.<sup>61</sup> 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).<sup>44,62</sup> 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).<sup>61,62</sup> Combination of markers were assessed in two studies.<sup>44,61</sup> 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.<sup>61</sup> In the other study, the combination of %HYPO and classical markers resulted in lower sensitivity and high specificity.<sup>44</sup>

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 to 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 to ferritin and TSAT in all studies.

**Table 2.7. Summary results of the ability of %HYPO to predict the response to IV iron treatment in HD CKD patients**

Study, Year [UI]	Population	Lab Analysis or Assay	%HYPO Cut-off (%)	N <sub>analyzed</sub> (% responders)	Sens, %	Spec, %	AUC (CI)	Ferritin AUC (CI)	Comparison %HYPO vs. Ferritin P value	TSAT AUC (CI)	Comparison %HYPO vs. TSAT P value	Other Results
Tessitore, 2010 <sup>62</sup> [20538788]	HD CKD	Advia 120 Hematology Analyser	>6	56 (38)	76	89	0.844 (0.737, 0.950)	0.552 (0.391, 0.713)	NS	0.593 (0.431, 0.754)	NS	OR = 1.60 [95% CI=1.08,2.39], P=0.02).
Bovy, 2007 <sup>44</sup> [17237481]	HD CKD	Advia 120 cell counter	> 10	32 (38)	67	95	0.937 (0.837, 1.00)	0.834 (0.685, 0.983)	NS	0.896 (0.778, 1.014)	NS	NR
Buttarelli, 2010 <sup>46</sup> [20472854]	HD CKD	Advia 120	≥ 5.8	59 (NR)	45	87	0.72 (0.58, 0.86)	0.53 (0.38, 0.69)	NR	0.56 (0.40, 0.72)	NR	NR
Fishbane, 1997 <sup>52</sup> [9211366]	HD CKD	Technicon H*3 hematology analyzer	> 10	32 (22)	43	80	NR	NR	NR	NR	NR	NR
Silva, 1998 <sup>67</sup> [9794562]	HD CKD	Technicon Mod. H2 System	> 10	33 (88)	10	100	NR	NR	NR	NR	NR	NR
Tessitore, 2001 <sup>61</sup> [11427634]	HD CKD	Advia 120 Hematology Analyser	> 6	125 (41)	82	95	0.93 (0.884, 0.976)	0.633 (0.514, 0.752)	P<0.001	0.753 (0.669, 0.837)	P<0.05	NR

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

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

<sup>b</sup> Reticulocyte count multiplied by the hematocrit and divided by 40.

## Soluble Transferrin Receptor

### Key Points (Table 2.8)

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.<sup>44,61</sup> 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,<sup>61</sup> and one at a medium risk of bias.<sup>44</sup> The response to IV iron treatment was defined differently in the two studies, either as an increase in Hb concentration  $\geq 1$  g/dL after intravenous iron treatment,<sup>44</sup> or as an increase in Hb  $>15$  percent from baseline.<sup>61</sup>

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 between the two studies) 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 2.8. 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**

Number of Studies (Total N Analyzed)	Risk of Bias	Consistency	Directness	Precision	Overall Strength of Evidence
2 (157 HD CKD patients)	1 medium risk 1 high risk	Consistent	indirect	Imprecise	Low

CKD=chronic kidney disease; HD=hemodialysis

### Detailed Synthesis (Tables 2.2 and Table 2.9)

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.<sup>44,61</sup> 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.<sup>44,61</sup> One study was rated as being at a high risk of bias,<sup>61</sup> and one at a medium risk of bias,<sup>44</sup> 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  $\geq 1$  g/dL after intravenous iron treatment,<sup>44</sup> or as an increase in Hb  $>15$  percent from baseline.<sup>61</sup> 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.<sup>61</sup>

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 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 2.9. Summary results of the ability of sTfR to predict the response to IV iron treatment in HD CKD patients**

Study, Year [UI]	Lab Analysis or Assay	sTfR Cut-off (pg)	N <sub>analyzed</sub> (% Responders)	Sens, %	Spec, %	AUC (CI)	Ferritin AUC (CI)	Comparison sTfR vs. Ferritin P value	TSAT AUC (CI)	Comparison sTfR vs. TSAT P value	Other Results
Bovy, 2007 <sup>44</sup> [17237481] Belgium	Enzyme-linked immunosorbent assays (Quantikine™ IVDTM, R&D Systems, Minneapolis, MN, USA)	>6.6 (Best cutoff)	32 (NR)	NR	NR	0.989 (0.922, 1.0)	0.834 (0.685, 0.983)	NS	0.896 (0.778, 1.014)	NS	NR
Tessitore, 2001 <sup>61</sup> [11427634]	Commercially available automated particle-enhanced immunophelometric (PETIA) assay (Dade Behring, Marburg, Germany), using highly purified sTfR isolated from human serum as a calibrator.	>1.5	125 (41)	81	71	0.7834 (0.668, 0.899)	0.633 (0.514, 0.752)	NS	0.753 (0.669, 0.837)	NS	NR
		sTfR >1.5 or TSAT <19		91	66	NR	NR	NR	NR	NR	NR

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

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

## Erythrocyte Zinc Protoporphyrin

### Key Points (Table 2.10)

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.<sup>61,65</sup> 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 to TSAT and ferritin, and none of the three markers was a significant predictor of response to IV iron treatment.<sup>61</sup>

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 2.10. Overall strength of evidence for the test performance of Erythrocyte Zinc Protoporphyrin (ZPP) comparing with that of classical markers of iron status to predict a response to IV iron treatment**

Number of Studies (Total N Analyzed)	Risk of Bias	Consistency	Directness	Precision	Overall Strength of Evidence
2 (187 HD CKD patients)	2 high risk	Not applicable (different reference standards)	Indirect	Imprecise	insufficient

CKD=chronic kidney disease; HD=hemodialysis

### Detailed Synthesis (Tables 2.2 and 2.11)

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.<sup>61,65</sup> 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.<sup>65</sup> 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.<sup>61</sup> 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,<sup>61</sup> or a 5 percent increase in Hct or a decrease in erythropoietin dose of more than 2000 units.<sup>65</sup> 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 to TSAT,<sup>65</sup> and the other study reported a higher sensitivity and lower specificity for ZPP as compared to TSAT.<sup>61</sup> When the three markers were assessed in a multivariate regression analysis in the latter study, the test accuracy of ZPP was comparable to TSAT and ferritin, and none of the three markers was a

significant predictor of response to IV iron treatment.<sup>61</sup> 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 to 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.<sup>65</sup>

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 2.11. Summary results of the test performance of erythrocyte zinc protoporphyrin (ZPP) in predicting a response to IV iron treatment in adult HD CKD patients**

Study, Year [UI]	Lab Analysis or Assay	N <sub>analyzed</sub> (% Responders)	ZPP Cut-off (pg)	Sens, %	Spec, %	AUC (CI)	Ferritin AUC (CI)	ZPP vs. Ferritin P value	TSAT AUC (CI)	ZPP vs. TSAT P value	Other Results
Fishbane, 1995 <sup>65</sup> [7872320]	Hematofluorometer, AVIV Biomedicals	62 (62)	>52	100	17	0.853 (0.760, 0.946) <sup>b</sup>	0.785 (0.672, 0.897) <sup>b</sup>	NR	0.665 (0.53, 0.80) <sup>b</sup>	NR	NR
			>66	90	35						
			>90	87	83						
			>103	80	91						
			>107	70	91						
			>109	60	91						
			>112	50	96						
			>122	40	96						
			>138	30	100						
			>140	20	100						
			>177	10	100						
>190	0	100									
Tessitore, 2001 <sup>61</sup> [11427634]	Fluorometer, Shimadzu, Rf-551	125 (41)	>52	81	69	0.77 (0.63, 0.91)	0.633 (0.51, 0.75)	NS	0.753 (0.67, 0.84)	NS	Not a significant predictor in the multi-variate regression analysis
			>90	14	97						
			ZPP >52 or %HYPO >6%	94	72						

CKD=Chronic kidney disease; AUC=Area under the curve; 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

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

<sup>b</sup>CI estimated from reported sensitivity and specificity pairs at different cutoffs using Watkins, M. W. (2000). *An EXCEL program for calculating and graphing the Receiver Operating Characteristic (ROC)* [Computer software]. State College, PA: Ed & Psych Associates.

<sup>c</sup>In patients whose ferritin level did not increase to more than 100 ng/mL at the end of 3 months, an additional 1,000 mg was administered using the same protocol.

## 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 2.2 and 2.12)

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).<sup>62</sup> 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  $\geq 1$  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 2.12. Summary results of the ability of serum hepcidin to predict the response to IV iron treatment in HD CKD patients**

Study, Year [UI]	Lab Analysis or Assay	N <sub>analyzed</sub> (% Responders)	Index Test	Sens, %	Spec, %	AUC (CI)	Ferritin AUC (CI)	Comparison Hepcidin vs. Ferritin P value	TSAT AUC (CI)	Comparison Hepcidin vs. TSAT P value	Other Results
Tessitore, 2010 <sup>62</sup> [20538788]	PBSCIIc mass spectrometer and copperloaded immobilized metal-affinity capture ProteinChip arrays (IMAC30-Cu2+)	56 (NR)	Hepcidin-20	NR	NR	0.541 (0.373, 0.710)	0.552 (0.391, 0.713)	NS	0.593 (0.431, 0.754)	NS	NR
			Hepcidin-25	NR	NR	0.517 (0.330, 0.672)	0.552 (0.391, 0.713)	NS	0.593 (0.431, 0.754)	NS	NR

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; TSAT=transferrin saturation; UI=universal identifier/pubmed ID

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

## 2b. Adverse Effects or Harms Associated with Testing

Only seven of the 27 identified studies reported information on harms.<sup>47,50,59,64-67</sup> 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 2.13**.

**Table 2.13. Adverse effects or harms reported in the 27 studies included in Key Question 2**

Author, Year [PMID]	Adverse effects or harms reported
Chuang, 2003 <sup>47</sup> [12543894]	No adverse reactions were found to be associated with iron therapy
Domrongkitchaiporn, 1999 <sup>50</sup> [10401012]	No adverse reaction developed during or immediately after intravenous iron infusion
Fishbane, 1995 <sup>65</sup> [7872320]	Development of bleeding episodes: 5/62 (8%); significant intercurrent illnesses: 5/62 (8%); death: 2/62 (3%)
Kaneko, 2003 <sup>66</sup> [12631092] <sup>a</sup>	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.
Mittman, 1997 <sup>59</sup> [9398141]	No adverse reactions were found to be associated with iron treatment.
Silva, 1998 <sup>67</sup> [9794562]	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
Van Wyck, 2005 <sup>64</sup> [16316362]	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.

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

<sup>a</sup> 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 3.1)

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 to 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 to 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 3.1. 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**

Number of Studies (Total N Analyzed)	Risk of Bias	Consistency	Directness	Precision	Overall Strength of Evidence
2 RCTs (354 adult CKD HD patients)	2 medium risk	Inconsistent (dose of epoetin treatment) Consistent (dose of iron treatment – post-hoc intermediate outcome)	Indirect	Imprecise	Low

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

## Detailed Synthesis

### Description of Included Studies (Table 3.2)

Two RCTs, with a total of 354 adult CKD patients (mean age: 60 years), undergoing hemodialysis were included.<sup>66,70</sup> One trial was conducted in the U.S., with a followup duration of 6 months,<sup>70</sup> and the other in Japan, with a followup duration of 4 months.<sup>66</sup> 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 3.3)

Both RCTs reported the dose of epoetin required to maintain the Hct target as the primary outcome.<sup>66,70</sup> The epoetin dose adjustment schedule was more frequent in the U.S. trial (every 2 weeks)<sup>70</sup> than the Japanese trial (twice per month, 3 days after the previous hemodialysis therapy).<sup>66</sup> 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 (Table 2.34). 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.<sup>70</sup> 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.<sup>66</sup> 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<sup>70</sup> and the secondary outcome in the Japanese trial.<sup>66</sup> 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).<sup>66</sup>

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.<sup>70</sup> 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 ( $\pm$ 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 ( $\pm$ 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.<sup>66</sup> 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 to 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;<sup>70</sup> and 6.3, 130.5, and 48.9 percent, respectively in the Japanese trial.<sup>66</sup> 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 3.4**).

## 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).<sup>70</sup> 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.<sup>66</sup> 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 3.2. 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**

Study, Year [UI] Country	N <sub>enrolled</sub> / N <sub>analyzed</sub>	Demographics	Duration of HD	Anemia and Iron Status Indices	Intervention	Comparator	Iron Treatment Regimen	Anemia Management Protocol Targets	Follow up Months	Risk of Bias
Fishbane, 2001 <sup>70</sup> [11737617] US	157/138	Male (%):54 Age (yr): 60 Race (%): -Caucasian 46 -African American 44 -Hispanic 7 Other 3	≥3 months	Hb (g/dL):NR Hct (%):35.6 ferritin (ng/mL):240.6 5 TSAT (%):23.5	Iron management based on serum CHr measured every 4 wks	Iron management based on ferritin or TSAT measured every 4 wks	IV iron dextran 100 mg for 10 consecutive treatments if presence of iron Tx trigger	Hb and Hct every 2 wks; the dose of EPO adjusted to maintain Hct 33-36% <sup>a</sup>	6	Medium
Kaneko, 2003 <sup>66</sup> [12631092] Japan	197/183	Male (%):61 Age (yr): 59 Race (%): NR	2 months to 26 years	Hb (g/dL):NR Hct (%):31.3 ferritin (ng/mL):247.4 TSAT (%):25.8	Iron management based on serum CHr measured twice a month	Iron management based TSAT measured twice a month	IV iron colloid with chondroitin sulfate 40 mg 3 times per wk for 2 wks if presence of iron Tx trigger	Hct twice per mo; the dose of EPO adjusted to maintain Hct 29.5-32.5% <sup>b</sup>	4	Medium

CHr=reticulocyte hemoglobin content; Hb=hemoglobin; Hct=hematocrit; 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 3.3. Dose of epoetin required to maintain hematocrit targets**

Study, Year [UI] Country	Arms (Trigger for Iron Tx)	N	Unit	Baseline	4 wks	8 wks	9 wks	11 wks	16 wks	24 wks	28 wks	P <sub>within</sub>	P <sub>between</sub>
Fishbane, 2001 <sup>70</sup> [11737617]	Iron management guided by serum CHR measured every 4 wks (CHR <29 pg)	64	Mean (SD), IU/week	12237 (12001)	NR	12200 (12049)	NR	NR	11300 (11785)	10933 (12095)	10949 (12154)	NS	NR
US	Iron management guided by ferritin or TSAT measured every 4 wks (ferritin <100 ng/mL or TSAT < 20%)	74	Mean (SD), IU/week	12232 (11029)	NR	12077 (11444)	NR	NR	12100 (11029)	11902 (11320)	11772 (11780)	NS	
Kaneko, 2003 <sup>66</sup> [12631092]	Iron management guided by serum CHR (CHR <32.5 pg) twice a month	94	Mean (SD), IU/week	4121 (2922)	5426 (3481)	NR	3957 (3320)	3638 (3276)	3606 (3347)	NA	NA	<0.05 at 4 wks	<0.05 from 11 wks
Japan	Iron management guided by TSAT (TSAT <20%) twice a month	89	Mean (SD), IU/week	4081 (3123)	4803 (3325)	NR	3051 (2730)	2528 (2730)	2629 (2640)	NA	NA	<0.01 from 11 wks	

NA=not applicable; NR=not reported; NS=not statistically significant; TSAT=Transferrin saturation; Tx=treatment; UI=universal identifier; wk=week

**Table 3.4. Changes in iron status markers**

Study, Year [UI] Country	Arms (Trigger for Iron Tx)	N	Followup Duration (wk)	Outcome	Unit	Baseline	Final	P <sub>within</sub>	P <sub>between</sub>
Fishbane, 2001 <sup>70</sup> [11737617]	Iron management guided by serum CHR measured every 4 wks (CHR <29 pg)	64	28	ferritin	Mean (SD), ng/mL	251.7 (231.3)	304.7 (290.6)	NS	0.05 at final
US	Iron management guided by ferritin or TSAT measured every 4 wks (ferritin <100 ng/mL or TSAT < 20%)	74		ferritin	Mean (SD), ng/mL	229.6 (178.8)	399.5 (247.6)	<0.05	
	Iron management guided by serum CHR measured every 4 wks (CHR <29 pg)	64	28	TSAT	Mean (SD), %	22.3 (11.7)	25.8 (16.6)	NS	0.04 at final
	Iron management guided by ferritin or TSAT measured every 4 wks (ferritin <100 ng/mL or TSAT < 20%)	74		TSAT	Mean (SD), %	24.7 (12.7)	29.4 (17.8)	<0.05	
	Iron management guided by serum CHR measured every 4 wks (CHR <29 pg)	64	28	CHR	Mean (SD), pg	30.8 (1.7)	30.8 (1.8)	NS	NS
	Iron management guided by ferritin or TSAT measured every 4 wks (ferritin <100 ng/mL or TSAT < 20%)	74		CHR	Mean (SD), pg	31.1 (1.8)	31.1 (1.8)	NS	
Kaneko, 2003 <sup>66</sup> [12631092]	Iron management guided by serum CHR (CHR <32.5 pg) twice a month	94	16	ferritin	Mean (SD),	234.5 (307.0)	279.5 (326.9)	<0.01	NS

Study, Year [UI] Country	Arms (Trigger for Iron Tx)	N	Followup Duration (wk)	Outcome	Unit	Baseline	Final	P <sub>within</sub>	P <sub>between</sub>
Japan	Iron management guided by TSAT (TSAT <20%) twice a month	89		ferritin	ng/mL Mean (SD), ng/mL	257.0 (453.4)	372.6 (518.1)	<0.0001	
	Iron management guided by serum CHr (CHr <32.5 pg) twice a month	94	16	TSAT	Mean (SD), %	25.5 (12.6)	28.2 (14.3)	NS	<0.05 at final
	Iron management guided by TSAT (TSAT <20%) twice a month	89		TSAT	Mean (SD), %	25.7 (15.6)	32.7 (14.9)	<0.0001	
	Iron management guided by serum CHr (CHr <32.5 pg) twice a month	94	16	CHr	Mean (SD), pg	33.2 (2.2)	34.4 (1.6)	<0.0001	NS
	Iron management guided by TSAT (TSAT <20%) twice a month	89		CHr	Mean (SD), pg	32.8 (2.4)	34.3 (1.9)	<0.0001	

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 4.1 and 4.2)**

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.<sup>68</sup> 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 ( $\leq 25\%$ ), 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  $\geq 2$  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  $\geq 19$  percent and a ferritin cutoff of  $\geq 726$  ng/mL were 29 and 70 percent, and 27 and 69 percent, respectively. The sensitivity and specificity pairs for a CHr cutoff of  $\geq 31.2$  pg and a sTfR cutoff of  $\geq 5.9$  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  $\geq 31.2$  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  $\geq 2$  g/dL Hb response in patients with baseline CHr  $\geq 31.2$  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.<sup>45</sup> 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).<sup>69</sup> 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 4.1. Characteristics of studies evaluating factors affecting test performance and clinical utility**

Study, Year [UI] Country, Design	Study Population	Groups	Intervention	N	Demographics	Anemia and Iron Status Indices	Followup Duration (wk)	Risk of Bias
Singh, 2007 <sup>68</sup> [17396118] US RCT	HD CKD	IV iron	IV ferric gluconate 1 g & 25% increase in weekly epoetin dose for 6 weeks	64	Male (%):58 Age (yr): 61 Race (%): -Caucasian 31 -African American 47 -Hispanic 14 -Asian/Pacific islander 8	Hb (g/dL): 10.4 Hct (%): NR ferritin (ng/mL): 759 TSAT (%): 18.5	6	Medium
		No IV Iron	25% increase in weekly epoetin dose for 6 weeks	65	Male (%):43 Age (yr): 59 Race (%): -Caucasian 31 -African American 51 -Hispanic 14 -Asian/Pacific islander 3 Other 2	Hb (g/dL): 10.2 Hct (%): NR ferritin (ng/mL): 765 TSAT (%): 19	6	
Ahluwalia, 1997(study 1; study 4) <sup>69</sup> [9328369] US Prospective cohorts	HD CKD	ESA naïve patient starting ESA treatment	One ferrous sulfate tablet (containing 50 mg elemental iron) per day with mean ESA dose of 162 IU/kg/week	17	Male (%):NR Age (yr): 46 Race (%):NR	Hb (g/dL):7.1 Hct (%):NR ferritin (ng/mL):98.5 TSAT (%):NR	12	Medium
		Increase in ESA dose for patient on maintenance ESA treatment	One 65 mg ferrous sulfate tablet per day with mean ESA dose of 121 IU/kg/week	16	Male (%):NR Age (yr): 46 Race (%):NR	Hb (g/dL):7.8 Hct (%):NR ferritin (ng/mL):59.5 TSAT (%):17	4	

ESA=erythropoiesis stimulating agents; Hb=hemoglobin; Hct=Hematocrit; IV=intravenous; IU=international units; NR=not reported; TSAT=transferrin saturation; wk=week; yr=year

**Table 4.2. Test accuracy of TSAT, ferritin, CHr and sTfR for predicting change in hemoglobin in subgroups of IV iron and no IV iron treatment**

Study, Year [UI] Country Design	Group	N	Reference Standard - Dx of Iron Deficiency	Index Test	Cutoff for Index Test	Sensitivity, %	Specificity, %	
Singh, 2007 <sup>68</sup> [17396118] US RCT	IV Iron	64	Hb change of $\geq 2$ g/dL	TSAT	$\geq 19\%$	48.5	54.8	
				ferritin	$\geq 726$ ng/mL	46.9	53.1	
				CHr	$\geq 31.2$ pg/cell	63.9	75	
				sTfR	$\geq 5.9$ mg/L	42.4	48.4	
	No IV Iron	65	Hb change of $\geq 2$ g/dL	TSAT	$\geq 19\%$	28.6	70	
				ferritin	$\geq 726$ ng/mL	27.3	68.8	
				CHr	$\geq 31.2$ pg/cell	26.7	68.6	
				sTfR	$\geq 5.9$ mg/L	35.3	77.4	
	Ahlwalia, 1997 (study 1; study 4) <sup>69</sup> [9328369] US (study 1;study4) Prospective cohorts	ESA naïve patient starting ESA treatment	16	> 2 g/dL increase in Hb at 3 months after initiation of rHuEPO therapy	sTfR	<6 mg/L	88	78
					ferritin	>50 $\mu$ g/L	88	44
				ferritin & sTfR	>50 $\mu$ g/L & <6 mg/L	75	78	
Increase in ESA dose for patient on ESA treatment		17	> 1 g/dL increase in Hb 4 weeks over the baseline level	sTfR	> 20 % increase in sTfR at 1 week	69	100	

CHr= reticulocytes hemoglobin content; Dx=Diagnosis; ESA=Erythropoiesis stimulating agents; IV=Intravenous; NR=not reported; 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 B**.

**Table B. Summary of the strength of evidence addressing Key Questions**

Key Questions	Strength of Evidence	Summary, Comments, and Conclusions
<b>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?</b>	Low / Insufficient (depending on the test comparisons, study populations, or test performance outcomes)	<ul style="list-style-type: none"> <li>• Among adult HD CKD patients, there is a low level of evidence that:               <ul style="list-style-type: none"> <li>○ 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 &lt;27 or &lt;28 pg) has better sensitivity and specificity to predict iron deficiency than classical markers (TSAT &lt;20 or ferritin &lt;100 ng/mL).</li> <li>○ 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 &gt;6% or &gt;10%) has a better sensitivity and specificity to predict iron deficiency (as defined by a response to IV iron treatment) than classical markers (TSAT &lt;20% or ferritin &lt;100 ng/mL).</li> <li>○ Soluble transferrin receptor (sTfR) has a similar test performance compared with classical markers (TSAT or ferritin) to predict a response to IV iron treatment.</li> </ul> </li> <li>• There is insufficient evidence regarding:               <ul style="list-style-type: none"> <li>○ Test performance of newer markers of iron status as an add-on to older markers.</li> <li>○ Test performance comparing erythrocyte zinc protoporphyrin (ZPP) and hepcidin to predict a response to IV iron treatment in adult HD CKD patients.</li> <li>○ 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.</li> </ul> </li> </ul>
<b>2a. What reference standards are used for the diagnosis of iron status in studies evaluating test accuracy?</b>	Not rated (descriptive data)	<ul style="list-style-type: none"> <li>• There is a lack of generally accepted reference standard tests for determining iron deficiency in the setting of CKD.<sup>16</sup> 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 “function 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.</li> </ul>

Key Questions	Strength of Evidence	Summary, Comments, and Conclusions
<b>2b. What are the adverse effects or harms associated with testing using newer and/or older markers of iron status?</b>	Insufficient	<ul style="list-style-type: none"> <li>Only 7 of the 27 studies reported information: <ul style="list-style-type: none"> <li>3 studies reported no adverse events associated with iron therapy during the study periods</li> <li>A total of 5 deaths reported. Studies did not attribute these deaths to either testing or any treatment.</li> <li>Most of the reported harms were attributed to iron therapy.</li> </ul> </li> </ul>
<b>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?</b>	Low	<ul style="list-style-type: none"> <li>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 to those guided by TSAT or ferritin.</li> <li>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.</li> <li>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.</li> <li>No study compared iron management guided by classical markers with that of newer markers (%HYPO, sTfR, Ret-He, ZPP, or hepcidin).</li> </ul>
<b>3a. What are the adverse effects or harms associated with the treatments guided by tests of iron status?</b>	Insufficient	<ul style="list-style-type: none"> <li>Only 1 RCT explicitly monitored the adverse events: <ul style="list-style-type: none"> <li>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).</li> <li>One patient in the TSAT group dropped out because of massive bleeding due to a femoral bone fracture and need for blood transfusion.</li> <li>There were no significant differences in the hospitalization or infection rates of the two iron management groups.</li> </ul> </li> </ul>
<b>Key Question 4. What factors affect the test performance and clinical utility of newer markers of iron status?</b>	Insufficient	<ul style="list-style-type: none"> <li>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: <ul style="list-style-type: none"> <li>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.</li> <li>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.</li> <li>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.</li> </ul> </li> <li>No study performed analyses by patient subgroups.</li> <li>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.</li> </ul>

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

We synthesized 27 studies to answer Key Question 2 (test performance of newer markers compared to 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 limits our confidence in evaluating the consistency 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  $\geq 1$  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 to 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.<sup>5,16</sup> 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.<sup>5,16</sup> 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 placement 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.<sup>61</sup> 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).<sup>66,70</sup> 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)<sup>70</sup> as compared to the Japanese trial (twice per month, 3 days after the previous HD therapy).<sup>66</sup> The U.S. trial also used

much higher doses of epoetin than the Japanese trial at baseline (12,232 vs. 4121 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.

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  $LR+ \geq 10$  and  $LR- \leq 10$ ) 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,<sup>71</sup> 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 C**.

**Table C. Research gaps and suggestions for future research**

Key Questions	Research Gaps	Suggestions for Future Research
<b>Key Question 2.</b> <b>What is the diagnostic test accuracy of newer markers of iron status as a replacement for or an add-on to classical laboratory markers?</b>	Insufficient evidence for the test performance of newer markers of iron status as an add-on to older markers	<ul style="list-style-type: none"> <li>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.</li> </ul>
	Many existing studies are at a high risk of bias, limiting their utility in informing clinical practice	<ul style="list-style-type: none"> <li>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.<sup>72</sup></li> <li>Studies assessing diagnostic accuracy should instead aim to enroll patients representative of the spectrum of disease typically seen in clinical practice.</li> <li>Future studies should provide details about the study base and sampling methods.</li> </ul>
<b>Key Question 3.</b> <b>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?</b>	There is no uniform iron management algorithms across studies	<ul style="list-style-type: none"> <li>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</li> <li>Assessing impact of the most promising iron management algorithms on both intermediate and patient outcomes through prospective observational studies or RCTs.</li> </ul>

Key Questions	Research Gaps	Suggestions for Future Research
<b>Key Question 4. What factors affect the test performance and clinical utility of newer markers of iron status?</b>	Insufficient evidence to draw conclusions regarding factors that may affect the test performance or clinical utility of laboratory markers of iron status	<ul style="list-style-type: none"> <li>• Future studies are need to evaluated the following factors, suggested by the experts: <ul style="list-style-type: none"> <li>○ Biological variation in diagnostic indices</li> <li>○ Use of different diagnostic reference standards</li> <li>○ Type of dialysis (i.e., peritoneal or hemodialysis)</li> <li>○ Patient subgroups (i.e., age, sex, comorbid conditions, erythropoiesis-stimulating agent resistance, protein energy malnutrition secondary to an inflammatory state, hemoglobinopathies [e.g., thalessemia and sickle cell anemia])</li> <li>○ Route of iron administration (i.e., oral or intravenous)</li> <li>○ Treatment regimen (i.e., repletion or continuous treatment)</li> <li>○ Interactions between treatments (i.e., patients treated with versus without ESA, patients treated with versus without iron-replacement therapy)</li> </ul> </li> </ul>
	Whether test performance and clinical utility of newer or classical markers of iron status vary by different CKD populations are not known	<ul style="list-style-type: none"> <li>• 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.</li> <li>• Power calculations should be performed to take into account for the planed subgroup analyses.</li> </ul>
<b>Cross-cutting issues (for Key Question 2, 3, and 4)</b>	There is no reference standard for determining iron deficiency in CKD patients	<ul style="list-style-type: none"> <li>• 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.</li> </ul>
	Existing studies were underpowered leading to imprecise estimates	<ul style="list-style-type: none"> <li>• 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.<sup>73</sup></li> </ul>
	There is no decision analysis to assess how using combinations of different markers to guide iron management strategies might influence clinical outcomes	<ul style="list-style-type: none"> <li>• Patient outcomes of interest are <ul style="list-style-type: none"> <li>○ Mortality</li> <li>○ Morbidity (e.g., cardiac or liver toxicity and infection)</li> <li>○ 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).</li> <li>○ 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)</li> </ul> </li> <li>• For studies assessing clinical outcomes, blinding to test results to the outcome assessors is essential to avoid bias.<sup>35,72</sup></li> </ul>

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

## Conclusions

Combining the evidence addressing Key Questions 2, 3, and 4, we can conclude that there is insufficient evidence to support the use of newer laboratory markers of iron status to replace TSAT or ferritin for assessing iron status, and that no single test (using either newer or classical markers) is adequate to determine iron status. 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 to those guided by TSAT or ferritin. These results suggest that CHr may be a suitable alternative marker of iron status for guiding iron treatment, and could potentially reduce the frequency of iron testing and potential harms from IV iron treatment.

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

1. National Kidney Foundation. What causes CKD? Available at <http://www.kidney.org/kidneydisease/ckd/index.cfm#what> Accessed December 2011.
2. K/DOQI clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification. *Am J Kidney Dis* 2002 Feb;39(2 Suppl 1):S1-266.
3. National Kidney Foundation. The Facts About Chronic Kidney Disease (CKD). Available at <http://www.kidney.org/kidneydisease/ckd/index.cfm#facts> Accessed December 2011.
4. Couser WG, Remuzzi G, Mendis S, et al. The contribution of chronic kidney disease to the global burden of major noncommunicable diseases. *Kidney Int* 2011 Dec;80(12):1258-70.
5. KDOQI Clinical Practice Guidelines and Clinical Practice Recommendations for Anemia in Chronic Kidney Disease. 2006.
6. Mix TC, Kazmi W, Khan S, et al. Anemia: a continuing problem following kidney transplantation. *Am J Transplant* 2003 Nov;3(11):1426-33.
7. Vanrenterghem Y, Ponticelli C, Morales JM, et al. Prevalence and management of anemia in renal transplant recipients: a European survey. *Am J Transplant* 2003 Jul;3(7):835-45.
8. United States Renal Data System. USRDS 2006 annual data report: atlas of end-stage renal disease in the United States. Bethesda, MD: National Institute of Diabetes and Digestive and Kidney Diseases; 2006.
9. Robertson HT, Haley NR, Guthrie M, et al. Recombinant erythropoietin improves exercise capacity in anemic hemodialysis patients. *Am J Kidney Dis* 1990 Apr;15(4):325-32.
10. Pickett JL, Theberge DC, Brown WS, et al. Normalizing hematocrit in dialysis patients improves brain function. *Am J Kidney Dis* 1999 Jun;33(6):1122-30.
11. Foley RN, Parfrey PS, Harnett JD, et al. The impact of anemia on cardiomyopathy, morbidity, and mortality in end-stage renal disease. *Am J Kidney Dis* 1996 Jul;28(1):53-61.
12. Harris DC, Chapman JR, Stewart JH, et al. Low dose erythropoietin in maintenance haemodialysis: improvement in quality of life and reduction in true cost of haemodialysis. *Aust N Z J Med* 1991 Oct;21(5):693-700.
13. Evans RW, Rader B, Manninen DL. The quality of life of hemodialysis recipients treated with recombinant human erythropoietin. Cooperative Multicenter EPO Clinical Trial Group. *JAMA* 1990 Feb 9;263(6):825-30.
14. Stevens ME, Summerfield GP, Hall AA, et al. Cost benefits of low dose subcutaneous erythropoietin in patients with anaemia of end stage renal disease. *BMJ* 1992 Feb 22;304(6825):474-77.
15. Collins AJ, Li S, Ebben J, et al. Hematocrit levels and associated Medicare expenditures. *Am J Kidney Dis* 2000 Aug;36(2):282-93.
16. Anaemia Management in Chronic Kidney Disease - Rapid Update 2011. National Clinical Guideline Centre; 2011.
17. Eschbach JW, Abdulhadi MH, Browne JK, et al. Recombinant human erythropoietin in anemic patients with end-stage renal disease. Results of a phase III multicenter clinical trial. *Ann Intern Med* 1989 Dec 15;111(12):992-1000.

18. Eckardt KU. Chronic kidney disease: Are elevated doses of ESAs associated with adverse outcomes? *Nature Reviews Nephrology* 2010 Oct;6(10):566-68.
19. Hayat A. Safety issues with intravenous iron products in the management of anemia in chronic kidney disease. *Clin Med Res* 2008 Dec;6(3-4):93-102.
20. IV. NKF-K/DOQI Clinical Practice Guidelines for Anemia of Chronic Kidney Disease: update 2000. *Am J Kidney Dis* 2001 Jan;37(1 Suppl 1):S182-S238.
21. Bain BJ. Bone marrow biopsy morbidity: review of 2003. *J Clin Pathol* 2005 Apr;58(4):406-08.
22. Winkel P, Statland BE, Bokelund H. Factors contributing to intra-individual variation of serum constituents: 5. Short-term day-to-day and within-hour variation of serum constituents in healthy subjects. *Clin Chem* 1974 Dec;20(12):1520-27.
23. Winkel P, Statland BE, Bokelund H. The effects of time of venipuncture on variation of serum constituents. Consideration of within-day and day-to-day changes in a group of healthy young men. *Am J Clin Pathol* 1975 Oct;64(4):433-47.
24. Wish JB. Assessing iron status: beyond serum ferritin and transferrin saturation. *Clin J Am Soc Nephrol* 2006 Sep;1 Suppl 1:S4-S8.
25. Tietz NW, Rinker AD, Morrison SR. When is a serum iron really a serum iron? The status of serum iron measurements. *Clin Chem* 1994 Apr;40(4):546-51.
26. Eckfeldt JH, Witte DL. Serum iron: would analytical improvement enhance patient outcome? *Clin Chem* 1994 Apr;40(4):505-07.
27. Braun J, Lindner K, Schreiber M, et al. Percentage of hypochromic red blood cells as predictor of erythropoietic and iron response after i.v. iron supplementation in maintenance haemodialysis patients. *Nephrol Dial Transplant* 1997 Jun;12(6):1173-81.
28. Macdougall IC, Cavill I, Hulme B, et al. Detection of functional iron deficiency during erythropoietin treatment: a new approach. *BMJ* 1992 Jan 25;304(6821):225-26.
29. Braun J. Erythrocyte zinc protoporphyrin. *Kidney Int Suppl* 1999 Mar;69:S57-S60.
30. Babitt JL, Lin HY. Molecular mechanisms of hepcidin regulation: implications for the anemia of CKD. *Am J Kidney Dis* 2010 Apr;55(4):726-41.
31. Canavese C, Bergamo D, Ciccone G, et al. Validation of serum ferritin values by magnetic susceptometry in predicting iron overload in dialysis patients. *Kidney Int* 2004 Mar;65(3):1091-98.
32. Centers for Medicare & Medicaid Services C. CMS PROPOSES NEW PROSPECTIVE PAYMENT SYSTEM FOR RENAL DIALYSIS FACILITIES. 2009.
33. Agency for Healthcare Research and Quality. *Methods Reference Guide for Effectiveness and Comparative Effectiveness Reviews, Version 1.0.* 2010.
34. Moher D, Liberati A, Tetzlaff J, et al. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med* 2009 Jul 21;6(7):e1000097. PMID: doi:10.1371/journal.pmed.1000097.
35. Whiting P, Rutjes A, Reitsma J, et al. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Medical Research Methodology* 2003;3(1):25. PMID: 10.1186/1471-2288-3-25.

36. Higgins JPT GSe. Cochrane handbook for systematic reviews of interventions. Version 5.0.2. [updated September 2009]. The Cochrane Collaboration.; 2009.
37. Altman DG, Bland JM. Presentation of numerical data. *BMJ* 1996 Mar 2;312(7030):572.
38. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem* 1993 Apr;39(4):561-77.
39. Florkowski CM. Sensitivity, specificity, receiver-operating characteristic (ROC) curves and likelihood ratios: communicating the performance of diagnostic tests. *Clin Biochem Rev* 2008 Aug;29 Suppl 1:S83-S87.
40. Glas AS, Lijmer JG, Prins MH, et al. The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol* 2003 Nov;56(11):1129-35.
41. Baldus M, Walter H, Thies K, et al. Transferrin receptor assay and zinc protoporphyrin as markers of iron-deficient erythropoiesis in end-stage renal disease patients. *Clinical Nephrology* 1998 Mar;49(3):186-92.
42. Bhandari S, Norfolk D, Brownjohn A, et al. Evaluation of RBC ferritin and reticulocyte measurements in monitoring response to intravenous iron therapy. *American Journal of Kidney Diseases* 1997 Dec;30(6):814-21.
43. Bhandari S, Turney JH, Brownjohn AM, et al. Reticulocyte indices in patients with end stage renal disease on hemodialysis. *Journal of Nephrology* 1998 Mar;11(2):78-82.
44. Bovy C, Gothot A, Delanaye P, et al. Mature erythrocyte parameters as new markers of functional iron deficiency in haemodialysis: sensitivity and specificity. *Nephrology Dialysis Transplantation* 2007 Apr;22(4):1156-62.
45. Bruynars C, Schiller B, Moran J. Reticulocyte hemoglobin equivalent (Ret He) and assessment of iron-deficient states. *Clinical & Laboratory Haematology* 2006 Oct;28(5):303-08.
46. Buttarello M, Pajola R, Novello E, et al. Diagnosis of iron deficiency in patients undergoing hemodialysis. *American Journal of Clinical Pathology* 2010 Jun;133(6):949-54.
47. Chuang CL, Liu RS, Wei YH, et al. Early prediction of response to intravenous iron supplementation by reticulocyte haemoglobin content and high-fluorescence reticulocyte count in haemodialysis patients. *Nephrology Dialysis Transplantation* 2003 Feb;18(2):370-77.
48. Cullen P, Soffker J, Hopfl M, et al. Hypochromic red cells and reticulocyte haemoglobin content as markers of iron-deficient erythropoiesis in patients undergoing chronic haemodialysis. *Nephrology Dialysis Transplantation* 1999 Mar;14(3):659-65.
49. Daschner M, Mehls O, Schaefer F. Soluble transferrin receptor is correlated with erythropoietin sensitivity in dialysis patients. *Clinical Nephrology* 1999 Oct;52(4):246-52.
50. Domrongkitchaiporn S, Jirakranont B, Atamasrikul K, et al. Indices of iron status in continuous ambulatory peritoneal dialysis patients. *American Journal of Kidney Diseases* 1999 Jul;34(1):29-35.
51. Eguchi A, Tsuchiya K, Tsukada M, et al. [Clinical usefulness of reticulocyte hemoglobin equivalent (RET-He) in patients at the pre-dialysis stage and in patients on peritoneal dialysis]. [Japanese]. *Nippon Jinzo Gakkai Shi* 2010;Japanese(2):132-40.
52. Fishbane S, Galgano C, Langley RC, Jr., et al. Reticulocyte hemoglobin content in the evaluation of iron status of hemodialysis patients. *Kidney International* 1997 Jul;52(1):217-22.

53. Fusaro M, Munaretto G, Spinello M, et al. Soluble transferrin receptors and reticulocyte hemoglobin concentration in the assessment of iron deficiency in hemodialysis patients. *Journal of Nephrology* 2005 Jan;18(1):72-79.
54. Gupta M, Kannan M, Gupta S, et al. Contribution of iron deficiency to anemia in chronic renal failure. *Indian Journal of Pathology & Microbiology* 2003 Oct;46(4):563-64.
55. Hukic F, Nuhbegovic S, Brkic S, et al. Biochemical markers of iron status in hemodialysis patients. *Medicinski Arhiv* 2010;64(4):219-22.
56. Kim JM, Ihm CH, Kim HJ. Evaluation of reticulocyte haemoglobin content as marker of iron deficiency and predictor of response to intravenous iron in haemodialysis patients. *International Journal of Laboratory Hematology* 2008 Feb;30(1):46-52.
57. Matuszkiewicz-Rowinska J, Ostrowski G, Niemczyk S, et al. [Red cell zinc protoporphyrin and its ratio to serum ferritin (ZPP/logSF index) in the detection of iron deficiency in patients with end-stage renal failure on hemodialysis]. [Polish]. *Polskie Archiwum Medycyny Wewnetrznej* 2003 Jul;110(1):703-10.
58. Mitsuiki K, Harada A, Miyata Y. Assessment of iron deficiency in chronic hemodialysis patients: investigation of cutoff values for reticulocyte hemoglobin content. *Clinical & Experimental Nephrology* 2003 Mar;7(1):52-57.
59. Mittman N, Sreedhara R, Mushnick R, et al. Reticulocyte hemoglobin content predicts functional iron deficiency in hemodialysis patients receiving rHuEPO. *American Journal of Kidney Diseases* 1997 Dec;30(6):912-22.
60. Miwa N, Akiba T, Kimata N, et al. Usefulness of measuring reticulocyte hemoglobin equivalent in the management of haemodialysis patients with iron deficiency. *International Journal of Laboratory Hematology* 2010 Apr;32(2):248-55.
61. Tessitore N, Solero GP, Lippi G, et al. The role of iron status markers in predicting response to intravenous iron in haemodialysis patients on maintenance erythropoietin. *Nephrology Dialysis Transplantation* 2001 Jul;16(7):1416-23.
62. Tessitore N, Girelli D, Campostrini N, et al. Hepcidin is not useful as a biomarker for iron needs in haemodialysis patients on maintenance erythropoiesis-stimulating agents. *Nephrology Dialysis Transplantation* 2010 Dec;25(12):3996-4002.
63. Tsuchiya K, Okano H, Teramura M, et al. Content of reticulocyte hemoglobin is a reliable tool for determining iron deficiency in dialysis patients. *Clinical Nephrology* 2003 Feb;59(2):115-23.
64. Van Wyck DB, Roppolo M, Martinez CO, et al. A randomized, controlled trial comparing IV iron sucrose to oral iron in anemic patients with nondialysis-dependent CKD. *Kidney International* 2005 Dec;68(6):2846-56.
65. Fishbane S, Lynn RI. The utility of zinc protoporphyrin for predicting the need for intravenous iron therapy in hemodialysis patients. *American Journal of Kidney Diseases* 1995 Mar;25(3):426-32.
66. Kaneko Y, Miyazaki S, Hirasawa Y, et al. Transferrin saturation versus reticulocyte hemoglobin content for iron deficiency in Japanese hemodialysis patients. *Kidney International* 2003 Mar;63(3):1086-93.
67. Silva J, Andrade S, Ventura H, et al. Iron supplementation in haemodialysis--practical clinical guidelines. *Nephrology Dialysis Transplantation* 1998 Oct;13(10):2572-77.

68. Singh AK, Coyne DW, Shapiro W, et al. Predictors of the response to treatment in anemic hemodialysis patients with high serum ferritin and low transferrin saturation. *Kidney International* 2007 Jun;71(11):1163-71.
69. Ahluwalia N, Skikne BS, Savin V, et al. Markers of masked iron deficiency and effectiveness of EPO therapy in chronic renal failure. *American Journal of Kidney Diseases* 1997 Oct;30(4):532-41.
70. Fishbane S, Shapiro W, Dutka P, et al. A randomized trial of iron deficiency testing strategies in hemodialysis patients. *Kidney International* 2001 Dec;60(6):2406-11.
71. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making* 1991 Apr;11(2):88-94.
72. Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. *Ann Intern Med* 2003 Jan 7;138(1):40-44.
73. Rothman K, Greenland S. *Modern Epidemiology*. Lippincott Williams and Wilkins; 1998.

## Abbreviations and Acronyms

%HYPO	Percentage of Hypochromic Erythrocytes
AAAC	American Association for Clinical Chemistry
AUC	Area Under the Curve
CHr	Hemoglobin Content of Reticulocytes
CKD	Chronic Kidney Disease
AHRQ	Agency for Healthcare Research and Quality
CI	confidence interval
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
KQ	Key Question
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