Effective Health Care Program

Future Research Needs Paper
Number 33

Biomarkers for Assessing and Managing Iron Deficiency Anemia in Late-Stage Chronic Kidney Disease: Future Research Needs

AHRQ
Agency for Healthcare Research and Quality
Advancing Excellence in Health Care • www.ahrq.gov

Future Research Needs Paper

Number 33

Biomarkers for Assessing and Managing Iron Deficiency Anemia in Late-Stage Chronic Kidney Disease: Future Research Needs

Identification of Future Research Needs From Comparative Effectiveness Review No. 83

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

Contract No. 290-2007-10055-I

Prepared by:
Tuft Evidence-based Practice Center
Boston, MA

Investigators:
Mei Chung, Ph.D., M.P.H.
Jeffrey A. Chan, B.S.
Denish Moorthy, M.B.B.S., M.S.
Nira Hadar, M.S.
Sara J. Ratichek, M.A.
Thomas W. Concannon, Ph.D.
Joseph Lau, M.D.

AHRQ Publication No. 13-EHC038-EF
January 2013

This report is based on research conducted by the Tufts Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. 290-2007-10055-I). The findings and conclusions in this document are those of the author(s), 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 researchers and funders of research make well-informed decisions in designing and funding research and thereby improve the quality of health care services. This report is not intended to be a substitute for the application of scientific judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical research and in conjunction with all other pertinent information, i.e., in the context of available resources and circumstances.

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

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

Suggested citation: Chung M, Chan JA, Moorthy D, Hadar N, Ratichek SJ, Concannon TW, Lau J. Biomarkers for Assessing and Managing Iron Deficiency Anemia in Late-Stage Chronic Kidney Disease: Future Research Needs. Future Research Needs Paper No. 33. (Prepared by the Tufts Evidence-based Practice Center under Contract No. 290-2007-10055-I.) AHRQ Publication No. 13-EHC038-EF. Rockville, MD: Agency for Healthcare Research and Quality. January 2013. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-based Practice Centers (EPCs), sponsors the development of evidence reports and technology assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. The reports and assessments provide organizations with comprehensive, science-based information on common, costly medical conditions and new health care technologies and strategies. The EPCs systematically review the relevant scientific literature on topics assigned to them by AHRQ and conduct additional analyses when appropriate prior to developing their reports and assessments.

An important part of evidence reports is to not only synthesize the evidence, but also to identify the gaps in evidence that limited the ability to answer the systematic review questions. AHRQ supports EPCs to work with various stakeholders to identify and prioritize the future research that is needed by decisionmakers. This information is provided for researchers and funders of research in these Future Research Needs papers. These papers are made available for public comment and use and may be revised.

AHRQ expects that the EPC evidence reports and technology assessments will inform individual health plans, providers, and purchasers as well as the health care system as a whole by providing important information to help improve health care quality. The evidence reports undergo public comment prior to their release as a final report.

We welcome comments on this Future Research Needs document. 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 e-mail to epc@ahrq.hhs.gov.

Carolyn M. Clancy, M.D.
Director
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

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

Elisabeth U. Kato, M.D., M.R.P.
Task Order Officer
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Acknowledgments

The authors gratefully acknowledge the stakeholders for their contributions to this project. Broad expertise and perspectives are sought. Divergent and conflicted opinions are common and perceived as healthy scientific discourse that results in a thoughtful, relevant report. Therefore, in the end, study questions, design and/or methodologic approaches do not necessarily represent the views of individual stakeholder panel participants.

The authors would like to thank Ramon C. Iovin, Ph.D., who served as the medical editor for this report, and Lina Kong Win Chang, B.S., who provided administrative support.

Contributors

Ali K. Abu-Alfa M.D., FASN
American University of Beirut
New Haven, CT

Jeffrey Berns, M.D.
Hospital of the University of Pennsylvania
Philadelphia, PA

James D. Chambers, Ph.D., MPharm, M.Sc.
Tufts Medical Center
Boston, MA

Michael F. Flessner, M.D., Ph.D.
National Institutes of Health
Bethesda, MD

Ameet R. Kini, M.D., Ph.D.
Loyola University
Chicago, IL

Elizabeth A. Koller, M.D.
Centers for Medicare & Medicaid Services
Baltimore, MD

Celeste C. Lee
National Kidney Foundation
Ann Arbor, MI

Poonhar Poon, M.S., R.D.
Tufts Medical Center
Boston, MA

Kimberly Smith, M.D., M.S.
Centers for Medicare and Medicaid Services
Baltimore, MD

Melissa Threlkeld, M.H.A., FACHE
National Kidney Foundation
Austin, TX

Contents

Tables

Figures

Appendixes

Executive Summary

Background

Anemia is a common complication of chronic kidney disease (CKD). The management of anemia in CKD patients must strike an appropriate balance between stimulating generation of erythroblasts (erythropoiesis) and maintaining sufficient iron levels for optimum hemoglobin (Hb) production.[1] As such, it is important to assess iron stores and the availability of iron for erythropoiesis, as adequate iron status is integral to both iron and anemia managements in CKD patients.

Classical iron status tests, of which ferritin and transferrin saturation (TSAT) are the most widely used, exhibit large biological variability in the context of underlying inflammation of CKD.[2-4] The accurate assessment of iron status is dependent on the validity and reliability of laboratory test results, and differences in test performance pose a dilemma regarding the most appropriate test to guide treatment decisions. Several novel biomarkers of iron status have been proposed as alternatives to the classical iron status tests. These include hemoglobin content of reticulocytes (CHr), reticulocyte hemoglobin equivalent (RetHe), percentage of hypochromic erythrocytes (%HYPO), erythrocyte zinc protoporphyrin (ZPP), soluble transferrin receptor (sTfR), and hepcidin. In addition, Superconducting Quantum Interference Devices (SQUIDs) are an alternative non-invasive means for detecting and quantifying liver iron content, via the paramagnetic properties of iron (magnetic resonance diminishes in the liver as iron concentration increases).

The Tufts Evidence-based Practice Center (EPC) conducted a Comparative Effectiveness Review (CER) to systematically evaluate studies that examined the impact on patient-centered outcomes of using the newer laboratory biomarkers as a replacement for or as an add-on to classical laboratory biomarkers of iron status for assessing iron status and the management of iron deficiency in adult and pediatric CKD patients (nondialysis and dialysis).[5] The Key Questions for the CER are presented below:

Key Question 1 (Overarching Question)

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

a. What reference standards are used for the diagnosis of iron status in studies evaluating test performance?

[a] Newer laboratory biomarkers: content of Hb in reticulocytes, percentage of hypochromic red blood cells, erythrocyte zinc protoporphyrin, soluble transferrin receptor, hepcidin, and superconducting quantum interference devices.

[b] Older laboratory biomarkers: bone marrow iron stores, serum iron, transferrin saturation, iron-binding capacity, and ferritin.

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?

Combining the evidence addressing Key Questions 2, 3, and 4, the CER concluded that there is currently insufficient data to determine if most newer laboratory biomarkers of iron status are better than classical markers for predicting iron deficiency as defined by a response to an iron challenge test. However, it may be that CHr and %HYPO have better predictive ability for a response to intravenous (IV) iron treatment than classical markers (TSAT <20% or ferritin <100 ng/mL) in HD CKD patients. In addition, results from two randomized controlled trials (RCTs) showed a reduction in the number of iron status tests and resulting IV iron treatments administered to patients whose iron management was guided by CHr, compared with those guided by TSAT or ferritin. These results suggest that CHr may reduce potential harms from IV iron treatment by lowering the frequency of iron testing; however, the evidence for the potential harms associated with testing or test-associated treatment is insufficient.

Nevertheless, the strength of evidence supporting these conclusions is low and there remains considerable clinical uncertainty regarding the use of newer markers in the assessment of iron status and the 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.

Table A summarizes the evidence gaps identified in the CER. One major evidence gap concerns the dearth of pediatric studies. Addressing this gap would require a specially-composed stakeholder group for determining specific Future Research Needs (FRN). For this reason, the current FRN project is focused on adult CKD patients (nondialysis and dialysis).

Table A. Research gaps and suggestions for future research

Key Questions	Research Gaps
Key Question 1 (overarching question). What is the impact on patient centered outcomes of using the newer laboratory biomarkers[a] as a replacement for or an add-on to the older laboratory biomarkers of iron status for assessing iron status and management of iron deficiency?	No studies were found that provided data directly addressing our overarching question regarding the impact of using newer laboratory biomarkers on patient-centered outcomes (e.g., mortality, morbidity, quality of life, and adverse effects). All gaps described in Key Questions 2, 3, and 4 are applicable to this overarching question.
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? 2a. What reference standards are used for the diagnosis of iron status in studies evaluating test accuracy? 2b. What are the adverse effects or harms associated with testing using newer and/or older markers of iron status?	**Populations:** • Insufficient evidence in adult PD CKD and ND CKD patients. • Whether test performance and clinical utility of newer or classical markers of iron status vary by different CKD populations is unknown. **Interventions and comparators:** • Insufficient evidence for the test performance of newer markers of iron status as an add-on to older markers. **Outcomes:** • There is a lack of a generally accepted reference standard test for determining iron deficiency in the setting of CKD.[1] • A small percentage (26%) of the identified studies reported information on harms, and most studies did not attribute harms to either testing or treatment. **Study Design Issues:** • Many existing studies are at a high risk of bias due to lack of demographic details of study populations and blinding of the diagnosis or patient characteristics. These biases limit studies' utility in informing clinical practice. • Existing studies were underpowered, leading to imprecise estimates.
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? 3a. What are the adverse effects or harms associated with the treatments guided by tests of iron status?	**Populations:** • Insufficient evidence in adult PD CKD and ND CKD patients. **Interventions and comparators:** • There are no uniform iron management algorithms or test-and-treat protocols across studies. • There are a lack of studies with longer followup durations; existing short-term RCTs had a followup duration less than 6 months. • No study compared iron management guided by classical markers with that of newer markers (%HYPO, sTfR, RetHe, ZPP, or hepcidin), except for CHr. **Outcomes:** • The two RCTs showed different findings regarding the dose of epoetin required to maintain the Hct target as the primary outcome, possibly due to differences in the Hct targets (an indication for the adequacy of anemia management). Thus, the findings from the two RCTs could not be combined together. Important intermediate outcomes suggested by the experts include: o Increase in Hb or hematocrit, or more consistent maintenance of Hb or hematocrit within the desired range o Use of ESA for maintenance of Hb within the desired range (stable dose in contrast to escalating dose resulting in net decreased ESA dose in hyporesponsive patients or actual decreased ESA dose in relatively responsive patients) o Adverse effects or harms associated with different management strategies

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

Key Questions	Research Gaps
Key Question 4. What factors affect the test performance and clinical utility of newer markers of iron status?	• Insufficient evidence to draw conclusions regarding factors that may affect the test performance or clinical utility of laboratory markers of iron status. Important factors suggested by the experts include: o Biological variation in diagnostic indices o Use of different diagnostic reference standards o Type of dialysis (i.e., peritoneal or hemodialysis) o Patient subgroups (i.e., age, sex, comorbid conditions, erythropoiesis-stimulating agent resistance, protein energy malnutrition secondary to an inflammatory state, hemoglobinopathies [e.g., thalessemia and sickle cell anemia]) o Route of iron administration (i.e., oral or intravenous) o Treatment regimen (i.e., repletion or continuous treatment) o Interactions between treatments (i.e., patients treated with vs. without ESA, patients treated with vs. without iron-replacement therapy) • All gaps described in Key Questions 2 and 3 are also applicable here.

Abbreviations: CHr = Hb content of reticulocytes; CKD = chronic kidney disease; ESA = erythropoiesis-stimulating agent; HB = hemoglobin; Hct = hematocrit; %HYPO = percentage of hypochromic erythrocytes; ND = nondialysis; PD = peritoneal dialysis; RCT = randomized controlled trial; RetHe = Reticulocyte hemoglobin equivalent; sTfR = soluble transferrin receptor; TSAT = transferrin saturation; ZPP = erythrocyte zinc protoporphyrin

Methods

Identification of Evidence Gaps

As the original authors of the CER, we generated the initial list of FRN topics based on the Research Needs section of the report, and then organized the list of evidence gaps by Key Question and PICOD (Population, Intervention, Comparator, Outcomes, and study Design) element. We then used an iterative process to identify additional FRN topics through Webinars and email correspondence with a stakeholder panel. The nominated topics were evaluated by EPC program staff to determine whether the nominated topics were related to one of the Key Questions of the CER or not. In general, nominated topics were deemed out-of-scope if the state of the evidence was not rigorously assessed as part of the CER. These out-of-scope topics were not entered to the final list of potential topics, but were rather enumerated in the Discussion of this report. We asked the stakeholder panel to prioritize the FRN topics following a formalized schema of prioritization criteria.

Criteria for Prioritization

Stakeholders (described in the next section) were asked to consider four dimensions of need as outlined in the Effective Health Care Program Selection Criteria: Importance, Desirability of Research/ Avoidance of Unnecessary Duplication, Feasibility, and Potential Impact.

Engagement of Stakeholders, Researchers, and Funders

To form the Stakeholder panel, we adapted a Tufts-developed "7Ps" model of stakeholder engagement,[6] which identifies seven primary stakeholder categories: Patients and the Public, Providers, Purchasers, Payers, Policymakers, Principal investigators, and Product makers. These categories are not necessarily mutually exclusive, and one stakeholder may belong to more than one category. For this study, product makers and purchasers were not included. Stakeholders were contacted by email with a brief description of the project and its purpose, a formal

invitation to serve as a stakeholder, and the executive summary of the original CER. Once their participation was confirmed, we sent additional materials to orient stakeholders including the Future Research Needs section of the original CER and a proposed outline for this project.

The first round of Webinars was held in June 2012. In these Webinars, we reviewed the purpose and processes of this project. We then reviewed the research gaps identified in the CER, and explained the criteria for prioritization. After this orientation, the stakeholders were asked to identify additional FRN topics of interest to them and to discuss their or others' suggested topics, the supporting rationale, or related research challenges. Topics nominated by stakeholders were incorporated into the topic list along with their supporting rationale, which we condensed from the discussion and subsequent emails. We combined duplicate or similar FRN topics together and disseminated the revised list of topics, along with minutes from all the Webinars, with an invitation to comment on the nominated topics and supporting rationales.

We held the second round of Webinars in late June and early July of 2012, during which we reviewed the refined list of nominated topics. Stakeholders were asked if their nominations were appropriately captured and accounted for, and were provided the opportunity to participate in further discussion.

Following this second round, we further edited the topic list based on stakeholder rationale. The finalized topic prioritization form and the minutes from the second set of webinars were sent to each stakeholder. Stakeholders were asked to review the minutes for accuracy and in order to hear the perspectives of others, and then to identify their *top five* priority topics along with corresponding justification for each topic based upon the Agency for Healthcare Research and Quality's (AHRQ's) Effective Health Care Program Selection Criteria. The five topics with the highest number of stakeholder endorsements were designated as the prioritized FRN.

Research Question Development and Research Design Considerations

We transformed the final list of FRN topics into research questions using standard PICOD criteria. (Stakeholders were not utilized during this process.) We discussed various alternatives for future research efforts aimed at answering each question, specifically considering the feasibility of addressing the potential research questions with respect to sample size, the time required, recruitment, and possible ethical concerns. In particular, we evaluated the advantages and disadvantages of various potential research designs with regards to:

Results

Based on the CER's Future Research section and our discussion with stakeholders, 17 Future Research Needs topics were nominated. We considered the five topics with endorsement by at least fifty percent of the nine voting stakeholders as the highest priority FRN topics. The topics chosen as the highest priority Future Research Needs are listed in Table B.

Table B. Top five Future Research Needs as indicated by participating stakeholders*

Topic	Topic Description
1	What is the best reference standard for diagnosing *(absolute or functional) iron deficiency in CKD patients*?
2	What is the diagnostic accuracy of newer markers of iron status as a replacement for, or an add-on to, classical lab markers in diagnosing *absolute iron deficiency*?
3	What is the diagnostic accuracy of newer markers of iron status as a replacement for, or an add-on to, classical lab markers in diagnosing *functional iron deficiency*?
4	Which is the best marker to *monitor the response to therapy and repletion status*?
5	What is the best biomarker to *monitor iron overload* as an adverse event resulting from treatment of iron deficiency?

*Prioritized topics (1–5) are ordered logically by clinical content.

Nomination of FRN Topic 1 highlights the current lack of a well-accepted reference standard for diagnosing iron deficiency in CKD patients, which is consistent with the findings of our CER. To compare the test performance among medical tests, a common reference standard is needed. When studies use different definitions of a reference standard, the results cannot be compared or "summed up" across studies. Without using the same definition of reference standard, conducting more studies on the test performance of existing or new medical test is unlikely to build up the body of evidence and therefore impact current practice.

Based on stakeholder discussion, it appears that iron staining of a bone marrow biopsy specimen is widely regarded as the "gold standard" for the diagnosis of (absolute) iron-deficient anemia, although this viewpoint is not universally accepted in the setting of CKD. Bone marrow iron may have limited clinical use due to the risks of infection or bleeding at the biopsy site. Despite these limitations, a bone marrow biopsy remains the most accurate measure that reflects stored iron, and thus should be used to define absolute iron deficiency. On the other hand, there is currently lack of a well-accepted reference standard for functional deficiency. Thus, we suggest that an expert panel be convened to standardize the definition for functional iron deficiency and determine which definition should be considered the preferred reference standard for diagnosing functional deficiency. The panel should also assess which intermediate outcomes (e.g., erythropoiesis-stimulating agent [ESA] or iron treatment dosages) or test characteristics (e.g., test availability/accessibility, cost) are appropriate to consider in determining the ideal definition for functional iron deficiency. In addition to representatives from all stakeholder categories, this panel should specifically include authoritative bodies and major professional organizations with relevant interests in iron deficiency, using a process similar to that used by the Centers for Disease Control and Prevention (CDC) Lipid Standardization Program (www.cdc.gov/labstandards/lsp.html).

FRN Topic 2 was Key Question 2 in the original CER. Although we did find studies comparing classical and newer tests for diagnosing absolute iron deficiency, they used classical laboratory biomarkers (alone or in combination with each other) as the reference standard for iron deficiency, essentially measuring the concordance between classical and newer biomarkers of iron status. Thus, we were unable to answer the question. We suggest that future research on this topic follow a prospective cohort design, as such studies would allow for multiple tests to be compared all together and potential biases could be minimized. In addition, depletion of bone marrow iron should be used as the reference standard for absolute iron deficiency, as iron staining of a bone marrow biopsy specimen is considered the most accurate measure that reflects stored iron. However, we expect that it will be difficult to recruit patients for such studies, because bone marrow aspiration is painful and poses risks to some patients.

FRN Topic 3 was also part of Key Question 2 in the original CER. Based on our post-hoc observation of this body of literature, we found that current studies often used a response to intravenous (IV) iron treatment as the reference standard for functional iron deficiency. 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. These variations in the reference standards in the published studies resulted in incomparable study results, and limited the strength of body of evidence. Therefore we suggest that it is vital to establish a preferred reference standard for functional iron deficiency (also see FRN Topic 1), before future research on this topic go forward. Future research on this topic should use the same reference standard, in order to grow the body of evidence on this important research question in a manner amenable to systematic review and meta-analysis, to compare results across studies.

Similar to FRN Topic 2, we also suggest that future research on this topic also utilize a prospective cohort design, and the agreed upon (by the consensus expert panel) definition should be used as the reference standard for diagnosing functional iron deficiency. If the reference standard for functional iron deficiency is defined by a response to ESA or iron treatment, a sufficient washout period (at least 4 weeks) is needed to stop ESA and iron treatment before ascertainment of the baseline test measurements. Stopping treatment may not be feasible for some clinical settings. In this case, studies should recruit only CKD patients who did not receive ESA and iron treatment within 1 month before baseline test measurements.

Our CER did not directly address FRN Topics 4 and 5, so we suggest that as a first step, prior to conducting new research studies on these topics, an expert panel be convened to determine and prioritize the elements needed for design future research on these topics (such as which biomarkers and outcomes are appropriate and relevant to consider in determining an ideal marker to monitor response to therapy and repletion status, or monitoring iron overload). We noted that markers that can separate cell hemoglobinization from cell production would provide researchers a better chance of understanding the underlying mechanism, potentially enabling better targeted treatments and therefore better managed anemia. Markers that have specificity to responses related to ESA or iron delivery would be optimal.

In terms of the expert panel composition, to balance viewpoints the panel should specifically include clinical chemists and hematologists. Following identification of the most appropriate biomarkers by the expert panel and determination of a reference standard, prospective cohort studies with consecutive CKD patients are again, as in our previous suggestions, the ideal study design that should be conducted for future research on these topics. Moreover, future research on these topics should assess clinical or patient-centered outcomes.

Discussion

The prioritization of topics for future research is a stakeholder-driven process. Our stakeholder panel represented a broad range of perspectives, from a well-informed patient advocate to clinical experts and policy makers. However, our stakeholder panel is unlikely to represent all perspectives because we did not use formal sampling methods to select our stakeholder members. The process of engagement though multiple teleconferences enabled us to get a well-rounded perspective; and we believe that the process of sharing the minutes of the discussions with all stakeholders enabled them to appreciate one another’s viewpoints. We were able to obtain input from all members of the stakeholder panel, and the final list showed a clear preference for the top five priorities.

There are several challenges for designing future research on the priority topics. The ideal study design and sampling populations would vary depending on the purpose of using a medical test (e.g., screening, diagnosis, prognosis, monitoring, etc.), in order to maximize the internal and external validity of a study. One of the challenges in the current FRN project is that an iron status test can be used for the purposes of screening, diagnosis, and/or monitoring for iron deficiency anemia in CKD patients. However, the same iron status tests can be used for the purposes of diagnosis and monitoring without considering the biases in interpretations of test results that are highly likely in these instances. Another challenge is that patients' iron status can change spontaneously (due to changes in diet, or due to unrelated metabolic and inflammatory conditions) or due to the treatment received. Therefore, it is important to control for these confounding factors in future research studies, which will require large sample sizes to reach sufficient statistical power.

Conclusions

Our CER and FRN stakeholder discussions pointed out the top future research gaps, and highlighted the great deal of confusion/uncertainty in using newer and/or classical laboratory biomarkers of iron status for the purposes of the diagnosis or monitoring of iron deficiency anemia in CKD patients. The chief factor causing this confusion and uncertainty is the lack of a well-accepted reference standard for iron deficiency anemia. The most effective first step would be to establish a common reference standard for iron deficiency anemia, considering two separate and distinct definitions: absolute versus functional iron deficiency. The ideal reference standard should be independent of the index tests and test-directed treatment to maximize the internal validity of study results.

References

1. Anaemia Management in Chronic Kidney Disease - Rapid Update 2011. National Clinical Guideline Centre; 2011. www.ncbi.nlm.nih.gov/books/NBK65530/pdf/TOC.pdf.
2. 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. PMID: 4430129.
3. 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. PMID: 1239188.
4. Wish JB. Assessing iron status: beyond serum ferritin and transferrin saturation. Clin J Am Soc Nephrol 2006 Sep;1 Suppl 1:S4-S8. PMID: 17699374.
5. Chung M, Moorthy D, Hadar N, et al. Biomarkers for Assessing and Managing Iron Deficiency Anemia in Late-Stage Chronic Kidney Disease. Comparative Effectiveness Review No. 83. (Prepared by the Tufts Evidence-based Practice Center under Contract No. 290-2007-10055-I.) AHRQ Publication No. 12(13)-EHC140-EF. Rockville, MD: Agency for Healthcare Research and Quality. October 2012. www.effectivehealthcare.ahrq.gov/reports/final.cfm.
6. Concannon T, Meissner P, Grunbaum J, et al. A New Taxonomy for Stakeholder Engagement in Patient-Centered Outcomes Research. Journal of General Internal Medicine 2012:1-7. PMID: 22528615

Background

Context

Anemia is a common complication of chronic kidney disease (CKD), which develops early in the course of CKD and becomes increasingly severe as kidney function deteriorates.[1] Iron deficiency anemia is a continuous process evolving in three stages. The first phase is the depletion of storage iron (stage I), where total body iron is decreased but hemoglobin (Hb) synthesis and red cell indices remain unaffected. Both these indices change when the supply of iron to bone marrow becomes problematic (iron deficient erythropoiesis, or stage II). In stage III the iron supply is insufficient to maintain a normal Hb concentration, and eventually iron deficiency anemia develops.

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 Hb production.[2] It is important to assess iron stores and the availability of iron for erythropoiesis, as adequate iron status is integral to both iron and anemia management in CKD patients. The major cause of iron deficiency is blood loss, particularly for dialysis patients. Dialysis patients are in a state of continuous iron loss from gastrointestinal bleeding (very common), blood drawing, and/or, most importantly, with hemodialysis, the dialysis treatment itself. A CKD patient who receives treatment with erythropoietic-stimulating agents (ESAs) for the anemia often develops iron deficiency, because the iron requirements for achieving a response to ESA treatment usually cannot be met by mobilization of the patient's iron stores alone. Therefore, supplemental iron therapy, either given orally or intravenously, is often needed among dialysis patients who receive recombinant human erythropoietin (EPO) or darbepoetin alfa treatment.[1] Thus, iron management (iron status assessment and iron treatment) is an essential part of the treatment of anemia associated with CKD, as there are concerns regarding the adverse effects associated with both elevated doses of ESAs and supplemental (intravenous or oral) iron.

Classical iron status tests, of which ferritin and transferring saturation (TSAT) are the most widely used, reflect either the level of iron in tissue stores or the adequacy of iron for erythropoiesis. 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.[3-5] Furthermore, results from a meta-analysis of 55 studies (published before 1990) showed that ferritin radioimmunoassay was the most powerful test (a mean area under the receiver operating characteristic curve of 0.95; 95% CI 0.94, 0.96), compared with mean cell volume determination, TSAT, and red cell volume distribution, for diagnosing adult patients with iron deficiency, but test performances varied between patients with and those without inflammatory (e.g., CKD patients) or liver disease.[6] The accurate assessment of iron status is dependent on the validity and reliability of laboratory test results, and differences in test performance pose a dilemma regarding the most appropriate test to guide treatment decisions.

In an attempt to find a more accurate and reliable test, several novel biomarkers of iron status have been proposed:

- **The Hb content of reticulocytes (CHr)/Reticulocyte hemoglobin equivalent (RetHe).** CHr and RetHe measurements are functionally equivalent,[7] 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);[8] 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.[9] 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.[10]
- **Soluble transferrin receptor (sTfR).** 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.
- **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.[11]
- Superconducting Quantum Interference Devices (SQUIDs) are a non-invasive method for the detection and quantification of liver iron content.[12] They operate by exploiting the paramagnetic properties of iron: magnetic resonance diminishes in the liver as iron concentration increases.

Scope of Comparative Effectiveness Review

Although a number of international guidelines have examined the use of both classical and new serum iron biomarkers, their recommendations differ.[1,2,13] In view of the considerable clinical uncertainty, the high biological variability associated with laboratory biomarkers, and the need for frequent assessment to guide treatment with ESAs, a systematic review of the relevant literature was deemed to be a priority. In order to address this knowledge gap, the Tufts Evidence-based Practice Center (EPC) conducted a Comparative Effectiveness Review (CER) to systematically evaluate the impact on patient-centered outcomes of using newer laboratory biomarkers[a] as a replacement for or as an add-on to older laboratory biomarkers of iron status[b] for assessing iron status and the management of iron deficiency in adult and pediatric CKD patients (nondialysis and dialysis).[14] Although studies that assess the overall impact of these tests on the clinical management process would provide the most direct evidence for this CER, they are often challenging or unfeasible to conduct due to the high patient and resource requirements.

[a] Newer laboratory biomarkers: content of Hb in reticulocytes, percentage of hypochromic red blood cells, erythrocyte zinc protoporphyrin, soluble transferrin receptor, hepcidin, and superconducting quantum interference devices.

[b] Older laboratory biomarkers: bone marrow iron stores, serum iron, transferrin saturation, iron-binding capacity, and ferritin.

Because it was expected that little such evidence would be found, 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 that addresses these three component Key Questions could thus inform the conclusions for the review's primary, overarching question (Key Question 1).

Key Question 1 (Overarching Question)

What is the impact on patient centered outcomes of using the newer laboratory biomarkers as a replacement for or an add-on to the older laboratory biomarkers of iron status for 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 as a replacement for or an add-on to the older markers in stages 3-5 CKD patients nondialysis and dialysis, and in patients with a kidney transplant?

a. What reference standards are used for the diagnosis of iron status 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., thalessemia 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 vs. without iron-replacement therapy)

- Other factors (based on additional information in the reviewed papers)

Each question had specific criteria for study inclusion based on the population, intervention, comparator, outcomes, and study design. Population criteria included studies in both adults and children with stage 3, 4, or 5 CKD; patients with CKD undergoing dialysis (hemodialysis [HD] or peritoneal dialysis [PD]); and patients with a kidney transplant. For interventions, eligible studies were those involving newer laboratory biomarkers 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 older laboratory biomarkers to diagnose and manage iron deficiency. We were interested in both patient-centered outcomes (such as mortality and morbidity [e.g., cardiac or liver toxicity and infection], quality of life, and adverse events or harms) and intermediate outcomes (such as improvement in Hb levels, dose of erythropoiesis-stimulating agents, and time in target Hb range).

Comparative Effectiveness Review Findings

A total of 30 articles met the study eligibility criteria, based on the populations, tests, and outcomes of interest, 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);[7,15-40] two reported intermediate outcomes comparing iron management guided by newer laboratory markers with iron management guided by classical markers (Key Question 3);[39,41] and three (in two articles) reported data on factors affecting test performance comparing newer with classical laboratory markers of iron status (Key Question 4).[42,43] Most studies enrolled only adult CKD patients undergoing HD. The main findings of this CER are presented in Table 1.

Combining the evidence addressing Key Questions 2, 3, and 4, it was concluded that there is insufficient data to determine if most newer laboratory biomarkers of iron status are better than classical markers for predicting 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 intravenous (IV) iron treatment than classical markers (TSAT <20% or ferritin <100 ng/mL) in HD CKD patients. In addition, results from two randomized controlled trials (RCTs) showed a reduction in the number of iron status tests and resulting IV iron treatments administered to patients whose iron management was guided by CHr compared with those guided by TSAT or ferritin. These results suggest that CHr may reduce potential harms from IV iron treatment by lowering the frequency of iron testing, although the evidence for the potential harms associated with testing or test-associated treatment is insufficient.

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

Table 1. Strength of evidence addressing Key Questions

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

Table 1. Strength of evidence addressing Key Questions (continued)

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

Abbreviations: CHr = Hb content of reticulocytes; CKD = chronic kidney disease; ESA = erythropoiesis-stimulating agent; Hct = hematocrit; HB = hemoglobin; HD = hemodialysis; %HYPO = percentage of hypochromic erythrocytes; IV = intravenous; ND = nondialysis; PD = peritoneal dialysis; RCT = randomized controlled trial; RetHe = Reticulocyte hemoglobin equivalent; sTfR = soluble transferrin receptor; TSAT = transferrin saturation; ZPP = erythrocyte zinc protoporphyrin

Identification of Evidence Gaps

The current Future Research Needs (FRN) project was undertaken in order to find the most important research gaps in the literature with regards to laboratory tests for assessing iron status and management of iron deficiency in CKD patients, identified during the synthesis of the aforementioned CER. The objectives of this project are to identify potential research questions and to suggest study designs for addressing these questions. These objectives were achieved by cataloguing the evidence gaps relevant to iron deficiency laboratory tests, establishing a stakeholder panel, engaging stakeholders in research topic nomination and prioritization, and developing research protocols for the most highly ranked topics. Table 2 summarizes the evidence gaps identified in our review. (Note: The gaps are not listed in the order of the CER Key Questions.)

One major evidence gap concerns the dearth of pediatric studies (only one study enrolled pediatric HD and PD CKD patients).[22] This suggests that there is a need for further refinement of FRN for pediatric CKD patients (nondialysis and dialysis). Since this would require a specially-composed stakeholder group, it was determined to be beyond the scope of this project. Thus, the current FRN project is focused on adult CKD patients (nondialysis and dialysis).

Table 2. Research gaps and suggestions for future research

Key Questions	Research Gaps
Key Question 1 (overarching question). What is the impact on patient centered outcomes of using the newer laboratory biomarkers as a replacement for or an add-on to the older laboratory biomarkers of iron status for assessing iron status and management of iron deficiency?	No studies were found that provided data directly addressing our overarching question regarding the impact of using newer laboratory biomarkers on patient-centered outcomes (e.g., mortality, morbidity, quality of life, and adverse effects). All gaps described in Key Questions 2, 3, and 4 are applicable to this overarching question.
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? 2a. What reference standards are used for the diagnosis of iron status in studies evaluating test accuracy? 2b. What are the adverse effects or harms associated with testing using newer and/or older markers of iron status?	**Populations**: • Insufficient evidence in adult PD CKD and ND CKD patients. • Whether test performance and clinical utility of newer or classical markers of iron status vary by different CKD populations is unknown. **Interventions and comparators**: • Insufficient evidence for the test performance of newer markers of iron status as an add-on to older markers. **Outcomes**: • There is a lack of a generally accepted reference standard test for determining iron deficiency in the setting of CKD.[2] • A small percentage (26%) of the identified studies reported information on harms, and most studies did not attribute harms to either testing or treatment. **Study Design Issues**: • Many existing studies are at a high risk of bias due to lack of demographic details of study populations and blinding of the diagnosis or patient characteristics. These biases limit studies' utility in informing clinical practice. • Existing studies were underpowered leading to imprecise estimates.
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? 3a. What are the adverse effects or harms associated with the treatments guided by tests of iron status?	**Populations**: • Insufficient evidence in adult PD CKD and ND CKD patients. **Interventions and comparators**: • There are no uniform iron management algorithms or test-and-treat protocols across studies. • There are lack of studies with longer followup durations; exsiting short-term RCTs had a followup duration less than 6 months. • No study compared iron management guided by classical markers with that of newer markers (%HYPO, sTfR, RetHe, ZPP, or hepcidin), except for CHr. **Outcomes**: • The two RCTs showed different findings regarding the dose of epoetin required to maintain the Hct target as the primary outcome, possibly due to differences in the Hct targets (an indication for the adequacy of anemia management). Thus, the findings from the two RCTs could not be combined together. Important intermediate outcomes suggested by the experts include: o Increase in Hb or hematocrit, or more consistent maintenance of Hb or hematocrit within the desired range o Use of ESA for maintenance of Hb within the desired range (stable dose in contrast to escalating dose resulting in net decreased ESA dose in hyporesponsive patients or actual decreased ESA dose in relatively responsive patients) o Adverse effects or harms associated with different management strategies

Table 2. Research gaps and suggestions for future research (continued)

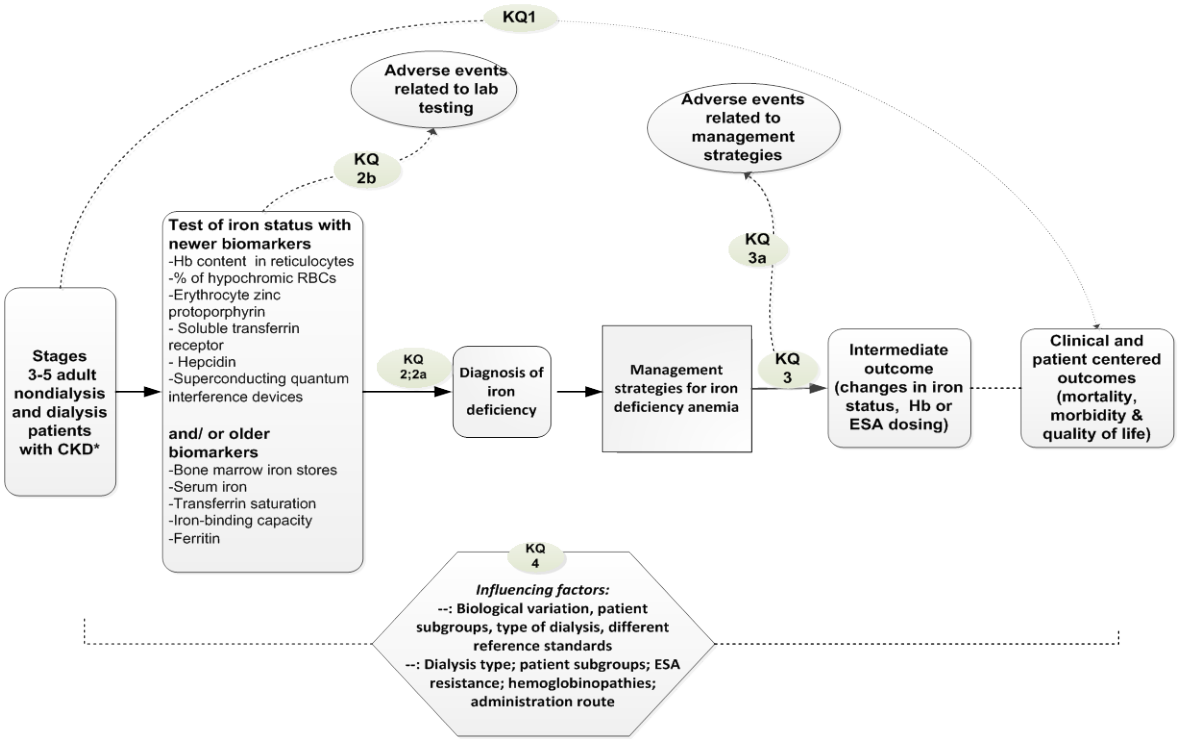
Key Questions	Research Gaps
Key Question 4. What factors affect the test performance and clinical utility of newer markers of iron status?	• Insufficient evidence to draw conclusions regarding factors that may affect the test performance or clinical utility of laboratory markers of iron status. Important factors suggested by the experts include: o Biological variation in diagnostic indices o Use of different diagnostic reference standards o Type of dialysis (i.e., peritoneal or hemodialysis) o Patient subgroups (i.e., age, sex, comorbid conditions, erythropoiesis-stimulating agent resistance, protein energy malnutrition secondary to an inflammatory state, hemoglobinopathies [e.g., thalessemia and sickle cell anemia]) o Route of iron administration (i.e., oral or intravenous) o Treatment regimen (i.e., repletion or continuous treatment) o Interactions between treatments (i.e., patients treated with vs. without ESA, patients treated with vs. without iron-replacement therapy) • All gaps described in Key Questions 2 and 3 are also applicable here.

Abbreviations: CHr = Hb content of reticulocytes; CKD = chronic kidney disease; ESA = erythropoiesis-stimulating agent; Hct = hematocrit; HB = hemoglobin; %HYPO = percentage of hypochromic erythrocytes; ND = nondialysis; PD = peritoneal dialysis; RCT = randomized controlled trial; RetHe = Reticulocyte hemoglobin equivalent; sTfR = soluble transferrin receptor; ZPP = erythrocyte zinc protoporphyrin

Analytic Framework

Figure 1 depicts the analytic framework used in structuring the CER as well as the FRN report. Broadly, it shows how the individual Key Questions are addressed within the context of the logical linkages between populations, interventions, comparators, and outcomes of interest.

Figure 1. Analytic framework



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

*The Comparative Effectiveness Review included both adult and pediatric CKD patients, but the current Future Research Needs project is focused on adult CKD patients (nondialysis and dialysis) only.

Methods

Identification of Evidence Gaps

As the original authors of the CER, we generated the initial list of FRN topics based on the Research Needs section of the report, and then organized the list of evidence gaps according to Key Questions and PICOD (Population, Intervention, Comparator, Outcomes, and study Design) elements. We then used an iterative process to identify additional FRN topics through Webinars and emails with a stakeholder panel (see Appendix A). The nominated topics were evaluated by EPC program staff to determine whether the nominated topics were related to one of the Key Questions of the CER. In general, nominated topics were deemed out-of-scope if the state of the evidence was not rigorously assessed as part of the CER. These out-of-scope topics were not entered to the final list of potential topics, but listed in the Discussion of this report. We asked the stakeholder panel to prioritize the FRN topics following a formalized schema of prioritization criteria.

Criteria for Prioritization

Stakeholders (described in the next section) were asked to consider four dimensions of need. These four dimensions are outlined in the Effective Health Care Program Selection Criteria (Appendix B). Briefly, they are:

- Importance
- Desirability of Research/Avoidance of Unnecessary Duplication
- Feasibility
- Potential Impact

The fifth dimension, Appropriateness, was not evaluated by the stakeholders, because the initial FRN topics based on the Research Needs section of the CER were already deemed to adequately meet this criterion by the Agency for Healthcare Research and Quality (AHRQ).

To further inform the selection criteria of Desirability of Research/Avoidance of Unnecessary Duplication, we also searched the National Library of Medicine Clinical Trial Registry (www.ClinicalTrials.gov) to identify ongoing or recently completed trials relevant to the CER questions (Appendix C). Relevant recently completed or registered studies identified in the searches were compared against the nominated FRN topics to assess if they would make future research on any nominated topic redundant, but none were judged to do so.

Engagement of Stakeholders, Researchers, and Funders

Although researchers and funders of research are the primary audience for FRN documents, the EPC solicits input from other stakeholders as well when identifying high-priority research gaps and FRN. Stakeholders are selected to provide broad expertise and a breadth of perspectives, as well as input on the kind of information that is helpful in health care decisionmaking. These stakeholders are engaged throughout the future research process. Their role is to (1) review the preliminary list of evidence gaps and possible future research topics derived from them, (2) to nominate additional topics to the list, (3) to discuss topic nominations, and (4) to participate in prioritization of the FRN topics. Stakeholders are not involved in translating the gaps into research questions and study designs, or composing or reviewing the

report. The final FRN document will be released for public comment. Stakeholders who participated in the FRN identification process are invited to provide comments on this report during the public posting. Public input may be incorporated into or otherwise reflected in the final report.

To form the panel, we adapted a Tufts-developed "7Ps" model of stakeholder engagement,[44] which identifies seven primary stakeholder categories. The stakeholder categories of the 7Ps model consist of:

1. **Patients and the Public.** This group represents current and potential consumers of patient-centered health care and population-focused public health programs. This group also includes caregivers, family members, and patient advocacy organizations, all of whom represent the interests of consumers or patients.
2. **Providers.** This group includes individuals (e.g., nurses, physicians, and other providers of care and support services) and organizations (e.g., hospitals, clinics, community health centers, pharmacies, emergency medical services agencies, schools) that provide care to patients and populations.
3. **Purchasers.** This group includes employers; the self-insured; Federal, state, and local governments; and other entities responsible for underwriting the costs of health care.
4. **Payers.** This group represents private insurers, government insurers (e.g., Medicare, Medicaid, the Veterans Administration), and others responsible for reimbursement for care.
5. **Policymakers.** This group includes entities such as the legislative and executive branches of the Federal and state governments, professional associations, and other intermediary groups that collect and distribute information to policymakers.
6. **Principal investigators.** This group consists primarily of researchers, and research funders.
7. **Product makers.** This group consists primarily of manufacturers and device makers.

These categories are not necessarily mutually exclusive, and one stakeholder may belong to more than one category.

For this project, we recruited stakeholders representing Patients and the Public, Providers, Payers, Policymakers, and Principal investigators and asked them to provide input foremost according to their designated stakeholder category. Product makers were not included because it was not possible to include all product makers (particularly since there are many laboratory-based tests available in individual labs), and selected participation might bias the FRN discussions. Purchasers (employers who purchase insurance policies) were not included because they were considered to share the payer perspective for the diagnostic tests in question.

We identified individuals to serve on the stakeholder panel through several means. We invited some individuals who had previously served in advisory roles on the original CER. We contacted the National Kidney Foundation Patient and Family Council to invite CKD patient/patient advocate stakeholders. We used professional contacts to identify public and private payers and a policy maker. Individuals were selected based on their particular stakeholder perspective. Our predetermined goal was to assemble a representative panel of 10 stakeholders across the appropriate stakeholder categories (Table 3).

Table 3. List of predetermined target stakeholders

Category	Subcategory	No. of Invited Stakeholders
Patients and the public	NKF Patient & Family Council	2
Providers	Clinicians – Nephrologist	3
	Clinical Chemist (Nominator)	
	Allied Health - Dietician	
Payers	Private Insurer	2
	CMS	
Policy Makers	NIDDK – Division of Kidney, Urologic & Hematologic Diseases	1
Principal investigators/researchers	Clinical Researcher	2
TOTAL		10

Abbreviations: CMS = Center for Medicare and Medicaid Services; NIDDK = National Institute of Diabetes and Digestive and Kidney Disease; NKF = National Kidney Foundation

All individuals involved in the project were required to submit a standard disclosure of interest form. Participation was only confirmed after review of the disclosure form. Stakeholders were asked to disclose any financial conflicts of interest greater than $10,000, and any other relevant business or professional conflicts of interest.

Individuals who met the criteria to participate as a stakeholder were contacted by email with a brief description of the project and its purpose, a formal invitation to serve as a stakeholder, and the executive summary of the original CER. Once their participation was confirmed, we sent additional materials to orient stakeholders, including the Future Research Needs section of the original CER and a proposed outline for this project.

The first round of Webinars was held in June 2012. All stakeholders attended. (Four duplicate Webinars were held to maximize participation.) During these Webinars, we reviewed the purpose and processes of this project. Subsequently stakeholders were asked to self-identify as a representative of a particular stakeholder category and directed to provide feedback chiefly with respect to their primary stakeholder category. We also directed stakeholders to restrict their FRN nominations to topics within the scope of the CER, specifically delineating out-of-scope topics such as pediatric CKD patient and cost effectiveness and cost utility of tests (based on the topic and study eligibility criteria of the CER). We then reviewed the research gaps identified in the CER and explained the criteria for prioritization. After this orientation, the stakeholders were asked to identify additional FRN topics of interest to them and to discuss their or others' suggestions, their supporting rationale, or related research challenges. Stakeholders were also given the option of nominating additional topics by email afterwards.

Topics nominated by stakeholders were incorporated into the topic list along with supporting rationale, which we condensed from the discussion and subsequent emails. We combined duplicate or similar FRN topics together and disseminated the revised list of topics, along with minutes from all the webinars, with an invitation to comment as to whether the nominated topics and supporting rationales were appropriately recorded and accounted for.

We held a second round of Webinars in late June and early July 2012, during which we reviewed the refined list of nominated topics. Again all stakeholders attended the second Webinar (held in duplicate.) Stakeholders were asked if their nominations were appropriately captured and accounted for, and were provided the opportunity for further discussion.

Following this second round, we further edited the topic list based on stakeholder input. The finalized topic prioritization form and the minutes from the second set of webinars were sent to each stakeholder. Stakeholders were asked to review the minutes for accuracy and in order to hear the perspectives of others, and then to identify their *top five* priority topics with corresponding justification for each topic based upon AHRQ's Effective Health Care Program Selection Criteria (Appendix B). The five topics with the highest number of stakeholder endorsements were designated as the prioritized FRN topics.

Use of the Effectiveness Health Care Program prioritization criteria was repeatedly emphasized throughout the prioritization process, including during discussion, nomination, and final topic selection. Upon the close of stakeholder prioritization, we identified the top five topics as those most frequently endorsed by stakeholders in their top five selections.

Research Question Development and Research Design Considerations

We transformed the final list of FRN topics into research questions using standard PICOD criteria. (Stakeholders were not utilized during this process.) We discussed various alternatives for future research efforts aimed at answering each question, specifically considering the feasibility of addressing the potential research questions with respect to sample size, the time required, recruitment, and possible ethical issues.

Results

Research Needs

We successfully enlisted the predetermined panel of stakeholders; however, despite multiple attempts, we were unable to recruit a stakeholder representing private insurers within the project timeline. A total of 10 stakeholders participated in the 2 sets of teleconference calls, and 9 stakeholders prioritized the list of FRN topics (the two stakeholders representing Center for Medicare and Medicaid Services jointly prioritized the final list of FRN topics; see Table 4).

Table 4. List of participating stakeholders

Category	Subcategory	No. of Stakeholders Who Prioritized the Final List of FRN topics
Patients and the public	NKF Patient & Family Council	2
Providers	Clinicians – Nephrologist	3
	Clinical Chemist (Nominator)	
	Allied Health - Dietician	
Payers	Private Insurer (0 participated)*	1
	CMS (2 participated)**	
Policy Makers	NIDDK – Division of Kidney, Urologic & Hematologic Diseases	1
Principal investigators/researchers	Clinical Researcher	2
TOTAL		9

Abbreviations: CMS = Center for Medicare and Medicaid Services; NIDDK = National Institute of Diabetes and Digestive and Kidney Disease; NKF = National Kidney Foundation

*Invited but declined.

**There were two stakeholders representing CMS but they jointly prioritized the final list of FRN topics.

Initially, we presented 10 FRN topics based on the CER Future Research section to the stakeholder panel for discussions. With input from the stakeholder panel, seven new topics were added. Of these, four were determined to be out of the scope of the current FRN project (see Discussion). Thus, our FRN identification process led to the nomination of 17 topics (Table 5).

We considered the five topics with endorsement by at least fifty percent of the nine voting stakeholders as the highest priority FRN topics. Two of the five topics are directly based on the evidence gaps identified in the CER, and the other three were nominated primarily based on the evidence gaps identified in the CER, with more refined definitions of iron deficiency and adverse outcomes.

Table 5. Nominated topics for Future Research Needs

Topic Rank*	Topic	No. of Stakeholder Votes
High-priority FRN Topics		
1	What is the best reference standard for diagnosing *(absolute or functional) iron deficiency in CKD patients*?	5
2	What is the diagnostic accuracy of newer markers of iron status as a replacement for, or an add-on to, classical lab markers in diagnosing *absolute iron deficiency*?	6
3	What is the diagnostic accuracy of newer markers of iron status as a replacement for, or an add-on to, classical lab markers in diagnosing *functional iron deficiency*?	5
4	Which is the best marker to *monitor the response to therapy and repletion status*?	5
5	What is the best biomarker to *monitor iron overload* as an adverse event resulting from treatment of iron deficiency?	5
Other Nominated FRN Topics		
6	What is the impact of managing iron status based on newer laboratory biomarkers, either alone or in addition to classical laboratory biomarkers, on patient-centered outcomes?	4
7	How does the test performance and clinical utility of iron status vary according to treatment subgroups (i.e. nondialysis CKD, hemodialysis, and peritoneal dialysis patients)?	3
8	What is the impact of managing iron status based on newer laboratory biomarkers, either alone or in addition to classical laboratory biomarkers, on intermediate outcomes?	2
9	What is the diagnostic accuracy of newer markers of iron status as a replacement for, or an add-on to, classical lab markers for iron deficiency?	2
10	Does the test performance and clinical utility of markers of iron status vary in erythropoiesis-stimulating agents and iron treatment-resistant patients?	2
11	What are the adverse effects or harms associated with testing or with treatments guided by tests of iron status?	2
12	Does the test performance and clinical utility of iron status vary according to patients with different comorbidities (e.g., diabetes, hypertension)?	2
13	Does the test performance and clinical utility of markers of iron status vary by CKD disease stage?	1
14	Does the test performance and clinical utility of markers of iron status vary by patient race/ethnicity?	1
15	In hemodialysis patients, does the test performance and clinical utility of markers of iron status vary by method of vascular access (i.e., catheter, arterial-venous shunt)?	0
16	Does the test performance and clinical utility of markers of iron status vary by sex?	0
17	Does the test performance and clinical utility of markers of iron status vary by patient age?	0

Abbreviation: CKD = chronic kidney disease

*Prioritized topics (1–5) are ordered logically by clinical content. Other nominated topics are listed in the order they were prioritized by the stakeholder panel.

High-Priority Future Research Needs Topic 1

What is the best reference standard for diagnosing (absolute or functional) iron deficiency in CKD patients?

Background

When evaluating iron deficiency in patients with CKD, it is useful to consider two separate and distinct definitions: absolute versus functional iron deficiency. Absolute/true/overt iron deficiency refers to the depletion of iron stores and the absence of stainable iron in the bone marrow, while functional iron deficiency refers to a clinical condition where stored iron is sufficient but circulating iron is not.[45] Functional iron deficiency can occur when ESA therapy stimulates red blood cell production beyond the available supply of iron necessary for hemoglobin synthesis; it can also be caused by chronic inflammation. Unlike absolute iron deficiency, the concept of functional iron deficiency does not imply depletion of iron stores.

Rather, iron stores may be nearly normal, but during ESA treatment, there may be insufficient immediately available iron to optimize ESA therapy. In ESA-treated CKD patients, the "treat-to-target" or "treat-to-goal" levels of iron tests are often used to define functional iron deficiency.[45]

The original CER found that, currently, there is a lack of generally accepted reference standard tests for determining iron deficiency in the setting of CKD.[2] In the 27 studies that were included for Key Question 2 (What reference standards are used for the diagnosis of iron deficiency in studies evaluating test performance?), studies used two distinct methods to operationalize a reference standard for assessing the diagnostic test accuracy: (1) classical laboratory biomarkers (TSAT, ferritin, or serum iron), alone or in combination with each other; and (2) a response to intravenous iron treatment. However, across studies, there are large variations in the definitions of these reference standards. Of the 15 studies that used classical markers of iron status to define "iron deficiency" as the reference standard in calculating the test accuracy of newer markers of iron status, the most commonly used definition was TSAT $\leq$20% and ferritin $\leq$100 ng/mL (7 studies), and TSAT $\leq$20% (7 studies). Of the 12 that used a response to IV iron treatment as the reference standard, the most commonly used definition for a response to IV iron treatment was an increase in hemoglobin concentration $\geq$1 g/dL after a (variable) period of IV iron treatment. 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.

To compare the test performance among medical tests, a common reference standard is needed. When studies used different definitions of a reference standard, the results cannot be compared nor "summed up" across studies. Synthesizing information on test performance metrics, such as sensitivity, specificity, predictive values, and likelihood ratios, is often an important part of a systematic review of a medical test, and it has a dual goal: to provide summary estimates for key quantities, and to explore and explain any observed dissimilarity (heterogeneity) in the results of the examined studies. Without using the same definition of reference standard, conducting more studies on the test performance of existing or new medical test is unlikely to add meaningfully to the body of evidence and therefore impact current practice.

Stakeholder Discussion

This FRN topic was raised during discussions with the stakeholder panel on the issues surrounding the lack of generally accepted reference standard tests for determining iron deficiency in the setting of CKD.[2] Generally, stakeholders agreed that it is important to use the same reference standard for iron deficiency in future research. Clinician stakeholders pointed out that, from a hematologist's perspective, bone marrow iron would be considered as the "gold standard" for diagnosing (absolute) iron deficiency. In contrast, a nephrologist needs to consider a reference standard that assesses iron storage as well as its availability or adequacy (i.e., both storage and function). Clinician stakeholders suggested that future research should focus on "functional iron deficiency." (Note: stakeholders did not provide specific definition for functional iron deficiency) However, they also mentioned that "a response to IV iron treatment" (one of the operationalized reference standards used in many existing studies, and often referred as the reference standard for functional iron deficiency in dialysis patients) may be a different research question than the reference standard for the diagnosis of iron deficiency.

Based on the stakeholders' discussions, it appears that iron staining of a bone marrow biopsy specimen is widely regarded as the "gold standard" for the diagnosis of (absolute) iron-deficient

anemia;[46] however, this viewpoint is not universally accepted in the setting of CKD.[2] Moreover, taking a bone marrow sample is invasive and carries the risks of infection or bleeding at the biopsy site.[47] These concerns limit the use of bone marrow iron tests in clinical practice. Despite these limitations, a bone marrow biopsy remains the most accurate measure that reflects stored iron, and thus should be used to define absolute iron deficiency (also see FRN topic 2).

For diagnosing functional iron deficiency, bone marrow iron may not be useful, because it reflects stored iron but not the iron readily available for erythropoiesis in patients treated with ESAs. To date, the only accepted means of measuring functional iron deficiency in clinical practice is to evaluate the erythropoietic response to iron administration.[46] The original CER found that there was no uniform regimen of IV iron in terms of dosage and iron formulation, and that there was a wide range of durations of IV iron treatment in published studies.

Proposed Study Designs

Existing evidence only allows indirect comparisons across studies, and cannot provide valid information on which definition of response to IV iron (or functional iron deficiency) is the "best" (or most accurate) reference standard due to large heterogeneity across study populations. Given that currently there is a lack of generally accepted reference standard for functional iron deficiency, we propose a *Consensus Expert Panel* to determine the "best" (or most practical) reference standard for diagnosing functional iron deficiency in CKD patients. The goal of this panel is to balance different concerns and to develop consensus across disciplines about how research should be conducted to address this high-priority FRN topic.

Consensus Expert Panel

Value of Study Design

We suggest that prior to conducting new research studies to assess other high-priority FRN topics (topics 3 through 5), an expert panel be convened to standardize the definition for functional iron deficiency and to determine which definition should be considered most suited for use as the reference standard in future research in CKD patients. Ideally, the "best" definition for functional iron deficiency would be highly correlated with patient-centered outcomes, such as mortality, morbidity, or quality of life. The panel should also assess which intermediate outcomes (e.g., ESA or iron treatment dosages) or test characteristics (e.g., test availability/accessibility, cost) are appropriate to consider in determining the ideal definition for functional iron deficiency. In addition to representatives from all stakeholder categories, this panel should specifically include authoritative bodies and major professional organizations that have relevant interests in iron deficiency, using a process similar to that employed by the Center for Disease Control and Prevention (CDC) Lipid Standardization Program (www.cdc.gov/labstandards/lsp.html).

Resource Use, Size, and Duration

A consensus expert panel would require relatively few resources, as it would rely on mostly volunteerism, and most of the required resources would be associated with the logistics for in-person conferences or scientific meetings.

Ability to Recruit

There should be little difficulty recruiting relevant stakeholders.

Ethical Issues

There should be no ethical issues in using an expert panel to investigate this evidence gap.

High-Priority Future Research Needs Topic 2

What is the diagnostic accuracy of newer markers of iron status as a replacement for, or an add-on to, classical lab markers in diagnosing absolute iron deficiency?

Background

This FRN topic was part of Key Question 2 in the original CER. Although the CER did find studies comparing classical and newer tests for diagnosing absolute iron deficiency, the identified studies used the classical laboratory biomarkers (alone or in combination with each other) as the reference standard for iron deficiency, essentially measuring the concordance between classical and newer biomarkers of iron status. 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 thus the CER was unable to provide an answer to the question.

Stakeholder Discussion

Much of the discussions with the stakeholder panel revolved around the issues regarding the lack of generally accepted reference standard tests for determining iron deficiency in the setting of CKD (described earlier in FRN topic 1). In brief, iron staining of a bone marrow biopsy specimen is widely regarded as the "gold standard" for the diagnosis of (absolute) iron-deficient anemia,[46] but taking a bone marrow sample is invasive and carries the risks of infection or bleeding at the biopsy site. These concerns limit the use of bone marrow iron tests in clinical practice. Despite these limitations, bone marrow iron remains the most accurate measure that reflects stored iron, and thus should be used to define absolute iron deficiency. In the original CER, no study was found examining the test accuracy of new markers of iron status as a replacement for, or an add-on to, classical laboratory biomarkers against bone marrow iron as the reference standard in late-stage CKD patients.

Proposed Study Designs

Prospective cohort studies are the ideal study design for future research on this topic, as they allow for multiple tests to be compared all together (which may not be feasible in randomized trials), and potential biases, such as spectrum bias and review bias (i.e., interpretation of the index test or reference standard influenced by knowledge of the results of the other test), can be minimized. Prospective cohort studies should recruit consecutive CKD patients who consent to have bone marrow aspiration for histologic examination. Depletion of bone marrow iron should be used as the reference standard for absolute iron deficiency (as described earlier in the Stakeholder Discussion section). A priori subgroup analysis should include at least the following three broad groups of CKD patients to explore the effect modifications by different CKD sub-populations: (1) late-stage CKD patients who are not on dialysis; (2) dialyzed CKD patients; and (3) patients with a kidney transplant. To rule out confounding factors, complete and detailed medical histories and patient characteristics should be evaluated to rule out other causes of anemia, such as malnutrition or abnormal hemoglobins (Thalassemia). Inflammation markers should also be measured and controlled for in statistical analyses. Because both ESA and iron treatment would confound the test results, a sufficient washout period (of at least 4 weeks) would

be necessary to stop ESA and iron treatment before ascertainment of the baseline test measurements.

All study patients should receive baseline test measurements including both newer and classical biomarkers of iron status, as well as any promising novel biomarkers of interest. Bone marrow aspiration should be conducted in all patients who received baseline test measurements on the same day or within few days after the baseline test measurements, and the results examined by two or more readers blinded to the results of other tests.

Resource Use, Size, and Duration

Cohort studies often require large samples, so it may be necessary to have multiple sites to enroll enough patients. Since it is difficult to obtain a meaningful estimate of true positive rate (i.e., true prevalence of absolute iron deficiency anemia), and sample size calculations are sensitive to the choice of true positive rate, we recommend performing a pilot study to obtain a meaningful estimate. The sample size calculation becomes more tedious to do when the same subject undergoes multiple tests, as the correlations must then be taken into account.[48]

Ability To Recruit

It is expected to be difficult to recruit sufficient number of patients, as bone marrow aspiration is painful and poses risks to patients. Using a washout period may not be feasible for some clinical settings. In this case, studies should recruit only CKD patients who did not receive ESA and iron treatment within 1 month before baseline test measurements.

Ethical Issues

Risks for bone marrow aspiration are significant concerns to some patients.

High-Priority Future Research Needs Topic 3

What is the diagnostic accuracy of newer markers of iron status as a replacement for, or an add-on to, classical lab markers in diagnosing functional iron deficiency?

Background

This FRN topic was also part of Key Question 2 in the original CER. Based on our post-hoc observation of this body of literature, we found that current studies often used a response to IV iron treatment as the reference standard for functional iron deficiency. The most commonly used definition for a response to IV iron treatment was an increase in hemoglobin concentration ≥1 g/dL after a period of IV iron treatment. 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.

The CER also found that, in adult hemodialysis CKD patients, CHr had a similar or better overall test accuracy compared with classical markers (TSAT or ferritin) to predict a response to IV iron treatment. Data suggested that CHr (with cutoff values of <27 or <28 pg) had a better sensitivity and specificity to predict a response to IV iron treatment than classical markers (TSAT <20 or ferritin <100 ng/mL). The %HYPO of red blood cells had a 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 also suggested that %HYPO (with cutoff values of >6% or >10%) had a better sensitivity and specificity to predict a response to IV iron

treatment than classical markers (TSAT <20% or ferritin <100 ng/mL). sTfR had a similar test performance compared with classical markers (TSAT or ferritin) to predict a response to IV iron treatment.

There is insufficient evidence regarding test performance of newer markers of iron status as an add-on to older markers, and test performance comparing newer with classical laboratory markers to predict a response to IV iron treatment in adult PD and nondialysis CKD patients, and in pediatric CKD patients.

Stakeholder Discussion

Much of the discussions with the stakeholder panel revolved around issues regarding the lack of generally accepted reference standard tests for determining iron deficiency in the setting of CKD (described earlier in FRN topic 1). In brief, stakeholders agreed that it is important to use the same reference standard for iron deficiency in future research; however, it appears that there is currently no consensus on whether "a response to IV iron treatment" should be considered as "functional iron deficiency" and that currently there is a lack of a well-accepted definition for a response to IV iron treatment. Stakeholders' comments were consistent with what we observed from the published research on this topic.

Proposed Study Designs

Unlike FRN topic 2 (diagnosing absolute deficiency), this topic cannot go forward until a reference standard for functional iron deficiency has been established. We therefore propose a *Consensus Expert Panel* to first determine the "best" (or most practical) reference standard for diagnosing functional iron deficiency in CKD patients (described in the FRN topic 1 in detail). Future research on this topic should use the same reference standard, in order to build on the body of evidence on this important research question and allow systematic review and meta-analysis to compare results across studies. If the reference standard for "functional iron deficiency" is defined by a response to ESA or iron treatment, a standardized protocol to ascertain the treatment response must be established, including the dosage and ESA and/or iron formulation and the followup duration to detect a response. There should be no change in patients' care to minimize the confounding, and the treatment response should be significantly different from biological variability in stable CKD patients.

Prospective cohort studies with consecutive CKD patients are the ideal study design for future research on this topic, as multiple tests can be compared all together (which may not be feasible in randomized trials), and potential biases, such as spectrum bias and review bias (i.e., interpretation of the index test or reference standard influenced by knowledge of the results of the other test), can be minimized. The agreed upon definition by the consensus expert panel should be used as the reference standard for diagnosing functional iron deficiency. *A priori* subgroup analysis should include at least the following three broad groups of CKD patients to explore the effect modifications by different CKD sub-populations: (1) late stage CKD patients who are not on dialysis; (2) dialyzed CKD patients; and (3) patients with a kidney transplant. To rule out confounding factors, complete and detailed medical histories and patient characteristics should be evaluated to rule out other causes of anemia, such as malnutrition or abnormal hemoglobins (Thalassemia). Inflammation markers should also be measured and controlled for in the statistical analyses. If the reference standard for functional iron deficiency is defined by a response to ESA or iron treatment, a sufficient washout period (of at least 4 weeks) is necessary to stop ESA and iron treatment before ascertainment of baseline test measurements.

All study patients should receive baseline test measurements including both newer and classical biomarkers of iron status, as well as any promising novel biomarkers of interest. Tests for assessing functional iron deficiency should be conducted in all patients who received baseline test measurements, and should be examined by two or more readers blinded to the results of other tests. If the reference standard for functional iron deficiency is defined by a response to ESA or iron treatment, newly conducted studies should adhere to the standardized protocol to ascertain the treatment response.

Resource Use, Size, and Duration

Cohort studies often require larger samples, so it may be necessary to have multiple sites to enroll enough patients. Since it is difficult to obtain a meaningful estimate of the true positive rate (i.e., the true prevalence of functional iron deficiency anemia) and sample size calculations are sensitive to the choice of true positive rate, we recommend performing a pilot study to obtain a meaningful estimate. The sample size calculation becomes more tedious when the same subject undergoes multiple tests, as the correlations must then be taken into account.[48]

Ability to Recruit

It should be relatively easy to recruit patients into a prospective cohort study. The major burden would be additional testing as per protocol. The reliance on observational data substantially reduces resource use and increases the feasibility of addressing this evidence gap. Using a washout period may not be feasible for some clinical settings. In this case, studies should recruit only CKD patients who did not receive ESA and iron treatment within 1 month before baseline test measurements.

Ethical Issues

There should be no ethical concerns for observational studies.

High-Priority Future Research Needs Topic 4

Which is the best marker to monitor the response to therapy and repletion status?

Background

The original CER did not directly address this FRN topic; however, this topic is closely related to FRN topic 3 (diagnosing functional iron deficiency). Since functional iron deficiency can be defined by a response to ESA or iron treatment, it is possible that this FRN topic is the same as FRN topic 3, although the purpose of the iron biomarkers is to monitor for iron deficiency anemia during treatment, not for diagnosing functional iron deficiency. It is important to note that studies included in Key Question 2 of the CER (diagnostic accuracy of newer markers of iron status as a replacement for, or an add-on to, classical lab markers in diagnosing functional iron deficiency) often did not explicitly state that the use of newer or classical lab markers was to diagnose functional iron deficiency. In fact, several studies that were included in the Key Question 2 of CER aimed to examine the test accuracy of iron biomarkers to "predict a response to treatment." Typically, "treat-to-target" or "treat-to-goal" levels of iron tests were used to monitor repletion status. It appears that there is a great deal of confusion/uncertainty in both the definition of functional iron deficiency and the purpose for iron status testing in existing research comparing the test performance of newer markers of iron status with that of classical markers.

Since monitoring tests are used repeatedly (usually in a given schedule) and more frequently than diagnostic tests, it would be sensible to use tests that have only a small diurnal variation so that real changes can be reflected accurately.

Stakeholder Discussion

The topic was nominated, based on stakeholder discussion and feedback, in the context of the "continuum of decision points in clinical management where biomarkers could be useful, but evidence is needed." In addition to diagnostic accuracy and safety, stakeholder discussion of the topic included additional characteristics of ideal biomarkers such as cost and availability. As with other high-priority FRN topics, establishing a reference standard was identified by stakeholders as necessary to assess test performance among relevant biomarkers.

Proposed Study Designs

Although the FRN topic poses a simple, noncomparative question, it is unclear which biomarkers and outcomes are most appropriate and relevant to study. Other potential biomarkers beyond the biomarkers of interest of the CER may be considered in future research. In the study design section below, we initially propose a *Consensus Expert Panel with Horizon Scan* to further specify the potentially limitless list of biomarkers and relevant outcomes (i.e., which outcomes determine which biomarker is the "best").

Consensus Expert Panel With Horizon Scan

Value of Study Design

We suggest that as a first step, prior to conducting new research studies to assess the best biomarker to monitor the response to therapy and repletion status, an expert panel should be convened to determine which markers appear most promising to explore in future research. Markers that can separate cell hemoglobinization from cell production would have a better chance of elucidating the underlying mechanism, and should the mechanism in question eventually become the treatment target, the anemia might as a consequence be more appropriately managed. Markers that have a specificity to the responses related to ESA or iron delivery would be optimal. The panel should assess which outcomes (e.g., specificity, sensitivity, adverse events) and test characteristics (e.g., test availability/accessibility, cost) are appropriate to consider in determining an ideal marker to monitor response to therapy and repletion status.

A *Horizon Scan*—a cursory summary of topic-relevant studies found using a systematic literature search—should be conducted to inform panel decisionmaking. The Horizon Scan could also be used to help assess the appropriateness of a CER on this topic. In addition to representatives from the current stakeholder categories, this panel should specifically include clinical chemists and hematologists. The main drawback to this approach would be that stakeholder opinion rather than evidence may well dictate the choice of biomarkers to study. However, given the resources that would be required to investigate all possible biomarkers, this tradeoff is likely to be reasonable.

Following identification of the most appropriate biomarkers by the expert panel, and determination of a reference standard, prospective cohort studies with consecutive CKD patients would pose the ideal study design for future research on this topic. We propose the same study design considerations as FRN topic 3, with the exception of study outcomes. Instead of test

performance outcomes, future research on this topic should assess clinical or patient-centered outcomes.

Resource Use, Size, and Duration

Initially, using an expert panel to identify promising biomarkers for future study would be more efficient than large studies assessing many different types/combinations of biomarkers, which would consume a large amount of resources. Expertise in literature search, assessment, and data extraction would be necessary to conduct the accompanying horizon scan.

Ability To Recruit

There should be little difficulty recruiting relevant stakeholders for the expert panel.

Ethical Issues

There should be no ethical issues in using an expert panel and horizon scan to investigate this evidence gap.

High-Priority Future Research Needs Topic 5

What is the best biomarker to monitor iron overload as an adverse event resulting from treatment of iron deficiency?

Background

Iron overload increases the risk for liver disease (cirrhosis, cancer), heart attack or heart failure, diabetes mellitus, osteoarthritis, osteoporosis, metabolic syndrome, hypothyroidism, hypogonadism, numerous undesirable symptoms, and in some cases premature death.

The original CER looked for evidence of the adverse effects or harms associated with the treatments guided by tests of iron status. However, it found insufficient evidence for this question.

Stakeholder Discussion

This topic was brought up during stakeholder discussion regarding the change in current dialysis practice, that is, decreasing ESA dosing while increasing iron treatment to reach the target hemoglobin level. The stakeholders stated that the increase in iron treatment will be accompanied by an increased risk of iron overload. In the opinions of the stakeholders, there are currently no reliable markers for monitoring iron overload.

Proposed Study Designs

Although the FRN topic poses a simple, non-comparative question, it is unclear which biomarkers are most appropriate and relevant for study. Other potential biomarkers beyond the biomarkers of interest in the CER may be considered in future research. In the study design section below, we initially propose a *Consensus Expert Panel with Horizon Scan* to further circumscribe the potentially limitless list of biomarkers.

Consensus Expert Panel With Horizon Scan

Value of Study Design

We suggest that as a first step, prior to conducting new research studies to assess the best biomarker to monitor iron overload as an adverse event, an expert panel should be convened to determine which markers appear most promising to explore in future research. Ideal markers would be significantly associated with clinically important outcomes. Horizon Scan could also be used to help assess the appropriateness of a systematic review on the associations between the markers of interest and clinically important outcomes.

Resource Use, Size, and Duration

Initially, using an expert panel to identify promising biomarkers for future study would be more efficient than large studies assessing many different types/combinations of biomarkers, which would consume a large amount of resources. Expertise in literature search, assessment, and data extraction would be necessary to conduct the accompanying horizon scan.

Ability To Recruit

There should be little difficulty recruiting relevant stakeholders for the expert panel.

Ethical Issues

There should be no ethical issues in using an expert panel and horizon scan to investigate this evidence gap.

Discussion

Challenges

The prioritization of topics for future research is a stakeholder-driven process. Our stakeholder panel represented a broad range of perspectives, from a well-informed patient advocate to clinical experts and policy makers. However, our stakeholder panel is unlikely to represent all perspectives because we did not use formal sampling methods to select our stakeholder members. The process of engagement though multiple teleconferences enabled us to get a well-rounded perspective, and we believe that the process of sharing the minutes of each discussion with all stakeholders enabled them to appreciate one another's viewpoint. We were able to obtain input from all members of the stakeholder panel, and the final selection of prioritized future research topics showed a clear preference for the top 5 listed.

Medical tests are used to help treatment decisions for one or more purposes in the disease prevention and management process: determining predisposition (who could develop the disease), screening (who has asymptomatic disease), diagnosis (who has symptomatic disease), staging (how advanced is the disease), prognosis (how progressive will the disease be), stratification (who will be a responder), efficacy (is the drug effective), monitoring (is the disease controlled), to measuring recurrence (relapse of disease). The ideal study design and sampling populations to investigate a medical test would vary depending on the purpose of the test, in order to maximize the internal and external validity of a study. For example, to evaluate the test performance of a diagnostic test, it is ideal to recruit consecutive patients with or without anemia and who have not received any treatment. In contrast, to evaluate the test performance of a monitoring test, it is ideal to recruit consecutive patients who have been diagnosed with and are receiving treatment for anemia.

One of the challenges in the current FRN project is that an iron status test can be used for the purposes of diagnosis or monitoring for the iron deficiency anemia in CKD patients, but the published studies don't often specify the purpose of the iron biomarkers evaluated. The same tests were being used for all three purposes without considering biases in interpretations of test results are highly likely to occur in this case. For example, treatment paradox occurs when treatment is started on the basis of the knowledge of the results of the index test, and the reference standard is applied after treatment has started. Another common bias arising in this situation is incorporation bias, which occurs when the index test is used to establish the final diagnosis. Thus, before future research on the top priority topics can be conducted, a truly independent reference standard of iron deficiency anemia must be established.

It should also be noted that patients' iron status can change spontaneously (due to changes in diet or unrelated metabolic and inflammatory conditions) or based on the treatment received. Therefore, it is important to control for these confounding factors in future research, which will require large sample sizes to reach sufficient statistical power.

Out-of-Scope Future Research Needs Topics

In addition to the 17 FRN topics eligible for prioritization, stakeholders identified 4 topics (Table 6) determined to be out-of-scope for the current FRN project. In general, topics were deemed out-of-scope if the state of the evidence was not rigorously assessed as part of the prior CER, such as studies of pediatric CKD patients, cost effectiveness and cost utility. Moreover, while these topics were areas of interest among stakeholders, some were not specific to anemia

biomarkers (e.g., Topic #1), did not address test effectiveness (e.g., Topic #2), and could be viewed as major research fields unto themselves.

Table 6. Topics outside scope of Future Research Needs project

#	Topic
1	What is the best method to translate the results of clinical tests to relevant management decisions?
2	What is the availability of newer biomarkers in routine clinical practice?
3	What is the impact of use of these tests on clinical practice?
4	What is the best marker to identify erythropoietin hyporesponsive patients?

Conclusions

Our CER and FRN stakeholder discussions identified the top future research gaps, and highlighted the great deal of confusion/uncertainty in using newer and/or classical laboratory biomarkers of iron status for the purposes of diagnosis or monitoring iron deficiency anemia in CKD patients. The chief factor causing this confusion and uncertainty is the lack of a well-accepted reference standard for iron deficiency anemia. The most effective first step would be to establish a common reference standard for iron deficiency anemia, considering two separate and distinct definitions: absolute and functional iron deficiency. The ideal reference standard should be independent of index tests and test-directed treatment to maximize the internal validity of study results.

References

1. KDOQI Clinical Practice Guidelines and Clinical Practice Recommendations for Anemia in Chronic Kidney Disease. Am J Kidney Dis. 2006 May;47(5 Suppl 3):S11-145. PMID: 16678659.
2. Anaemia Management in Chronic Kidney Disease - Rapid Update 2011. National Clinical Guideline Centre; 2011. www.ncbi.nlm.nih.gov/books/NBK65530/pdf/TOC.pdf.
3. 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. PMID: 4430129.
4. 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. PMID: 1239188.
5. Wish JB. Assessing iron status: beyond serum ferritin and transferrin saturation. Clin J Am Soc Nephrol 2006 Sep;1 Suppl 1:S4-S8. PMID: 17699374.
6. Guyatt GH, Oxman AD, Ali M, et al. Laboratory diagnosis of iron-deficiency anemia: an overview. J Gen Intern Med 1992 Mar;7(2):145-53. PMID: 1487761.
7. 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. PMID: 16999719.
8. 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. PMID: 9198047.
9. 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. PMID: 1739797.
10. Braun J. Erythrocyte zinc protoporphyrin. Kidney Int Suppl 1999 Mar;69:S57-S60. PMID: 10084287.
11. 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. PMID: 20189278.
12. 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. PMID: 15070565.
13. 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. PMID: 11229970.
14. Chung, M, Moorthy, D, Hadar, N, et al. Biomarkers for Assessing and Managing Iron Deficiency Anemia in Late-Stage Chronic Kidney Disease. Comparative Effectiveness Review No. 83. (Prepared by the Tufts Evidence-based Practice Center under Contract No. 290-2007-10055-I.) AHRQ Publication No. 12(13)-EHC140-EF. Rockville, MD: Agency for Healthcare Research and Quality. October 2012. www.effectivehealthcare.ahrq.gov/reports/final.cfm.
15. 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. PMID: 9543601.
16. 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. PMID: 9398126.
17. 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. PMID: 9589378.

18. 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. PMID: 17237481.

19. 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. PMID: 20472854.

20. 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. PMID: 12543894.

21. Cullen P, Soffker J, Hopfl M, et al. Hypochromic red cells and reticulocyte haemglobin content as markers of iron-deficient erythropoiesis in patients undergoing chronic haemodialysis. Nephrology Dialysis Transplantation 1999 Mar;14(3):659-65. PMID: 10193816.

22. 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. PMID: 10543327.

23. 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. PMID: 10401012.

24. 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. PMID: 20415234.

25. 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. PMID: 9211366.

26. 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. PMID: 15772926.

27. 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. PMID: 15025343.

28. Hukic F, Nuhbegovic S, Brkic S, et al. Biochemical markers of iron status in hemodialysis patients. Medicinski Arhiv 2010;64(4):219-22. PMID: 21246919.

29. 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. PMID: 18190467.

30. 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. PMID: 14682204.

31. 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. PMID: 14586744.

32. 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. PMID: 9398141.

33. 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. PMID: 19624802.

34. 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. PMID: 11427634.

35. 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. PMID: 20538788.

36. 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. PMID: 12608554.

37. 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. PMID: 16316362.

38. 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. PMID: 7872320.

39. 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. PMID: 12631092.

40. Silva J, Andrade S, Ventura H, et al. Iron supplementation in haemodialysis--practical clinical guidelines. Nephrology Dialysis Transplantation 1998 Oct;13(10):2572-77. PMID: 9794562.

41. 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. PMID: 11737617.

42. 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. PMID: 17396118.

43. 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. PMID: 9328369.

44. Concannon TW, Meissner P, Grunbaum JA, et al. A new taxonomy for stakeholder engagement in patient-centered outcomes research. J Gen Intern Med 2012 Aug;27(8):985-91. PMID: 22528615.

45. Fishbane S. Iron management in nondialysis-dependent CKD. Am J Kidney Dis 2007 Jun;49(6):736-43. PMID: 17533016.

46. Hung SC, Tarng DC. Bone marrow iron in CKD: correlation with functional iron deficiency. Am J Kidney Dis 2010 Apr;55(4):617-21. PMID: 20338462.

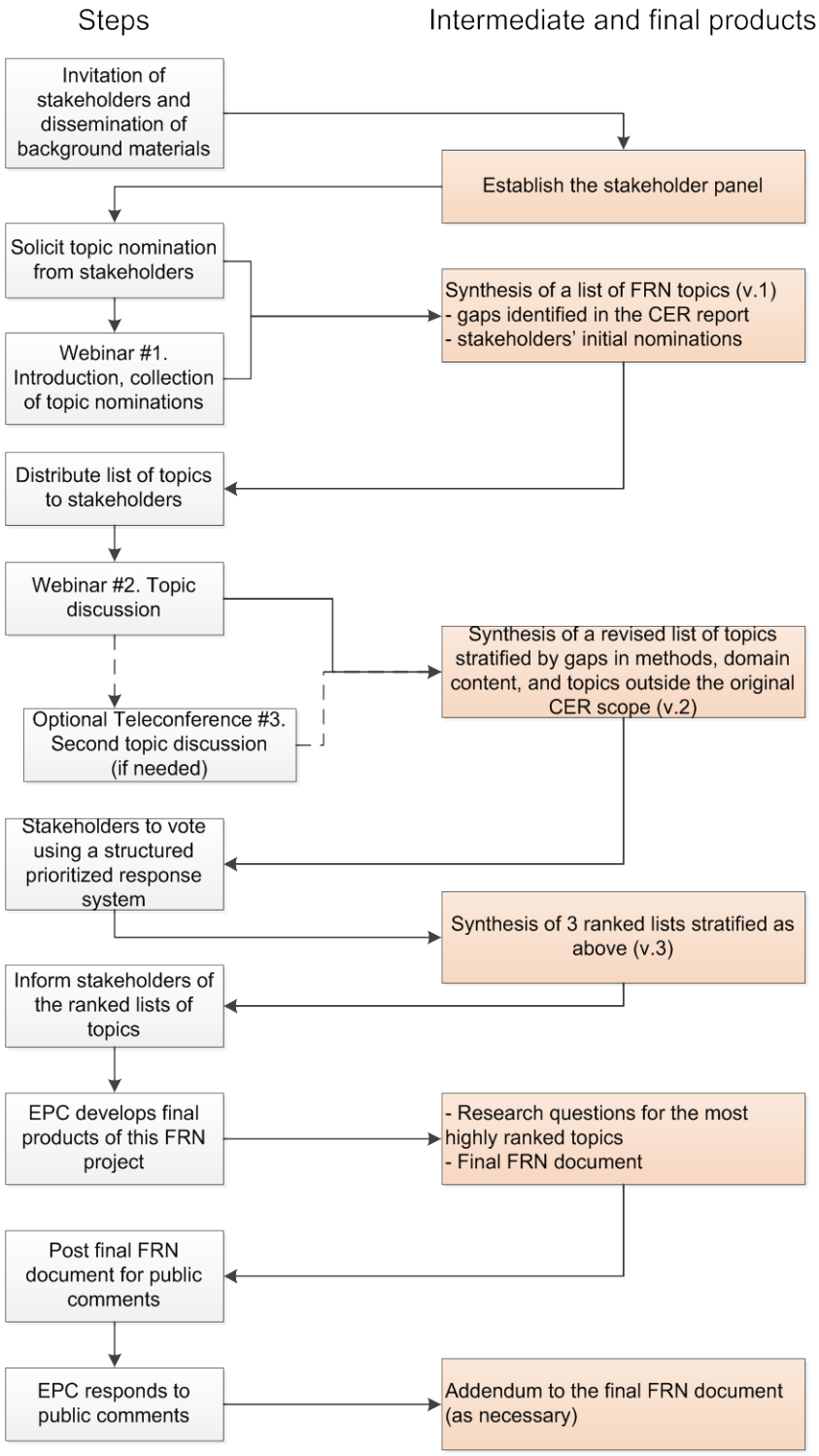
47. Bain BJ. Bone marrow biopsy morbidity: review of 2003. J Clin Pathol 2005 Apr;58(4):406-08. PMID: 15790706.

48. Liu A, Schisterman EF, Mazumdar M, et al. Power and sample size calculation of comparative diagnostic accuracy studies with multiple correlated test results. Biom J 2005 Apr;47(2):140-50. PMID: 16389911.

Abbreviations

AHRQ	Agency for Health Research and Quality
CER	Comparative Effectiveness Review
CHr	Hemoglobin content of reticulocytes
CKD	Chronic Kidney Disease
EPC	Evidence-based Practice Center
ESA	Erythropoiesis-stimulating agent
FRN	Future Research Needs
Hb	Hemoglobin
HD	Hemodialysis
IV	Intravenous
FRN	Future research needs
%HYPO	Percentage of hypochromic erythrocytes
PD	Peritoneal dialysis
PICOD	Population, Intervention, Comparator, Outcomes, and study Design
RBC	Red blood cell
RCT	Randomized controlled trial
RetHe	Reticulocyte hemoglobin equivalent
sTfR	Soluble transferrin receptor
TSAT	Transferrin saturation
ZPP	Zinc protoporphyrin

Appendix A. Future Research Needs Process



Abbreviations: CER = Comparative Effectiveness Review; EPC = Evidence-based Practice Center; FRN = Future Research Needs

Appendix B. Effective Health Care Program Selection Criteria

Appropriateness:

- Represents a health care drug, intervention, device, technology or health care system/setting available (or soon to be available) in the United States.
- Relevant to 1013 enrollees (Medicare, Medicaid, S-CHIP, other federal health care programs.
- Represents one of the priority conditions designated by the U.S. Department of Health and Human Services (HHS).

Importance:

- Represents a significant disease burden, large proportion or priority population.
- Is of high public interest; affects health care decision-making, outcomes, or costs for a large proportion of the U.S. population or for a priority population in particular.
- Was nominated/strongly supported by one or more stakeholder groups.
- Represents important uncertainty for decisionmakers.
- Incorporates issues around both clinical benefits and potential clinical harms.
- Represents important variation in clinical care, or controversy in what constitutes appropriate clinical care.
- Represent high costs to consumers, patients, health care systems or payers; due to common use, high unit costs, or high associated costs.

Desirability of New Research/Duplication:

- Would not be redundant (i.e., the proposed topic is not already covered by available or soon-to-be available high quality systematic review by AHRQ or others).

Feasibility:

- Effectively uses existing research and knowledge by considering adequacy of research for conducting a systematic review, and newly available evidence.

Potential Impact:

- Potential for significant health impact, significant economic impact, potential change, potential risk from inaction, addressing inequities and vulnerable populations, and/or addressing a topic with clear implications for resolving important dilemmas in health and health care decisions made by one or more stakeholder groups.

Appendix C. Yield of Ongoing Studies

Search on ClinicalTrials.gov for ongoing research found eight studies with a search of: (iron deficiency OR response to iron) AND ((chronic kidney disease) OR dialysis) | Open Studies | Exclude Unknown | Adult, Senior

NCT ID	Title	Recruitment	Interventions	Enrollment
NCT01227616	A Phase IV Trial of Ferumoxytol for the Episodic Treatment of Iron Deficiency Anemia in Patients With Chronic Kidney Disease	Not yet recruiting	Drug: Ferumoxytol	300
NCT01102413	Iron Isomaltoside 1000 (Monofer®) in Non-Dialysis Dependent Chronic Kidney Disease and With Renal-Related Anaemia	Recruiting	Drug: Monofer | Drug: Iron Sulphate	350
NCT00830037	A Clinical Trial of Oral Versus IV Iron in Patients With Chronic Kidney Disease	Recruiting	Drug: IV Iron | Drug: Ferrous Sulfate	200
NCT01509690	Impact of a Multidisciplinary Intensive Management Clinic on Outcomes in Multi-Ethnic Asian Incident Hemodialysis Patients	Recruiting	Other: Multidisciplinary, intensive and collaborative care	150
NCT01587924	4 Week Switch Study in Hemodialysis-dependent Subjects With Anemia Associated With Chronic Kidney Disease	Recruiting	Drug: GSK1278863 | Drug: rhEPO	68
NCT01414075	Study of FG-4592 to Correct Anemia in New Dialysis Patients	Recruiting	Drug: FG-4592	60
NCT01532349	Vitamin D as a Modifier of Serum Hepcidin in Children With Chronic Kidney Disease	Not yet recruiting	Drug: Cholecalciferol	28
NCT01138241	Tenofovir Renal Toxicity and Glomerular Filtration Rate (GFR) Validation	Recruiting	Observational study	700