



Effective Health Care Program

Serum Free Light Chain Analysis for the Diagnosis, Management, and Prognosis of Plasma Cell Dyscrasias

Executive Summary

Background

Plasma cell dyscrasias (PCDs) are a group of neoplastic disorders characterized by the uninhibited expansion of a monoclonal population of malignant plasma cells.¹ Multiple myeloma (MM) is the most common malignant plasma cell tumor, accounting for about 1 percent of all cancer types,¹ and the second most common hematologic malignancy in the United States. With an age-adjusted incidence rate of 5.5 cases per 100,000 population,² an estimated 19,900 new diagnoses and 10,790 deaths due to myeloma occurred in 2007, according to the American Cancer Society.³ Although the median survival has improved to 5 years with current standards of treatment,⁴ the annual costs of modern therapies can range from \$50,000 to \$125,000 per patient.^{5,6}

In PCDs, each abnormally expanded clone of malignant plasma cells produces an excess of either intact immunoglobulin or free light chains (FLCs) of a single type; either type of excess molecule is called a monoclonal protein (M protein) or paraprotein. Measurement of M proteins (either complete immunoglobulins or FLCs) is integral to diagnosing PCDs, monitoring disease response to therapy and adjusting treatment, and determining disease progression or relapse.

Effective Health Care Program

The Effective Health Care Program was initiated in 2005 to provide valid evidence about the comparative effectiveness of different medical interventions. The object is to help consumers, health care providers, and others in making informed choices among treatment alternatives. Through its Comparative Effectiveness Reviews, the program supports systematic appraisals of existing scientific evidence regarding treatments for high-priority health conditions. It also promotes and generates new scientific evidence by identifying gaps in existing scientific evidence and supporting new research. The program puts special emphasis on translating findings into a variety of useful formats for different stakeholders, including consumers.

The full report and this summary are available at www.effectivehealthcare.ahrq.gov/reports/final.cfm.

The serum FLC (SFLC) assay (i.e., the Freelite[®] assay, The Binding Site Ltd., Birmingham, United Kingdom) was introduced in 2001 to measure the FLC component in serum.⁷ The assay works



by recognizing an epitope that is detectable only on light chains that are not bound to the heavy chain of the immunoglobulin molecule—the FLCs—in the serum. This is the sole SFLC assay the U.S. Food and Drug Administration (FDA) has approved for use in the United States.

The International Myeloma Working Group (IMWG) considers the SFLC assay to be an adjunct to traditional tests.⁸ The assay could allow for quantitative monitoring of response and remission after treatment and provide prognostic information,^{9,10} potentially reducing the need for frequent bone marrow biopsies. Quantifying plasma cells in the marrow is needed for monitoring progression of monoclonal gammopathy of undetermined significance (MGUS) to MM and for defining and stringent monitoring of disease remission.⁸ The SFLC assay has the potential for use in conjunction with serum protein electrophoresis (SPEP) and serum immunofixation electrophoresis (SIFE) to replace urine tests that require 24-hour collection (i.e., urine protein electrophoresis [UPEP] and urine immunofixation electrophoresis [UIFE]), which could simplify diagnosis and disease monitoring.^{8,11} The SFLC assay may also be the only means of detecting a disease marker in some disease settings: (1) nonsecretory MM (NSMM), in which SFLCs are often the only marker of the disease;¹² (2) AL amyloidosis (in which amyloid [A] proteins derived from immunoglobulin light chains [L] are deposited in tissue), in which low M protein concentrations may not be detected by means of conventional techniques; and (3) light chain MM (LCMM), in which the M protein consists only of FLCs.⁸ Thus, in addition to detecting a wider spectrum of PCDs than traditional tests, the assay may help detect earlier stages of the disease, and because of the short half-life of SFLCs (2 to 6 hours, vs. 21 days for complete immunoglobulins¹³), the assay may also help detect relapses and treatment failures earlier than by reliance on M protein concentrations alone.¹⁰

Although the SFLC assay has been in use for a decade, how best to incorporate it into practice remains unclear.¹⁴ Given the assay's biological validity and ease of use compared with cumbersome urine collections, clinicians seem to have widely adopted the test as an adjunct to the panel they use to diagnose PCDs. Its use is also being evaluated in patient management. PCDs are a heterogeneous group of disorders that require a panel of tests for accurate diagnosis. Different tests will perform differently across the variety of disease subgroups and across different disease settings, and their results need to be evaluated with this in mind. Ascertainment of the

assay's comparative effectiveness will allow for its use to be refined and recommendations for its use optimized. This comparative effectiveness review (CER) addresses these aspects, noting that evaluations of the SFLC assay's clinical utility should allow for different clinical settings and phases of disease as well as different disease populations.

Objectives

The aim of this CER is to evaluate the present body of evidence addressing the relative effectiveness of the SFLC assay as compared with traditional tests for the diagnosis, management, and prognosis of PCDs. We sought to answer a set of questions focusing on the SFLC assay versus traditional testing in very specific clinical settings to focus on comparative effectiveness. Our goals were to evaluate the SFLC assay as an add-on test in diagnostic settings and to compare it with existing tests in other settings such as for disease monitoring and prognosis. Panels of Key Informants and Technical Experts, who helped identify the important areas for evidence review (as discussed in the Methods section), vetted these questions. To address these areas in an unbiased way that would permit summary of the relevant data, studies had to meet a specific, predefined set of criteria related to population, intervention (diagnostic test/disease monitoring), comparator, and outcome.

This CER evaluates the SFLC assay as a diagnostic and prognostic tool adjunctive to the standard diagnostic tests for various PCDs. It addresses five Key Questions (KQs) that pertain to the (1) diagnosis of PCDs, (2) prognosis (i.e., progression from MGUS to MM and overall and disease-free survival in patients with a malignant PCD), (3) change in treatment decisions, (4) assessment of response to treatment, and (5) reduction of the need for other diagnostic tests (e.g., bone marrow biopsy).

Key Questions

KQ1. Does adding the SFLC assay and the kappa/lambda ratio to traditional testing (serum/urine electrophoresis or IFE), compared with traditional testing alone, improve the diagnostic accuracy for PCDs (MGUS, MM, NSMM, or AL amyloidosis) in undiagnosed patients suspected of having a PCD?

KQ2. As compared with traditional tests, how well does the SFLC assay independently predict progression to MM in patients with MGUS?

KQ3. In patients with an existing diagnosis of PCD (MM, NSMM, or AL amyloidosis), does the use of the SFLC

assay result in different treatment decisions as compared with traditional tests?

- Does the use of the SFLC assay affect the management of patients by allowing for earlier institution of specific therapies?
- Does the use of the SFLC assay influence the duration of treatment?
- Does the use of the SFLC assay influence the type of treatment (e.g., radiation therapy)?

KQ4. In patients with an existing diagnosis of PCD (MM, NSMM, or AL amyloidosis), is the SFLC assay better than traditional tests in indicating how the patient responds to treatment and of outcomes (overall survival, disease-free survival, remission, light chain escape, and quality of life)?

KQ5. In patients with an existing diagnosis of PCD (MM, NSMM, or AL amyloidosis), does the use of the SFLC assay reduce the need for other diagnostic tests (e.g., bone marrow biopsy)?

Analytic Framework

To guide the development of the KQs, we generated an analytic framework (Figure A) that maps the specific linkages associating the population (patients with PCD symptoms) and subgroups of interest (e.g., individual PCDs or clinical settings) with the additional tests (i.e., SFLC assay in addition to traditional testing) and the comparator (traditional tests alone), as well as the outcomes of interest (diagnostic accuracy, prognosis, disease management, reduction of other diagnostic tests, and response to treatment). This framework depicts the chain of logic that evidence must support to link the use of the SFLC assay to improved health outcomes.

Methods

Input From Stakeholders

During a topic refinement phase, the initial questions were refined with input from a panel of Key Informants. Key Informants included representatives from the American Association for Clinical Chemistry; experts in renal amyloidosis, clinical chemistry, geriatrics, and general internal medicine; patient advocates; and representatives from the Centers for Medicare and Medicaid Services and a nationwide health insurance company. After a public review of the proposed KQs, we convened a Technical Expert Panel (TEP) consisting of experts (some of whom were Key Informants) in MM and/or AL amyloidosis,

clinical chemistry, and general medicine), who served in an advisory capacity to help refine KQs, identify important issues, and define parameters for the review of evidence. Discussions among the relevant EPC staff, Task Order Officer, and Key Informants, and subsequently, the TEP occurred during a series of teleconferences and via email. In addition, input from the TEP was sought during compilation of the report, when questions arose about the scope of the review.

Data Sources and Selection

The evidence presented was obtained through a systematic review of the published scientific literature, using established methodologies as outlined in AHRQ's Methods Guide for Effectiveness and Comparative Effectiveness Reviews¹⁵ and Methods Guide for Medical Test Reviews.¹⁶

We conducted literature searches of studies in MEDLINE[®], the Cochrane Central Register of Controlled Trials, and the Cochrane Database of Systematic Reviews. All English-language studies with adult human participants were screened to identify articles relevant to each KQ. The reference lists of related systematic reviews as well as selected narrative reviews and primary articles were also reviewed for relevant studies. Our search included variations of the terms “immunoglobulin light chain,” “monoclonal light chain,” “serum free light chain,” and “Bence Jones protein.”

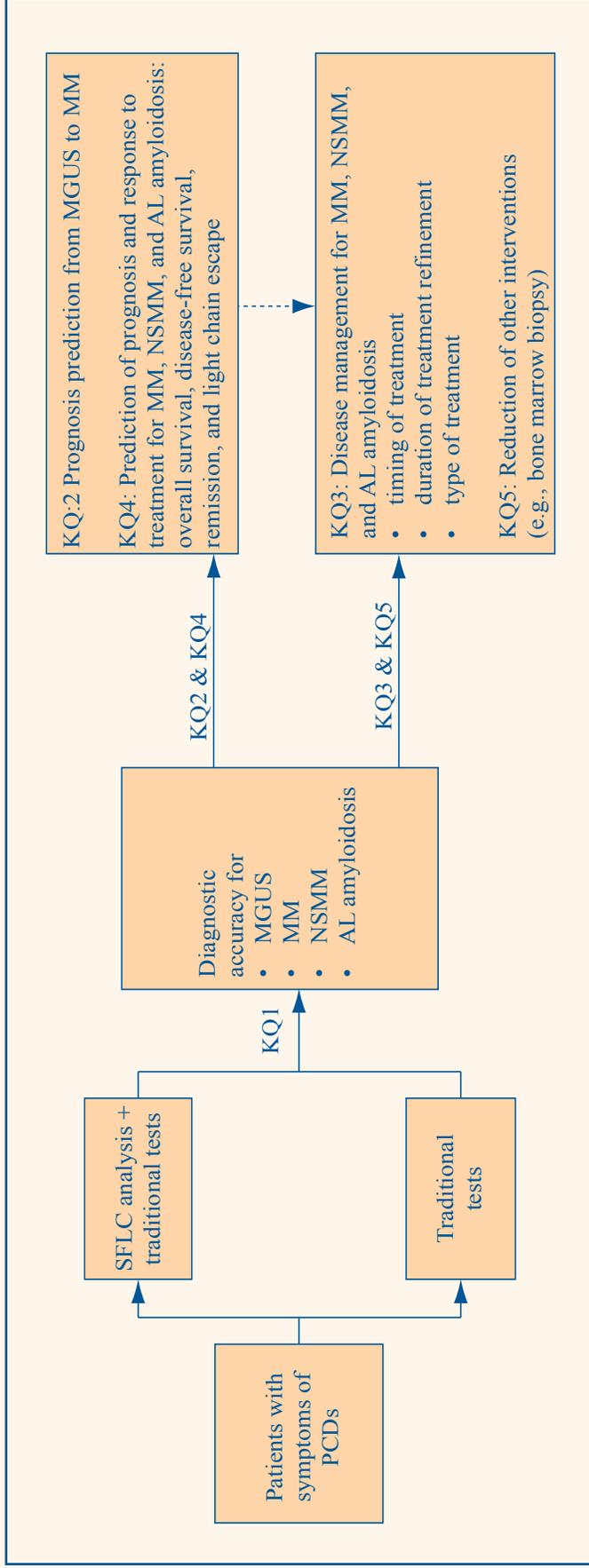
We included published, peer-reviewed articles only. Two team members independently screened the abstracts to ascertain their eligibility. Relevant abstracts were retrieved in full text for detailed evaluation.

Below are the eligibility criteria for study inclusion. No restrictions were placed on the particular type of study designs eligible in each of the KQs, but an overarching requirement was that the study be designed to address the comparative effectiveness of the SFLC assay—that is, compare the assay with (predefined) traditional tests: SPEP, UPEP, SIFE, and UIFE, and other tests in common use in a diagnostic panel for PCDs (e.g., bone marrow, skeletal survey).

The eligibility criteria for study populations included the following:

- KQ1: studies that addressed adults (≥ 18 years of age) who had not been diagnosed with a PCD, with or without kidney failure, but who were suspected of having PCD
- KQ2: studies of patients with MGUS

Figure A. Analytic framework for SFLC analysis for the diagnosis, management, and prognosis of PCDs



AL amyloidosis = systemic amyloidosis in which amyloid [A] proteins derived from immunoglobulin light chains [L] are deposited in tissue; KQ = Key Question; MGUS = monoclonal gammopathy of undetermined significance; MM = multiple myeloma; NSMM = nonsecretory multiple myeloma; PCD = plasma cell dyscrasia; SFLC=serum free light chain

- KQ3–5: studies of patients with an existing diagnosis of PCD (MM, NSMM, or AL amyloidosis), with or without disease measurable by means of traditional testing

For interventions (diagnostic tests/disease monitoring), eligible studies were those involving the SFLC assay as well as the FLC kappa/lambda ratio. For comparators, eligible studies were those involving any kind of traditional testing (i.e., SPEP, UPEP, SIFE, or UIFE; sizing and typing of serum M protein; bone marrow biopsy; or detection of skeletal lesions).

For outcomes, eligible studies were those with the following data:

- KQ1: measures of diagnostic accuracy, such as sensitivity, specificity, predictive values, likelihood ratios, or area under the receiver-operating-characteristics curve
- KQ2: progression to MM
- KQ3: timing, duration, and type of treatment
- KQ4: overall survival, disease-free survival, response to treatment or remission (categorized as partial, complete, or stringent complete on the basis of treatment-induced decline in M protein or FLC concentrations^{8,17}), light chain escape, or quality of life
- KQ5: clinic visits, bone marrow biopsies, or skeletal surveys

Studies could have any length of followup^{8,17} or any setting (primary or specialty care, in-facility or home, inpatient or outpatient).

Data Extraction and Risk-of-Bias Assessment

We extracted study data into customized forms. Together with information on study design, patient and test characteristics, outcome definitions, and study results, we rated the risk of bias (methodological quality) of each study from A (highest quality, least likely to have significant bias), to C (lowest quality, most likely to have significant bias).

In the present report, the majority of studies were related to testing diagnostic performance and predicting outcomes; therefore, we adapted criteria from formal quality-assessment schemes for diagnostic accuracy studies—STAndards for the Reporting of Diagnostic accuracy studies (STARD, www.stard-statement.org)—and observational epidemiologic studies—STrengthening

the Reporting of OBServational studies in Epidemiology (STROBE, www.strobe-statement.org).

We followed the Methods Guide to grade the strength of the body of evidence (mostly a measure of risk of bias) for each KQ, with modifications, on the basis of our level of confidence that the evidence reflected the true effect for the major comparisons of interest. The strength of evidence was defined as low, medium, high, or insufficient on the basis of the number of studies; consistency across the studies; and precision of the findings. We required at least two quality A studies for a high rating, a moderate rating can reflect fewer than two quality A studies, a low rating involves quality B or quality C studies, and an insufficient rating indicates that evidence is either unavailable or does not permit a conclusion.

Data Synthesis and Analysis

We summarized all included studies in narrative form and in summary tables. We included diagnostic performance parameters, risk estimates, and their 95% confidence intervals (CI) and p-values, where applicable. We provided mainly descriptive analyses¹⁸ and undertook a qualitative synthesis of studies that addressed the predictive role of the SFLC assay. We did not conduct any meta-analyses of the studies, as there was marked heterogeneity in their designs, populations, and comparisons.

Results

The literature search yielded 3,036 citations, of which 2,711 were excluded at the abstract level because FLCs were not studied; the diagnosis was not relevant to the KQs; or the report was a narrative review, conference proceeding, single case study, or animal study. The remaining 325 articles were retrieved for full-text review, upon which 310 were excluded, because they did not address the relevant test, population, diagnosis, or comparison of interest or because they were narrative reviews, commentaries, single case studies, or letters to the editor without primary data. Most of the exclusions were studies that did not meet all the predefined criteria and/or did not provide data comparing the performance of the SFLC assay with the predefined traditional tests (serum or urine tests [SPEP, UPEP, SIFE, or UIFE], bone marrow evaluation, or skeletal survey). A total of 15 studies that both were comparative and met all the CER eligibility criteria were included.

KQ1: Does adding the SFLC assay and the kappa/lambda ratio to traditional testing (serum/urine electrophoresis or IFE), compared with traditional testing alone, improve the diagnostic accuracy for PCDs (MGUS, MM, NSMM, or AL amyloidosis) in undiagnosed patients suspected of having a PCD?

Three studies evaluated the SFLC assay in combination with traditional tests in undiagnosed patients suspected of having a PCD. Reviewers gave all three studies a B quality rating because of their retrospective design and because they did not provide formal statistical comparisons and confidence intervals. All three studies compared test results with the diagnosis of disease verified by medical records on the basis of a panel of criteria. The addition of the SFLC assay to traditional tests in a diagnostic panel increased the sensitivity of the assay for detection of PCDs in all three studies (from 0.64–0.87 to 0.96–1.00 for SPEP and to 0.92–0.94 for SIFE); however, the statistical significance of this increase was not addressed in any of the studies and the effect on specificity was inconsistent. The studies were heterogeneous with regard to design and comparator, such that meta-analysis could not be performed for quantitative data synthesis. In the light of these results, we rated the strength of evidence to evaluate the effect of adding SFLC testing to traditional testing on diagnostic performance as insufficient.

KQ2: As compared with traditional tests, how well does the SFLC assay independently predict progression to MM in patients with MGUS?

No studies compared the use of the SFLC assay with traditional tests to determine whether the use of the SFLC assay predicts progression from MGUS to MM. Therefore, we rated the strength of evidence as insufficient for this question.

KQ3: In patients with an existing diagnosis of PCD (MM, NSMM, or AL amyloidosis), does the use of the SFLC assay result in different treatment decisions as compared with traditional tests?

No studies compared the use of the SFLC assay with traditional tests to determine whether treatment decisions were different with regard to timing, duration, or type of treatment. Therefore, we rated the strength of evidence as insufficient for this question.

KQ4: In patients with an existing diagnosis of PCD (MM, NSMM, or AL amyloidosis), is the SFLC assay better than traditional tests in indicating how the patient responds to treatment and of outcomes (overall survival, disease-free survival, remission, light chain escape, and quality of life)?

Eleven studies evaluated the SFLC assay and traditional testing in parallel and examined their relationship to clinical outcomes in PCDs. No direct comparisons between the SFLC assay and traditional tests were performed. Three studies were conducted with patients who had AL amyloidosis and eight with patients who had MM. Three studies reported industry-associated funding or authorship. Nine studies were retrospective, and one was prospective; the remaining study lacked enough detail to determine the study design. Followup times varied from 3 months to 13 years, with sample sizes of 40 to 399 patients. Among studies reporting patient characteristics, the median age ranged from 54 to 72 years, and the study populations were 44 to 65 percent male.

Patients With AL Amyloidosis

Three retrospective studies examined the use of the SFLC assay with patients who had AL amyloidosis and reported the use of SFLC assay in evaluating treatment response and predicting prognosis. These studies measured SFLC responses and paraprotein responses to treatment with traditional testing (electrophoresis or IFE) and examined their relationship to outcomes. Paraprotein reduction was usually reported as part of a “hematologically complete” response.

Although the three studies reported the SFLC assay may aid in assessing treatment response and monitoring outcomes in AL amyloidosis patients, no direct comparisons with traditional tests (electrophoresis or IFE) were performed. We rated all three studies as quality C, because of limitations in study design, including selection/spectrum bias as well as (in one study) small sample size. Overall, because of a lack of direct comparisons and poor study quality, current evidence on the effectiveness of the SFLC assay as compared with traditional tests for assessing treatment response and outcome is inconclusive. We therefore rated the strength of evidence underlying this comparison as insufficient.

Patients With MM

Eight studies enrolled patients with MM and compared the use of the SFLC assay and other traditional tests in

evaluating treatment response and predicting prognosis. Six were retrospective analyses of cohorts; one was prospective; and the other study had an unspecified design. We graded the study quality as B in three of the eight studies because of their retrospective designs without adjustments for potential confounders and as C in the other five studies because of their small sample sizes, limited information about study design, and/or potential selection bias. None of the three B-quality studies performed direct statistical comparisons of relative strength of prediction. The three outcome categories covered in the studies are discussed in the next paragraphs.

Assessment and Prediction of Treatment Response.

Four studies addressed the use of SFLC assay in the assessment of treatment response, and one study addressed the prediction of treatment response. The traditional test comparators used to assess treatment response (in parallel with the SFLC assay) differed in each study (i.e., SPEP, UPEP, total kappa/lambda ratio measured by nephelometry, bone marrow evaluation with immunophenotyping, or standard response criteria [e.g., from IMWG]). The heterogeneity in the tests and study designs across the five studies precluded any clear conclusion regarding assessment and prediction of treatment response.

Of the four studies that used SFLC assay test results to assess treatment response, one study, of C quality, found that 22 of 102 patients had discordant findings regarding achievement of a treatment response after induction therapy, defined according to the SFLC ratio and the immunophenotypic response. Another study, of B quality, found that after 2 months of therapy, treatment response was achieved by 23 percent of patients using the paraprotein definition, compared with 62 percent using the SFLC definition. In a smaller C-quality study, the majority of patients achieved treatment response as defined by both M protein criteria and SFLC criteria at the same time; in the minority of patients, however, the SFLC response occurred earlier than M protein response. A fourth study reported an abnormal SFLC ratio before relapse and a positive IFE test in nine patients, but it was rated of C quality because of limited information about study design, SFLC response definitions, and results. The poor quality and heterogeneity in the comparator used, as well as a lack of data for further synthesis, made it difficult to draw conclusions regarding the comparison between SFLC and traditional test comparators in the assessment and prediction of treatment response.

Only one study, of C quality, reported data on prediction of treatment response, so conclusions are premature until more studies are performed. This study applied an SFLC and M protein–based model to predict response to VDD (bortezomib, pegylated liposomal doxorubicin, and dexamethasone) used to treat newly diagnosed, histologically confirmed MM. The model predicted that either (1) a 90 percent or greater reduction of serum M protein level or involved SFLC level or (2) normalization of the SFLC ratio predicted a very good partial response (VGPR) or better response, with 92 percent sensitivity and 93 percent specificity after two cycles of treatment with VDD. Sensitivity increased to 96 percent after three cycles of VDD treatment. Neither the rate of decline in M protein nor the involved SFLC concentration independently predicted VGPR at the end of six cycles of VDD (at 90 percent sensitivity and specificity). When the involved SFLC was replaced by urine M protein in the predictive model, the sensitivity, specificity, and predictive value were all less than 90 percent.

Relationship Between Baseline SFLC Measurements and Survival.

For this outcome, the small number of included studies and the heterogeneity in the test comparator precluded a clear conclusion regarding the SFLC assay and prediction of survival. Two studies examined the relationship of baseline SFLC concentrations and survival. One, of B quality, evaluated the predictive ability between the SFLC assay and traditional testing (baseline concentrations of serum and urine M protein). The overall and event-free survival rates were significantly lower among patients with higher (> 75 mg/dL) versus those with lower (≤ 75 mg/dL) SFLC concentrations (overall survival: $p=0.016$, event-free survival: $p=0.08$), but neither serum nor urine M protein concentrations were predictive of survival. The other study, of C quality, compared the SFLC ratio with clinical stage (per Durie–Salmon staging and the International Staging System [ISS]¹⁹); both were found to be independent predictors (both $p<0.001$), and an abnormal SFLC ratio was also significantly associated with 3- and 5-year disease-specific survival rates ($p=0.0001$).

Relationship Between Post-Treatment SFLC Measurements and Survival.

Three studies examined the relationship between post-treatment SFLC measurements and survival. Because of the differences in comparators analyzed and heterogeneity in data analyses, we could not draw any conclusions. One study, of C quality, analyzed the SFLC ratios after induction therapy and

reported that after stratification of patients on the basis of immunofixation status, the 3-year progression-free survival rate, time to progression, and overall survival did not differ between patients with normal and abnormal SFLC ratios post-treatment.

A second study, of B quality, analyzed immunofixation results and SFLC ratios after stem-cell transplantation. Overall and event-free survival did not differ between patients with and those without a normal SFLC ratio or between patients with and those without a normal SIFE test. However, a normal SFLC ratio at 3 months post treatment was significantly associated with longer event-free survival ($p=0.02$) but not with overall survival ($p=NS$).

In the third study, also of B quality, patients with a percentage reduction in SFLC level in the top tertile after transplantation had nearly twice the risk of death than patients with a smaller reduction. However, there was no significant relationship between the tertiles of percentage reductions in serum and urine M protein values and overall or event-free survival.

Summary for MM. Eight studies reported on the use of the SFLC assay and traditional tests in measuring treatment response and predicting prognosis in patients with MM. However, none of the studies formally compared the predictive capability of the SFLC assay with that of traditional tests. Most were retrospective cohort studies, and only three were of quality B (with the rest being quality C). The studies were heterogeneous with respect to population, intervention (diagnostic test/disease monitoring), and comparator, as well as degree of adjustment for confounders. Taken together, these factors limit the conclusions that can be drawn about the definitive use of the SFLC assay in prognosis prediction, and we rated the strength of evidence as insufficient for comparisons with traditional testing in patients with MM.

KQ5: In patients with an existing diagnosis of PCD (MM, NSMM, or AL amyloidosis), does the use of the SFLC assay reduce the need for other diagnostic tests (e.g., bone marrow biopsy)?

One C-quality retrospective study assessed the need for bone marrow examination, with the SFLC assay used to define the completeness of response to treatment. As currently defined in the European Group for Blood and Marrow Transplantation and IMWG uniform response criteria, a complete response in a patient with MM requires a bone marrow examination showing less than 5 percent plasma cells, in addition to negative SIFE and UIFE

results; the addition of normalization of the SFLC ratio defines stringently complete remission.^{17,20} This study reported on 29 patients with MM and negative SIFE and UIFE tests who also had a bone marrow aspirate or biopsy as well as data on the SFLC ratio. The authors concluded it was not possible to eliminate the need for bone marrow testing to evaluate response. Because of the preliminary nature of the data, we rated the strength of evidence as insufficient for addressing this question.

Discussion

Since its introduction in 2001, the SFLC assay has been used for screening and diagnosing PCDs, disease prognostication, and quantitative monitoring of treatment course. In the present review, we assessed the comparative effectiveness of the SFLC assay as an adjunct to traditional tests such as SPEP and SIFE for the diagnosis of PCD in populations suspected of having the disease. We also ascertained the assay's ability, relative to traditional testing, to predict progression of MGUS to MM, prognosticate for malignant PCDs, determine treatment decisions, and eliminate the need for other diagnostic tests. Table A summarizes the main findings addressing the five KQs of this CER.

Our results reveal a paucity of evidence to clarify the comparative effectiveness of the role of the SFLC assay for the diagnosis, management, and prognosis of PCDs. We identified only 15 studies in our literature search, those having met all the inclusion criteria to address the KQs. Across the studies, there was considerable clinical heterogeneity with regard to variation in type or stage of disease and phase of treatment. Moreover, although in the 15 studies the SFLC assay and traditional testing were commonly conducted in parallel, they were not formally compared. That is, the studies did not include statistical comparisons of predictive value by comparing areas under a receiver-operating-characteristic curve or strength of association within models using measures such as likelihood ratios. The study heterogeneity observed, with variations in study design and population as well as inconsistency in the comparisons being made, may also reflect uncertainties associated with the role of the assay in research and clinical practice. Finally, the majority of the studies were of poor quality. All these factors limited the validity of the studies and the conclusions that could be drawn from them. The insufficient evidence to answer those questions indicates areas needing targeted research in the future. We also found that much of the available research did not meet stringent reporting standards, and this finding should inform the conduct of future studies.

Specific summaries of the state of the evidence for each KQ are presented below.

SFLC Assay and Diagnostic Testing (KQ1)

The addition of SFLC testing to traditional tests of electrophoresis and/or IFE for the diagnostic screening of patients suspected of having a PCD was evaluated in three studies, all quality B. The studies were all retrospective, were conducted in a hospital laboratory setting, and comprised adults suspected to have a monoclonal gammopathy. They used archived laboratory samples that had been obtained for SPEP or UPEP. All three studies reported that adding the SFLC assay to traditional tests increased diagnostic sensitivity, although the effect on diagnostic specificity was inconsistent.

Several limitations and potential biases in these studies make it difficult to present clear conclusions regarding the comparative effectiveness of the SFLC assay and limit the studies' utility for informing clinical practice. We found that demographic details, including racial breakdown and comorbid conditions, were underreported. Quantitative synthesis across the studies was not possible because of variation in the methods used to select patients, the types of PCDs examined, and the specific comparisons addressed, as well as whether patients with MGUS were included. Most studies did not report whether data assessors were blinded to diagnosis or a test group, increasing the likelihood of misclassification bias. In several studies, study samples were obtained from large repositories in laboratories, populations were selected on the basis of the need for performing SPEP, and data were analyzed only for those with parallel SFLC and traditional test results. The effects of such convenience sampling are difficult to assess. The possibility of multiple samples from the same patient being analyzed without accounting for nonindependence was also not explicitly discussed. Few studies were designed a priori as studies of diagnostic-test performance with an adequately powered sampling scheme, and not all studies included evaluation of significance or precision in the form of hypothesis testing or estimation of confidence intervals.

The diagnosis of PCDs is based on a set of criteria, including the results of the screening tests. Thus, there are potentially several types of biases that can affect diagnostic test studies for PCDs that should be considered when interpreting the results. Incorporation bias can occur because the result from the reference test itself (e.g., SPEP or SIFE) is needed to reach a diagnosis of PCD. Selection bias could occur if study samples from large laboratory repositories are selected on the basis of

the need to perform SPEP and the availability of parallel SFLC and traditional test results. The diagnostic performance of the SFLC assay varies depending on the type and distribution of PCDs in the study sample, the production of monoclonal light chains being closely dependent on the biology of the disease. Hence, the diagnostic accuracy of the SFLC assay has to be interpreted in the light of the specific PCD being diagnosed. Finally, variation in disease severity studied can lead to spectrum bias. Measures recommended to maximize the quality of test interpretation include repeat testing and targeted followup of false positives, as well as blinding of diagnosis or test group to diminish the likelihood of misclassification bias. However, such safeguards were seldom emphasized in the studies reviewed.

The purpose of this review was to examine the value added by SFLC testing to existing traditional tests; the population of interest was undiagnosed patients. Diagnostic studies using data only from patients already known to have PCDs were excluded from this CER (see Appendix B). We understand that studies of patients known to have PCDs have already been used to inform clinical practice. However, data from already diagnosed patients could potentially bias the evidence, as they reflect the extreme end of the spectrum of disease severity, for which the proportion of patients with a positive test is overestimated. Moreover, without studying a nondiseased population, true negatives cannot be assessed. Certain study designs such as the case-control approach, with different enrollment strategies for the disease and control groups, could exaggerate the reported sensitivity and specificity, invoking the possibility of spectrum bias.

SFLC Assay and Treatment Response and Survival (KQ4)

Eleven studies, three with patients with AL amyloidosis and eight with patients with MM, evaluated SFLC testing compared with traditional testing for assessing treatment response and in relation to five outcomes (overall survival, disease-free survival, remission, light chain escape, or quality of life). The studies varied in their inclusion criteria and treatments analyzed, as well as in the proportions of patients with newly diagnosed or relapsed disease and the types of traditional tests used as a comparator for the SFLC assay.

In the three studies of AL amyloidosis, a reduction in the SFLC concentration after treatment was associated with improved survival. However, it was not possible to determine whether SFLC testing is superior to traditional

testing, since SFLC responses and M protein responses were not compared directly. All three were given a quality C grade, as they were small and retrospective with evidence of selection bias. The strength of evidence underlying this comparison was therefore rated as insufficient.

The eight reviewed studies of patients with MM were mostly retrospective cohort studies, and only three were of quality B. They addressed the use of SFLC assay in assessing or predicting response to treatment and the relationship between baseline or post-treatment SFLC level and survival, as well as overall survival. The traditional test comparators reported varied in each study. Discordance of the SFLC response and the response as assessed by traditional testing was reported, although SFLC response occurred before a response on traditional tests. Studies that addressed changes in SFLC or M protein relative to survival showed conflicting results. We rated as insufficient the strength of evidence for SFLC response being a better predictor of survival than traditional testing. Limiting our consideration to the B quality studies did not qualitatively change the pattern of observations outlined above or the grading of the strength of evidence.

The strength of evidence for this KQ was insufficient for both AL amyloidosis and MM for all outcomes examined. Limitations in the literature reviewed included suboptimal reporting standards and a paucity

of information regarding high-risk subgroups such as patients with renal involvement, as well as patients across the disease spectrum (e.g., encompassing a range of types of PCD, or those without measurable disease vs. those with only SFLC production). Also, many of the studies were conducted in either single centers or as ancillary studies to preexisting trials. All these issues limited the applicability of the findings to the general PCD population and subgroups of interest.

SFLC Assay in Outcome Prediction, Treatment Decisions, and Reducing Diagnostic Tests (KQ2, KQ3, and KQ5)

We did not find any studies comparing the SFLC assay with traditional tests in predicting progression of MGUS to MM (to address KQ2). No studies compared the use of the SFLC assay with traditional tests to determine whether treatment decisions changed (with regard to timing, duration, or type of treatment) to address KQ3.

A single study explored whether the use of the SFLC assay compared with traditional testing would reduce the need for bone marrow examination in assessing response to treatment. Ten percent of patients with normalization of the SFLC ratio still had 5 percent or more of plasma cells in marrow, indicating the continued need for bone marrow testing. Since this conclusion is based on one study only, more detailed evaluation is needed.

Table A. Summary of findings for KQs 1–5

KQ	Strength of Evidence	Summary, Comments, and Conclusions
KQ1: Do the SFLC assay and the SFLC ratio improve diagnostic accuracy for PCDs when combined with traditional tests, compared with traditional tests alone, in undiagnosed patients with suspected PCD?	Insufficient (favoring use of the SFLC assay and ratio)	<ul style="list-style-type: none"> Three retrospective studies (all quality B) directly evaluated the SFLC assay in the context of diagnosing PCDs. All 3 compared test results to the diagnosis of disease verified by medical records. Although these studies showed an increase in sensitivity with the addition of the SFLC assay, because of the heterogeneity in design, patient selection, and comparators used, meta-analysis could not be performed. The effect on specificity was inconsistent. Conclusions: The SFLC assay appears to increase the sensitivity for diagnosis of PCD, although the effect on specificity was inconsistent. We rated the strength of evidence as insufficient, favoring the addition of the SFLC assay and ratio to the diagnostic test panel for PCDs.
KQ2: As compared with traditional tests, how well does the SFLC assay independently predict progression to MM in patients with MGUS?	Insufficient	<ul style="list-style-type: none"> No studies directly compared the use of the SFLC assay with traditional tests to determine whether it provided better prediction of progression to MM Conclusions: Because of the lack of directly applicable data, we rated the evidence as insufficient.

Table A. Summary of findings for KQs 1–5 (continued)

KQ	Strength of Evidence	Summary, Comments, and Conclusions
<p>KQ3: In patients with an existing diagnosis of PCD, does the use of the SFLC assay result in different treatment decisions with regard to timing, type, or duration of therapy as compared with traditional tests?</p>	<p>Insufficient</p>	<ul style="list-style-type: none"> No studies directly compared the use of the SFLC assay with traditional tests to determine whether treatment decisions were different with regard to timing, duration, or type of treatment. Conclusions: Because of the lack of directly applicable data, we rated the evidence as insufficient.
<p>KQ4: In PCD patients, is the SFLC assay a better indicator of response to treatment, and of outcomes (overall survival, disease-free survival, remission, light chain escape, and quality of life) than traditional tests?</p>	<p>Insufficient for SFLC response as a better predictor of survival than M protein response in AL amyloidosis and in MM; also insufficient for other outcomes specified</p>	<ul style="list-style-type: none"> One prospective study, 10 retrospective studies, and 1 study of unclear design (3 quality B, 8 quality C) evaluated the SFLC assay used in parallel with traditional tests in relationship to clinical outcomes, including survival. Three studies involved patients with AL amyloidosis and evaluated response to treatment as a predictor of outcomes; the other 8 studies involved patients with MM and evaluated either responses of SFLC or M protein to treatment or baseline levels of SFLC or M protein as predictors of clinical outcomes. The 3 retrospective studies in AL amyloidosis showed that patients with greater reductions in abnormal SFLC concentrations (a >50% or >90% reduction vs. lesser reductions) after treatment (either chemotherapy or stem-cell transplantation) had better survival outcomes. The relationship between quantitative reduction in M protein and outcomes was inconsistent across studies. The prevalence of measurable disease limited the utility of the SFLC assay, precluding its use in patients without elevated levels before treatment. Five of the 8 studies that enrolled patients with MM addressed the use of the SFLC assay in the assessment or prediction of treatment response. The traditional test comparators differed in each study. Four of the studies included patients who achieved an SFLC response earlier than a response by traditional tests; 2 examined the relationship between baseline SFLC concentrations and survival; 3 examined the relationship between post-treatment SFLC level and survival. Studies that addressed changes in SFLC or M protein relative to survival showed conflicting results. Conclusions: Although SFLC response to therapy appeared to be a consistent predictor of outcomes in AL amyloidosis, there was no evidence that the SFLC assay was superior to traditional tests, as direct comparisons were unavailable. Similarly, there was no evidence to ascertain whether SFLC response was a better predictor of outcomes than traditional tests in MM. We rated the strength of evidence as insufficient for the SFLC response as a better predictor of survival in AL amyloidosis and insufficient for the SFLC response as a better predictor of survival in MM.

Table A. Summary of findings for KQs 1–5 (continued)

KQ	Strength of Evidence	Summary, Comments, and Conclusions
<p>KQ5: In PCD patients, does the use of the SFLC assay reduce the need for other diagnostic tests (e.g., bone marrow biopsy)?</p>	<p>Insufficient to support the theory that use of the SFLC assay reduces the need for other diagnostic tests</p>	<ul style="list-style-type: none"> • One study (quality C) addressed this question. • The study was a retrospective review of patients with a negative IFE test after treatment of MM who had a concomitant evaluable bone marrow aspiration or biopsy. A subset of patients also had data on the SFLC ratio; among those whose ratio normalized, the percentage of clonal plasma cells in the bone marrow was examined. A total of 14% of patients with a negative IFE test had $\geq 5\%$ plasma cells in bone marrow, as did 10% with a normal SFLC ratio. • The authors recommended that, even if the SFLC assay is used, bone marrow examination should not be eliminated for the assessment of response.

AL amyloidosis = systemic amyloidosis in which amyloid [A] proteins derived from immunoglobulin light chains [L] are deposited in tissue; IFE = immunofixation electrophoresis; KQ = Key Question; MGUS = monoclonal gammopathy of undetermined significance; MM = multiple myeloma; PCD = plasma cell dyscrasia; SFLC = serum free light chain

Limitations

The present systematic review is subject to several important limitations. Few studies were available for specific comparisons between SFLC testing and traditional testing; the studies showed wide clinical heterogeneity stemming from the variation in the populations, diagnostic tests, and outcomes examined; and many were rated as poor quality. Comparators selected for the review were those that were in general use at the time of the review and did not include newer advances such as positron emission tomography. Finally, most studies were underpowered with respect to PCDs, for which the comparative role of the SFLC assay would have been the most meaningful, such as AL amyloidosis, LCMM, or NSMM.

Applicability

MGUS and other PCDs are known to be more common in African-Americans than in Caucasians in the United States, but no studies that were included in our review addressed whether race modified the applicability of the SFLC assay for diagnosis and monitoring of disease. African-American patients with MGUS have been found to have different laboratory findings than Caucasians, although the biologic differences underlying this and the effect on prognosis is unknown.²¹

Studies that addressed SFLC testing as a treatment marker for monitoring disease were often underpowered and failed to identify PCD subgroups as distinct risk categories. Given the biologic basis of the test, the comparative role of the SFLC assay is likely to be the most meaningful if disease expression is influenced by the function of a

malignant clone of plasma cells that make light chains. Such a situation may apply to certain types of disease (e.g., AL amyloidosis, LCMM, or NSMM) or stages of disease (e.g., response to treatment, relapse, or light chain escape). There were no studies that specifically targeted these settings.

Implications for Future Research

Uncertainties remain regarding the applications of the SFLC assay, both within and beyond the 2009 IMWG consensus guidelines.⁸ Areas of uncertainty span the comparative effectiveness of the adjunctive role of the assay for the diagnosis of PCDs and the adjunctive and independent role of the assay in making therapeutic decisions and monitoring disease progression, recognizing response and remission, and predicting clinical outcomes and prognosis among patients with diagnosed PCDs. The available data do not completely answer important clinical questions relevant to patient management; further research is needed to help elucidate these issues. However, given the widespread use and acceptance of SFLC testing in practice and the clinical impression of its effectiveness, the role of future research into the assay's comparative effectiveness should be targeted toward populations and settings that may greatly increase its utility.

SFLC Assay in Diagnostic Testing

Prospectively designed cohort studies, representative of the clinically relevant population in which a PCD may be suspected, are needed to provide a more accurate assessment of the effect of adding SFLC to traditional testing. Studies only involving patients diagnosed with

PCD would reflect the extreme end of the spectrum of disease severity, overestimating the proportion of patients with a positive test. Without a population with no PCD, true negatives cannot be assessed. The higher sensitivity of the SFLC assay potentially increases the number of false-positive results; hence, a more systematic study of the false-positive rate of the SFLC assay in different settings is needed, as is study of the best approach to resolve the discordance of a positive SFLC result but a negative result on traditional tests. Studies should have an a priori calculation of the sample size needed to determine the desired precision and should include inferences based on formal statistical testing of estimates of diagnostic accuracy.

Other important issues relate to validity of the published reference ranges, within-patient inconsistency in SFLC concentrations, and the harms of testing—questions that were outside the scope of this review. In addition, the lack of a suitable reference standard for PCD diagnosis and the need for a panel of tests to satisfy the criteria for diagnosis complicate the ability to make valid inferences from the data. Finally, conditions such as polyclonal gammopathy and diminished kidney function can produce false-positive test results in the SFLC assay, and certain settings such as antigen excess and technical variations in commercial assays can produce false-negative results as well. As new diagnostic tests emerge for PCDs (e.g., positron emission tomography) and modifications of the SFLC assay evolve (e.g., “N Latex” SFLC assay), future research is needed to elucidate how these tests affect the clinical use of the SFLC assay.

Although the elimination of the need for 24-hour urine collection would add tremendous value to the diagnostic testing protocol, this approach needs to be validated in undiagnosed populations, where the danger of false negatives for the SFLC assay can be thoroughly vetted. Therefore the question of the SFLC assay being able to replace 24-hour urine collections in a diagnostic panel remains as an evidence gap.

SFLC Assay in Risk Stratification and in Determining Prognosis

In addition to its diagnostic use, the SFLC assay is being used to monitor the course of PCDs characterized by light chain production. Definitions of FLC response are largely empirical in the current guidelines for AL amyloidosis and MM and have not been validated. Research is needed to address the best definition of FLC response and the relationship of FLC response to hematological response

and M protein response, progression-free survival, and overall survival. Similarly, a range of definitions have been used to describe the predictive clinical findings of the SFLC assays, including the absolute concentrations of the involved light chain, the difference between the concentrations of each type of light chain, and the SFLC ratio. These definitions are not standardized, and it remains unclear which is optimal in a variety of clinical situations.

Future studies should also clarify whether SFLC measurement can replace the 24-hour UPEP or UIFE in disease monitoring and the potential of the SFLC assay to obviate invasive testing such as bone marrow aspiration or biopsy or radiation exposure from skeletal surveys. In addition, there is a need to examine the role of the SFLC assay in risk stratification across the spectrum of PCDs, from MGUS to MM and its variants as well as AL amyloidosis. There is a growing awareness that specific gene rearrangements are associated with FLC production across the spectrum of PCDs. Risk stratification according to findings on the SFLC assay may therefore provide a marker for the biological variability of the PCD. Such insight could provide guidance about the timing, duration, or type of treatment decisions used. This could be a major area for future research.

Reporting on the SFLC Assay

Finally, there is a need to standardize the reporting of SFLC results for diagnostic test performance studies or of cohort studies in this area. At a minimum, studies should consistently report complete information on the mode of enrollment and on population characteristics, including demographic data. Future studies of SFLC testing should also report details on frequency and periodicity of measurements to account for within-patient variability.

Conclusions

We did not find sufficient evidence to determine whether the addition of the SFLC assay to traditional testing would increase the diagnostic accuracy of PCD and whether it would help prognosticate the disease course. Its precise role and optimal use across the spectrum of PCDs and clinical settings still need to be defined. Potential areas where its benefit may be seen are in diagnosis and prognosis, monitoring of therapy, and aiding treatment decisions. Future research should focus on standardization of patient inclusion criteria, testing of diagnostic and disease monitoring algorithms, and defining outcome and response definitions.

References

1. Kyle RA, Rajkumar SV. Epidemiology of the plasma-cell disorders. *Best Pract Res Clin Haematol.* 2007;20:637-64.
2. Ries LAG, Harkins D, Krapcho M, et al. SEER cancer statistics review, 1975-2003. NCI, Bethesda, MD; 2005. http://seer.cancer.gov/csr/1975_2003.
3. Jemal A, Siegel R, Ward E, et al. Cancer Statistics, 2007. *CA: A Cancer Journal for Clinicians.* 2007;57(1):43-66.
4. Katzel JA, Hari P, Vesole DH. Multiple Myeloma: Charging Toward a Bright Future. *CA: A Cancer Journal for Clinicians.* 2007;57(5):301-18.
5. Cook R. An economic perspective on treatment options in multiple myeloma. *Managed Care Oncol.* 2007;2007(Spring):10-12.
6. Messori A, Trippoli S, Santarlasci B. Pharmacotherapy of multiple myeloma: an economic perspective. *Expert Opin Pharmacother.* 2003 Apr 1;4(4):515-24. PMID: 12667114.
7. Bradwell AR, Carr-Smith HD, Mead GP, et al. Highly sensitive, automated immunoassay for immunoglobulin free light chains in serum and urine. *Clin Chem.* 2001 Apr;47(4):673-80.
8. Dispenzieri A, Kyle R, Merlini G, et al. International Myeloma Working Group guidelines for serum-free light chain analysis in multiple myeloma and related disorders. [Review] [43 refs]. *Leukemia.* 2009 Feb;23(2):215-24.
9. Mead GP, Carr-Smith HD, Drayson MT, et al. Serum free light chains for monitoring multiple myeloma. *British Journal of Haematology.* 2004 Aug;126(3):348-54.
10. Sanchorawala V, Seldin DC, Magnani B, et al. Serum free light-chain responses after high-dose intravenous melphalan and autologous stem cell transplantation for AL (primary) amyloidosis. *Bone Marrow Transplantation.* 2005 Oct;36(7):597-600.
11. Katzmann JA. Serum free light chains: quantitation and clinical utility in assessing monoclonal gammopathies. *Clin Lab News.* 2006 Jun;June:12-14.
12. Kyle RA, Durie BGM, Rajkumar SV, et al. Monoclonal gammopathy of undetermined significance (MGUS) and smoldering (asymptomatic) multiple myeloma: IMWG consensus perspectives risk factors for progression and guidelines for monitoring and management. *Leukemia.* 2010 Jun;24(6):1121-27.
13. Bradwell AR. Serum free light chain measurements move to center stage. *Clin Chem.* 2005 May;51(5):805-07.
14. Whitlock EP, Lopez SA, Chang S, et al. AHRQ Series Paper 3: Identifying, selecting, and refining topics for comparative effectiveness systematic reviews: AHRQ and the Effective Health-Care program. *J Clin Epidemiol.* 2010;63:491-501.
15. Agency for Healthcare Research and Quality. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. AHRQ Publication No. 10(12)-EHC063-EF. Rockville MD: Agency for Healthcare Research and Quality. April 2012.
16. Agency for Healthcare Research and Quality. Methods Guide for Medical Test Reviews. Rockville, MD; 2010. www.effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?productid=558&pageaction=displayproduct.
17. Durie BG, Harousseau JL, Miguel JS, et al. International uniform response criteria for multiple myeloma.[Erratum appears in *Leukemia.* 2007 May;21(5):1134].[Erratum appears in *Leukemia.* 2006 Dec;20(12):2220]. *Leukemia.* 2006 Sep;20(9):1467-73.
18. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA.* 1994 Mar 2;271(9):703-07.
19. Kyrtonis MC, Vassilakopoulos TP, Kafasi N, et al. Prognostic value of serum free light chain ratio at diagnosis in multiple myeloma. *British Journal of Haematology.* 2007 May;137(3):240-43.
20. Gertz MA, Comenzo R, Falk RH, et al. Definition of organ involvement and treatment response in immunoglobulin light chain amyloidosis (AL): a consensus opinion from the 10th International Symposium on Amyloid and Amyloidosis, Tours, France, 18-22 April 2004. [Review] [100 refs]. *American Journal of Hematology.* 2005 Aug;79(4):319-28.
21. Weiss BM, Minter A, Abadie J, et al. Patterns of monoclonal immunoglobulins and serum free light chains are significantly different in black compared to white monoclonal gammopathy of undetermined significance (MGUS) patients. *American Journal of Hematology.* 2011 Jun;86(6):475-78.

Full Report

This executive summary is part of the following document:

Rao M, Yu W, Chan J, Patel K, Comenzo R, Lamont JL, Ip S, Lau J. Serum Free Light Chain Analysis for the Diagnosis, Management, and Prognosis of Plasma Cell Dyscrasias. Comparative Effectiveness Review No. 73. (Prepared by the Tufts Evidence-based Practice Center under Contract No. 290-2007-10055 I.) AHRQ Publication No. 12-EHC102-EF. Rockville, MD: Agency for Healthcare Research and Quality. August 2012.
www.effectivehealthcare.ahrq.gov/reports/final.cfm.

For More Copies

For more copies of Serum Free Light Chain Analysis for the Diagnosis, Management, and Prognosis of Plasma Cell Dyscrasias. Executive Summary No. 73 (AHRQ Pub. No. 12-EHC102-1), please call the AHRQ Publications Clearinghouse at 1-800-358-9295 or email ahrqpubs@ahrq.gov.

