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

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

Contract No. 290-2007-10055 I

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

Investigators:
Madhumathi Rao, M.D., Ph.D., Project Lead, Investigator
Winifred W. Yu, Ph.D., M.S., R.D., Research Associate
Jeffrey Chan, B.S., Research Associate
Kamal Patel, M.P.H., M.B.A., Research Associate
Raymond Comenzo, M.D.
Jenny L. Lamont, M.S., Medical Editor
Stanley Ip, M.D.
Joseph Lau, M.D.

AHRQ Publication No. 12-EHC102-EF
August 2012
Preface

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

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

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

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

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

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

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

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

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

Elisabeth Kato, M.D., M.R.P.
Task Order Officer
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality
Key Informants
James Adamson, M.D.
Senior VP and Medical Officer for National Accounts
Arkansas Blue Cross Blue Shield
Little Rock, AR

Michael Katz
Vice President of International Myeloma Foundation
Los Angeles, CA

Susan Miller Maynard, B.S., M.S., Ph.D., F.A.C.B.
Director of Chemistry and Toxicology
Carolinas Medical Center
Charlotte, NC

Mary E. O’Donnell
President/CEO of Amyloidosis Foundation
Clarkston, MI

Mitchell G. Scott, Ph.D.
Professor of Pathology and Immunology
Co-Medical Director, Clinical Chemistry
Washington University School of Medicine
St. Louis, MO

Technical Expert Panel
Laura M. Dember, M.D.
Professor of Medicine
Renal, Electrolyte and Hypertension Division
University of Pennsylvania Perelman School of Medicine
Philadelphia, PA

Shaji Kumar, M.D.
Associate Professor of Medicine at Mayo Clinic College of Medicine
Mayo Clinic
Rochester, MN

Barbarajean Magnani, Ph.D., M.D.
Pathologist-in-Chief
Department of Pathology and Laboratory Medicine, Tufts Medical Center
Professor of Pathology, Tufts University School of Medicine
Boston, MA

Susan Miller Maynard, B.S., M.S., Ph.D., F.A.C.B.
Director of Chemistry and Toxicology
Carolinas Medical Center
Charlotte, NC

Philip McCarthy Jr., M.D.
Professor of Oncology
Director of Blood & Marrow Transplant Program
Roswell Park Cancer Institute
Buffalo, NY

Martha Q. Lacy, M.D.
Professor of Medicine at Mayo Clinic College of Medicine
Mayo Clinic
Rochester, MN
Peer Reviewers
Adam D. Cohen, M.D.
Assistant Professor of Medical Oncology
Fox Chase Cancer Center
Philadelphia, PA

Rochelle Fu, Ph.D.
Associate Professor of Biostatistics
Oregon Evidence-based Practice Center
Department of Public Health and Preventive Medicine
Oregon Health and Science University
Portland, OR

Jerry Katzmann, Ph.D.
Associate Professor of Biochemistry & Molecular Biology and Laboratory Medicine
Mayo Clinic
Rochester, MN
Serum Free Light Chain Analysis for the Diagnosis, Management, and Prognosis of Plasma Cell Dyscrasias

Structured Abstract

Objectives. To summarize the literature regarding the role of the serum free light chain (SFLC) assay in diagnosis as an adjunct to traditional tests (defined as serum and urine electrophoresis or immunofixation electrophoresis), compared with traditional testing alone, and its role compared with traditional tests in the management of patients with plasma cell dyscrasias (PCDs).

Data Sources. MEDLINE®, the Cochrane Central Register of Controlled Trials, and the Cochrane Database of Systematic Reviews from January 2000 through January 2012.

Methods. We used established systematic review methods, selecting only published, peer-reviewed, English-language articles on the basis of predetermined eligibility criteria. A standardized protocol was used to extract details on designs, diagnoses, interventions (diagnostic tests/disease monitoring), outcomes, and study methods. We considered studies of adults with suspected and diagnosed PCDs, specifically monoclonal gammopathy of undetermined significance (MGUS) or multiple myeloma (MM), which includes light chain MM, nonsecretory MM, and AL amyloidosis. The comparison and outcomes of interest were the role of the SFLC assay as an adjunct to traditional tests for diagnosis of PCDs, and the effectiveness of the SFLC assay versus traditional tests for studying progression to MM, treatment response, and prognosis.

Results. The literature search yielded 3,036 citations, with 2,711 excluded at the abstract level. The remaining 325 articles were retrieved for full-text review, upon which 310 were excluded, most often because studies did not meet all the predefined eligibility criteria or were not comparative. A total of 15 studies were included. Three retrospective, fair-quality studies evaluated the SFLC assay as an adjunct to traditional testing in populations suspected of having a PCD. Three retrospective, poor-quality studies of AL amyloidosis, and eight studies (three of fair quality and five of poor quality) of MM, six of which were retrospective, evaluated either baseline or post-treatment concentrations of SFLC or monoclonal protein as predictors of clinical outcomes. Overall, because of the small number of studies and their poor methodological quality and considerable clinical heterogeneity, the strength of evidence was rated as insufficient regarding: (1) the value of adjunct SFLC testing on diagnostic accuracy in undiagnosed patients, (2) the role of the SFLC assay as a better predictor of outcome in PCDs or of progression of MGUS to MM, and (3) the role of the SFLC assay as a better indicator for therapeutic decisionmaking compared with traditional testing alone and as a substitute for other diagnostic tests.

Conclusions. The role of the SFLC assay remains to be defined. The evidence was rated as insufficient to suggest that the assay may increase sensitivity when used as an adjunct to traditional testing for diagnosis of PCDs or that it was more effective for predicting and monitoring treatment response and for predicting patient survival. Methodological limitations of the studies reviewed preclude definitive conclusions regarding these potential uses. Future research should focus on standardization of diagnostic testing and monitoring algorithms in
prespecified patient populations, with adherence to accepted definitions of outcomes and responses.
Contents

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

Introduction ................................................................................................................................................................1
  Plasma Cell Dyscrasias ........................................................................................................................................... 1
  SFLC Assay, Guidelines, and Current Use ............................................................................................................. 1
    SFLC Assay ......................................................................................................................................................... 1
    Guidelines ............................................................................................................................................................ 2
    Clinical Effectiveness and Use in Practice ........................................................................................................ 3
  Context of This Comparative Effectiveness Review ............................................................................................ 3
  Key Questions ........................................................................................................................................................ 3

Methods .................................................................................................................................................................. 5
  AHRQ Task Order Officer .................................................................................................................................... 5
  External Expert Input ........................................................................................................................................... 5
  Analytic Framework ............................................................................................................................................ 5
  Literature Search .................................................................................................................................................. 6
  Study Selection and Eligibility Criteria ................................................................................................................ 6
  Data Extraction and Data Management ............................................................................................................... 6
  Assessment of Risk of Bias .................................................................................................................................. 7
  Data Synthesis ....................................................................................................................................................... 8
  Grading the Body of Evidence for Each KQ ......................................................................................................... 9
  Peer Review and Public Commentary .............................................................................................................. 9

Results .................................................................................................................................................................. 10
  Literature Search ................................................................................................................................................... 10
  Study Quality Grade and Overall Strength of Evidence ...................................................................................... 10
  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? .................................................................................................................... 12
    Results ............................................................................................................................................................... 12
    Summary .......................................................................................................................................................... 12

  KQ2: As compared with traditional tests, how well does the SFLC assay independently predict progression to MM in patients with MGUS? ........................................................................................................................................... 16

  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? ....................................................................................................................... 16

  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)? ....................................................................................................................... 16
    Results ............................................................................................................................................................... 16

  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)? ......................................................................................... 30
    Results ............................................................................................................................................................... 30
    Summary .......................................................................................................................................................... 30
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.\textsuperscript{1} Multiple myeloma (MM) is the most common malignant plasma-cell tumor, accounting for about 1 percent of all cancer types,\textsuperscript{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,\textsuperscript{2} an estimated 19,900 new diagnoses and 10,790 deaths due to myeloma occurred in 2007, according to the American Cancer Society.\textsuperscript{3} Although the median survival has improved to 5 years with current standards of treatment,\textsuperscript{4} the annual costs of modern therapies can range from $50,000 to $125,000 per patient.\textsuperscript{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.

The serum FLC (SFLC) assay (i.e., the Freelite\textsuperscript{®} assay, The Binding Site Ltd., Birmingham, United Kingdom) was introduced in 2001 to measure the FLC component in serum.\textsuperscript{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.\textsuperscript{8} The assay could allow for quantitative monitoring of response and remission after treatment and provide prognostic information,\textsuperscript{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.\textsuperscript{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.\textsuperscript{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\textsuperscript{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.\textsuperscript{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 immunoglobulins\textsuperscript{13}), the assay may also help detect relapses and treatment failures earlier than by reliance on M protein concentrations alone.\textsuperscript{10}

Although the SFLC assay has been in use for a decade, how best to incorporate it into practice remain unclear.\textsuperscript{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.

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.

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
- **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 concentrations8,17), light chain escape, or quality of life
- **KQ5**: clinic visits, bone marrow biopsies, or skeletal surveys

Studies could have any length of followup8,17 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 analyses 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])

**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.\textsuperscript{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.

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. 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. 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.
Table A. Summary of findings for KQs 1–5

<table>
<thead>
<tr>
<th>KQ</th>
<th>Strength of Evidence</th>
<th>Summary, Comments, and Conclusions</th>
</tr>
</thead>
</table>
| 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) | • 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              | • 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.                                                                 |
| 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? | Insufficient              | • 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.                                                                 |
**Table A. Summary of findings for KQs 1–5 (continued)**

<table>
<thead>
<tr>
<th>KQ</th>
<th>Strength of Evidence</th>
<th>Summary, Comments, and Conclusions</th>
</tr>
</thead>
</table>
| 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? | 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 | - 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. |

| KQ5: In PCD patients, does the use of the SFLC assay reduce the need for other diagnostic tests (e.g., bone marrow biopsy)? | Insufficient to support the theory that use of the SFLC assay reduces the need for other diagnostic tests | 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 ≥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.
References


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.


Plasma Cell Dyscrasias

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,\(^2\) the American Cancer Society estimated that there were 19,900 new diagnoses and 10,790 deaths due to myeloma in 2007.\(^3\) 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.\(^5,6\)

Plasma cells arise from B cells in the bone marrow and produces immunoglobulins that constitute the body’s normal humoral immune response. The immunoglobulin molecule is composed of a heavy chain and a light chain. Plasma cells normally produce light chains in excess that do not bind to heavy chains to form a complete immunoglobulin molecule and instead enter the bloodstream as free light chains (FLCs).

In PCDs, each abnormally expanded clone of malignant plasma cells produce an excess of either intact immunoglobulin or 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.

PCDs range in severity. The mildest and most common PCD is the precancerous monoclonal gammopathy of undetermined significance (MGUS), affecting approximately 3 percent of the general population 50 years of age or older.\(^1\) MGUS can progress to asymptomatic MM (also called smoldering or indolent MM) or symptomatic MM. The M proteins produced in MM are either intact immunoglobulins or FLCs or both. Rarer MM variants include light chain MM (LCMM, formerly known as Bence Jones myeloma), characterized by expanded FLC-producing clones, and oligosecretory or nonsecretory MM (NSMM), in which few detectable light- or heavy-chain M proteins are secreted. Other PCDs include systemic (primary) AL amyloidosis, (also called light chain amyloidosis) in which amyloid [A] proteins derived from immunoglobulin light chains [L] are deposited in tissue, as well as macroglobulinemia, solitary plasmacytoma, and plasma-cell leukemia. AL amyloidosis can be a complication of MM but is often considered a distinct disorder related to a relatively stable, slow-growing plasma-cell clone and organ dysfunction.

SFLC Assay, Guidelines, and Current Use

SFLC Assay

The serum free light chain (SFLC) assay (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 approved by the U.S. Food and Drug Administration (FDA) and is
classified as an immunoglobulin light chain-specific immunological test system. It measures kappa and lambda light chains separately and detects low concentrations of FLCs—less than 1 mg/dL in serum and less than 200 mg/day in urine. The other main advantage is the ability to measure the ratio of kappa chains to lambda chains, for which the normal range is 0.26 to 1.65. An abnormal ratio provides a useful index of clonality, as clonal disorders produce disproportionately high concentrations of a single type of light chain. In a given case of PCD, if kappa chains are in excess, the kappa/lambda ratio is greater than 1.65; if the lambda chains are in excess, the ratio is less than 0.26.

Guidelines

The International Myeloma Working Group (IMWG) recommends the following actions and tests for evaluation of a patient suspected of having a myeloma: a complete history taking and physical examination; routine laboratory testing including serum protein electrophoresis (SPEP), serum immunofixation electrophoresis (SIFE), nephelometric quantitation of immunoglobulins, and measurement of serum FLCs (SFLCs); bone marrow aspiration and biopsy with immunophenotyping, conventional cytogenetics, and fluorescence in situ hybridization; and imaging. Thus, testing for M protein is only one part—albeit an integral part—of a suite of tests done to diagnose PCDs.

M protein measurement and typing are traditionally achieved through the use of SPEP and/or urine protein electrophoresis (UPEP) and SIFE and/or urine immunofixation electrophoresis (UIFE), plus immunoglobulin quantification. These traditional tests have relatively low sensitivity, especially regarding concentrations of SFLCs. This lack of sensitivity results in many undetected cases of PCDs involving excess FLCs. It is likely that up to 3 percent of cases of NSMM, LCMM, or AL amyloidosis are not detected by traditional tests. To increase the chance of detection of FLCs in urine, 24-hour urine collection has been recommended, along with procedures to concentrate urine samples. Yet these adaptations can be cumbersome for patients and providers, affecting compliance and test accuracy.

In general, for diagnosis, SPEP is estimated to detect an immunoglobulin peak in 82 percent of patients with MM. The addition of SIFE increases the sensitivity to 93 to 95 percent, which is further increased to 97 percent by performing UPEP and UIFE.

It has been suggested that the SFLC assay could play an adjunctive role in screening, diagnosis, monitoring, and prognosis of PCDs in high-risk populations. The IMWG currently considers the SFLC assay to be an adjunct to traditional tests. The assay could allow for quantitative monitoring of response and remission after treatment and provide prognostic information, potentially reducing the need for frequent bone marrow biopsy for purposes of quantifying plasma cells, which is required as part of stringent monitoring for MGUS progression to MM or defining disease remission. It could potentially be used in conjunction with SPEP and SIFE to replace urine tests that require 24-hour collection (UPEP and UIFE), which could simplify diagnosis and disease monitoring. The SFLC assay may also be the only means of detecting a disease marker in some disease settings: NSMM, where SFLCs are often the only marker of the disease; AL amyloidosis, where low M protein concentrations may not be detected by means of conventional techniques; and LCMM, where the M protein consists only of FLCs. 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 immunoglobulins), the assay may also help detect relapses and treatment failures earlier than by reliance on M protein concentrations alone.
Clinical Effectiveness and Use in Practice

Although the SFLC assay has been in use for a decade, how best to incorporate it into practice remain unclear. The test appears to have been widely adopted by clinicians as an adjunct to the panel of tests used to diagnose PCDs, given the assay’s biological validity and ease of use as compared with cumbersome urine collections. Its use is also being evaluated in patient management. The SFLC assay has successfully been used to define disease subcategories and improve risk stratification. The test is efficient in the diagnosis of AL amyloidosis, as is reflected in the International Society of Amyloidosis Consensus Response criteria.

But uncertainties regarding the optimal use of the SFLC assay remain. 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 its comparative effectiveness will allow for the use of the assay to be refined and recommendations optimized; these aspects are addressed in the present comparative effectiveness review (CER). Evaluations of clinical utility should take into consideration different clinical settings and phases of disease as well as different disease populations.

Context of This Comparative Effectiveness Review

The aim of this CER is to evaluate the body of evidence that exists to address 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 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. These questions were vetted by panels of Key Informants and Technical Experts who assisted in identifying the important areas for evidence review (as discussed in the Methods section). 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 (PICO). Many articles in the literature address clinical but not comparative effectiveness and therefore did not meet our stated goals.

Key Questions

Five KQs were formulated in consultation with American Association for Clinical Chemistry (AACC) and the Agency for Healthcare Research and Quality (AHRQ).

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)?
Methods

This CER evaluates the SFLC assay as an adjunctive diagnostic and prognostic tool for various PCDs in addition to the standard diagnostic tests for PCDs. The evidence presented was obtained through a systematic review of the published scientific literature using established methodologies as outlined in the AHRQ’s Methods Guide for Effectiveness and Comparative Effectiveness Reviews and Methods Guide for Medical Test Reviews.

AHRQ Task Order Officer

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

External Expert Input

During a topic refinement phase, the initial questions were refined with input from a panel of Key Informants. Key Informants included representatives from AACC; experts in renal amyloidosis, clinical chemistry, and general internal medicine and geriatrics; 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), which served in an advisory capacity to help refine KQs, identify important issues, and define parameters for the review of evidence. Discussions among the EPC, TOO, 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.

Analytic Framework

The five KQs take into account the patient populations, interventions (diagnostic tests/disease monitoring), comparators, outcomes, timing, and settings (PICOTS) that are clinically relevant to the use of the SFLC analysis. Specifically, they pertain to the diagnosis of PCDs, prognosis (i.e., progression from MGUS to MM as well as overall and disease-free survival in patients with a malignant PCD), change in treatment decisions, assessment of response to treatment, and reduction of the need for other diagnostic tests (e.g., bone marrow biopsy).

To guide the development of the KQs, we generated an analytic framework (Figure 1) that maps the specific linkages associating the population (patients with PCD symptoms) and subgroups of interest to the additional tests (i.e., SFLC analysis in addition to traditional testing) and comparator (traditional tests alone), and 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.
Figure 1. 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 gammapathy of undetermined significance, MM=multiple myeloma, NSMM=nonsecretory multiple myeloma, PCD=plasma cell dyscrasia, SFLC=serum free light chain.

Literature Search

We conducted literature searches of studies published from January 1, 2000, through January 31, 2012, in MEDLINE®, the Cochrane Central Register of Controlled Trials, and the Cochrane Database of Systematic Reviews. A start year of 2000 was chosen because the SFLC assay was approved by the FDA in 2001; any reports of clinical use of the assay prior to 2000 would not be representative of the approved test. 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” (see Appendix A for complete search strings). TEP members were also invited to provide additional search terms.

Study Selection and Eligibility Criteria

We included published, peer-reviewed articles only. We did not use unpublished data, non–English-language studies, abstracts, or conference proceedings. The consensus of the TEP was not to include unpublished data or studies in the form of single case reports. Case series were included on the basis of the prevalence of the type of PCD (with lower thresholds applied for rarer forms), as long as extractable quantitative data were present. Sample size thresholds were chosen primarily on the basis of practical consideration of available resources and time, taking into consideration the likely yield of available literature. We did not contact authors for additional data.

Abstracts were manually screened, using Abstrackr, by two members of the team independently to ascertain whether they met the predefined eligibility criteria (see next paragraph) and exclusions and were reviewed by a second member of the team. Articles that were excluded after full-text screening are listed, with the reasons for exclusion, in Appendix B. Articles whose abstracts were relevant, as well as those that did not clearly signal inclusion or exclusion, were retrieved in full text for detailed evaluation to determine eligibility. During full-text evaluation, equivocal articles were read by at least two team members.
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 evaluation, skeletal survey). (Newer tests [e.g., positron emission tomography26] that were not in general use were not addressed.)

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 to have a PCD;
- **KQ2**: studies of patients with MGUS;
- **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 concentrations11,27), light chain escape, or quality of life; and
- **KQ5**: clinic visits, bone marrow biopsies, or skeletal surveys.

Studies could have any length of followup11,27 or any setting (primary or specialty care, in-facility or home, inpatient or outpatient).

### Data Extraction and Data Management

Eight articles were extracted simultaneously by all researchers for training purposes. Subsequently, each study was extracted by one methodologist and this extraction was reviewed and confirmed by at least one other methodologist. Any disagreements were resolved by discussion in team meetings. Data were extracted into tables in Microsoft Word, designed to capture all elements relevant to the KQs. Briefly, we extracted bibliographic data, eligibility criteria, enrollment years, and sample size for all studies. We also extracted population characteristics such as basic demographic data—age, sex, and race or ethnic group—as well as any factors that may have a role in the outcome of PCDs, such as type of PCD, presence of anemia, light chain or M protein type and concentration, organ involvement, treatment and other pertinent characteristics, and test-related characteristics such as diagnostic performance. The forms were tested on several articles and revised before commencement of full data extraction.

### Assessment of Risk of Bias

For assessment of risk of bias, we used predefined methods for evaluating study quality pertinent to risk of bias that are common within the EPC Program.23,28,29 Briefly, we used a three-
category (A, B, or C) grading system to denote the methodological quality of each study. This system involves a generic grading scheme that is applicable to varying study designs including randomized controlled trials, nonrandomized comparative trials, and cohort and case–control studies.

In the present report, the majority of the studies were related to testing of diagnostic performance and prediction of 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). The modified checklists used for quality assessment are provided in Appendix C, along with how each study fulfilled those criteria and the quality grade assigned to each.

The specific criteria of each grade are as follows:

- **A (good).** Quality A studies are those judged to have the least likelihood of bias and their results are considered valid. They possess, at a minimum, the following: a representative study population with both disease and nondiseased groups, no verification bias, a clear description of the reference test (if applicable), and no selection bias. Ideally, the population, setting, interventions (diagnostic tests/disease monitoring), and comparison groups are well defined and there is appropriate measurement of outcomes, appropriate statistical and analytic methods and reporting, complete and consistent overall reporting, clear accounting of dropouts, and a low dropout rate. For this review of diagnostic test studies, only studies with a sample size of at least 100 patients in total could receive a grade of A; these studies could be either prospective or retrospective.

- **B (fair).** Quality B studies are susceptible to some bias but not sufficiently to invalidate results. They do not meet all the minimum criteria in category A, owing to some deficiencies, but none of these are likely to introduce major bias. Quality B studies may be missing information, making it difficult to assess limitations and potential problems.

- **C (poor).** Quality C studies have a substantial risk of bias that may invalidate the reported findings. These studies have serious errors in design, analysis, or reporting and contain discrepancies in reporting or have large amounts of missing information.

Quality assessment was performed by the team member responsible for primary data extraction. The quality grade was confirmed by at least one other team member.

**Data Synthesis**

We summarized all included studies in narrative form and in summary tables (all of which are in the Results section) that succinctly describe the important features of the study population, design, intervention (diagnostic test/disease monitoring), outcomes, results, and study quality. We included diagnostic performance parameters, risk estimates, and their 95 percent confidence intervals and p values where applicable. Results are presented in separate summary tables for each KQ.

We conducted mainly descriptive analyses 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.
Grading the Body of Evidence for Each KQ

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 assessed the consistency of the data as either “no inconsistency” or “inconsistency present” (or not applicable if only one study). The direction, magnitude, and statistical significance of all studies were evaluated in assessing consistency, and logical explanations were provided in the presence of equivocal results. We also assessed the precision of the evidence on the basis of the degree of certainty surrounding an effect estimate. A precise estimate was considered an estimate that would allow for a clinically useful conclusion. An imprecise estimate was one for which the confidence interval is wide enough to preclude a conclusion.

Ratings were defined as follows:

- **High.** There is high confidence that the evidence reflects the true effect. Further research is very unlikely to change our confidence in the estimate of effect. No important scientific disagreement exists across studies. At least two quality A studies are required for this rating. In addition, there must be evidence regarding objective clinical outcomes.
- **Moderate.** There is moderate confidence that the evidence reflects the true effect. Further research may change our confidence in the estimate of effect and may in fact change the estimate. Little disagreement exists across studies. Moderately rated bodies of evidence contain fewer than two quality A studies or such studies lack long-term outcomes of relevant populations.
- **Low.** There is low confidence that the evidence reflects the true effect. Further research is likely to change the confidence in the estimate of effect and is likely to change the estimate. Underlying studies may report conflicting results. Low rated bodies of evidence could contain either quality B or C studies.
- **Insufficient.** Evidence is either unavailable or does not permit a conclusion. There are sparse or no data. In general, when only one study has been published, the evidence is considered insufficient, unless the study is particularly large, robust, and of good quality.

These ratings provide a shorthand description of the strength of evidence supporting the major questions we addressed. However, they by necessity may oversimplify the many complex issues involved in appraising a body of evidence. The individual studies involved in formulating the composite rating may differ in their design, reporting, and quality. The strengths and weaknesses of the individual reports, as described in detail in the text and tables, should also be considered.

Peer Review and Public Commentary

Experts in MM and/or AL amyloidosis and clinical chemistry and individuals representing stakeholder and user communities were invited to provide external peer review of this CER; AHRQ and an associate editor also provided comments. The draft report was posted on the AHRQ website for 4 weeks to elicit public comment. We addressed all reviewer comments, revising the text as appropriate, and documented everything in a disposition of comments report that will be made available 3 months after the Agency posts the final CER on the AHRQ Web site.
Results

Literature Search

The literature search yielded 3036 citations (Figure 2). Of these, 2711 were excluded at the abstract level. The remaining 325 articles were retrieved for full-text review, upon which 310 were excluded. Most of the exclusions were studies that did not meet all of the predefined PICO 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). (See Appendix B for the list of rejected articles and the rationale for their rejection.) A total of 15 studies that were both comparative and met all the CER eligibility criteria were included.

All included studies either used the Freelite assay for measuring SFLCs or referred to measurement of SFLCs or to a nephelometric technique for their measurement. We targeted any data describing, or permitting the inference of, a comparison between any single or group of traditional tests (SPEP, UPEP, SIFE, or UIFE) used to detect PCDs (particularly MGUS, MM [including LCMM and NSMM], or AL amyloidosis) and the same single test or group of tests with an SFLC assay added. Studies of diagnosis, progression, and treatment of PCDs were all of interest.

Study Quality Grade and Overall Strength of Evidence

Table 1 summarizes the relevance and quality of the 15 studies reviewed in detail. The studies are organized by which KQ they addressed and the quality grade they were assigned. The criteria met by each study and its quality grade are provided in Appendix C.

<table>
<thead>
<tr>
<th>Quality</th>
<th>KQ1</th>
<th>KQ2</th>
<th>KQ3</th>
<th>KQ4</th>
<th>KQ5</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Quality B</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Quality C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Total studies</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>1</td>
<td>15</td>
</tr>
</tbody>
</table>

Overall strength of evidence: Insufficient Insufficient Insufficient Insufficient Insufficient

KQ=Key Question.
Figure 2. Summary of search and selection of articles

3,036 citations identified from Jan. 2000 - Jan. 2012 in MEDLINE, the Cochrane Central Register of Controlled Trials, and the Cochrane Database of Systematic Reviews for articles on SFLC analysis for the diagnosis, management, or prognosis of PCDs, published in English.

2,711 Abstracts excluded for ≥1 of the following reasons:
- FLC not studied
- Diagnosis not relevant
- Narrative review
- Conference proceeding
- Single case study
- Animal study

325 Articles retrieved for full-text review

310 Articles excluded for ≥1 of the following reasons:
- Not relevant re-test, population, or comparison
- Narrative review or commentary
- Study of single case
- Letter without data

15 Articles included:
- 3 for KQ1
- 0 for KQ2
- 0 for KQ3
- 11 for KQ4
- 1 for KQ5

FLC=free light chain; KQ=Key Question; PCD=plasma cell dyscrasia; PICO=population, intervention (diagnostic test/disease monitoring), comparator, and outcome; SFLC=serum free light chain
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?

Results

Three studies\textsuperscript{31-33} evaluated the addition of SFLC testing to traditional testing for the diagnosis of PCDs in undiagnosed patients suspected of having a PCD. The study characteristics and findings are listed in Tables 2, 3, and 4.

Each study was rated B quality because of the retrospective design and because formal statistical comparisons and confidence intervals were not provided. All three studies compared test results with the diagnosis of disease verified by medical records on the basis of a panel of criteria. One study reported industry-associated funding and also was the only study of the three to report the demographic characteristics of the study population\textsuperscript{31}

Abadie 2006\textsuperscript{31} examined the diagnostic accuracy of the SFLC assay, with or without SPEP, in 312 consecutive, predominantly male veterans without a prior diagnosis of PCD. Fifteen percent of the patients were found on diagnostic testing to have a malignant PCD. The use of SPEP alone had a diagnostic sensitivity of 0.88 and a specificity of 0.98, with 15 false negatives (12 for MM and 1 each for Waldenstrom’s macroglobulinemia, AL amyloidosis, and lymphoma). SPEP used in combination with the SFLC assay increased the sensitivity to 1.00 and the specificity to 0.99, although use of the SFLC assay alone showed four false negatives (two for MM and two for “potential MM”).

Piehler 2008\textsuperscript{32} measured SFLCs, as well as performing SPEP, in 332 patients suspected of having monoclonal gammopathy (i.e., a PCD or other conditions such as hematological disorders associated with a monoclonal band). Twenty-seven percent of patients had a PCD, including 2.1 percent with LCMM, 6.6 percent with MM, 0.6 percent with amyloidosis, and 13.6 percent with MGUS. Use of the SFLC assay plus SPEP resulted in a diagnostic sensitivity of 0.96 and specificity of 0.78; whereas SPEP alone had a sensitivity of 0.87 and specificity of 0.98.

Vermeersch 2008\textsuperscript{33} explored the use of the SFLC assay in 833 consecutive patients suspected of having a PCD and compared various tests and combinations of tests (Table 4). Three percent of patients had a malignant PCD and 19 percent had MGUS. The highest diagnostic sensitivity, 0.94, was achieved by using the SFLC assay plus SIFE. SIFE alone had a sensitivity of 0.92. The SFLC assay plus SPEP (with SIFE performed only if SPEP was positive, for confirmation) achieved a sensitivity of 0.82, whereas SPEP plus SIFE without the SFLC assay had a sensitivity of 0.79. SPEP plus SIFE had a specificity of 1.00, as did SIFE alone; the SFLC assay plus either SPEP (with SIFE for confirmation) or SIFE had a specificity of 0.97.

Summary

Three retrospective studies evaluated the SFLC assay in combination with traditional tests in undiagnosed patients suspected of having a PCD. 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). The statistical significance of the increase in sensitivity was not addressed in any of the studies; 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. We rated the strength of evidence to evaluate the effect of adding SFLC testing to traditional testing on diagnostic performance as insufficient.
Table 2. Characteristics of studies addressing KQ1

<table>
<thead>
<tr>
<th>Author Year [PMID]</th>
<th>Index Test/ Comparator Test</th>
<th>Sample Size</th>
<th>Funding</th>
<th>Enrollment Period</th>
<th>Prospective Study?</th>
<th>Diagnosis Documented in Medical Records</th>
<th>Quality Grade and Issues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abadie 2006³¹ [16682511]</td>
<td>SFLC+SPEP, SPEP</td>
<td>312</td>
<td>Kit/reagents provided by industry</td>
<td>2004–2005</td>
<td>No</td>
<td>Yes</td>
<td>B (no CI provided, consecutive sampling, no major biases)</td>
</tr>
<tr>
<td>Piehler 2008³² [18801937]</td>
<td>SFLC+SPEP, SPEP</td>
<td>489</td>
<td>nd</td>
<td>2005–2006</td>
<td>No</td>
<td>Yes</td>
<td>B (no CI provided, consecutive recruitment, no major biases)</td>
</tr>
<tr>
<td>Vermeersch 2008³³ [18729849]</td>
<td>SFLC+SIFE, SFLC+SIFE+SPEP, SIFE, SIFE+SPEP</td>
<td>833</td>
<td>None</td>
<td>2004–2006</td>
<td>No</td>
<td>Yes</td>
<td>B (no CI provided, well-described sample, no major biases)</td>
</tr>
</tbody>
</table>

CI=confidence interval, IFE=immunofixation electrophoresis, KQ=Key Question, nd=no data, SFLC=serum free light chain [note this can refer to the light chain itself or the assay], SIFE=serum immunofixation electrophoresis; SPEP=serum protein electrophoresis, UIFE=urine immunofixation electrophoresis, UPEP=urine protein electrophoresis.

Table 3. Characteristics of patients in studies addressing KQ1

<table>
<thead>
<tr>
<th>Author Year [PMID]</th>
<th>Sample Size</th>
<th>Enrollment Method</th>
<th>Diagnosed Before Study?</th>
<th>PCD Prevalence</th>
<th>Age (yr)</th>
<th>Percent Male</th>
<th>Treated?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abadie 2006³¹ [16682511]</td>
<td>312</td>
<td>Consecutive</td>
<td>No</td>
<td>Malignant PCD, 15%</td>
<td>67 (mean)</td>
<td>97</td>
<td>nd</td>
</tr>
<tr>
<td>Piehler 2008³² [18801937]</td>
<td>489</td>
<td>Selection of those with SPEP testing results</td>
<td>No</td>
<td>Any PCD, 27% LCMM, 2.1% MM, 6.6% AL amyloidosis, 0.6% MGUS, 13.6%</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td>Vermeersch 2008³³ [18729849]</td>
<td>833</td>
<td>Consecutive</td>
<td>No</td>
<td>Malignant PCD, 3%; MGUS, 19%</td>
<td>nd</td>
<td>nd</td>
<td>NA</td>
</tr>
</tbody>
</table>

AL amyloidosis=systemic amyloidosis in which amyloid [A] proteins derived from immunoglobulin light chains [L] are deposited in tissue, KQ=Key Question, LCMM=light chain myeloma, MGUS=monoclonal gammopathy of undetermined significance, MM=multiple myeloma, NA=not applicable, nd=no data, PCD=plasma cell dyscrasia.
### Table 4. Results of studies addressing KQ1

<table>
<thead>
<tr>
<th>Author Year (PMID)</th>
<th>Sample Size</th>
<th>Diagnosis</th>
<th>Index Test</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Other Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abadie 2006(^\text{31}) [16682511]</td>
<td>312</td>
<td>PCD</td>
<td>SFLC</td>
<td>0.88 (0.75, 0.97)</td>
<td>0.98 (0.96, 0.99)</td>
<td>Considered MGUS as false positive</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SPEP</td>
<td>0.64 (0.49, 0.77)</td>
<td>0.81 (0.76, 0.86)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SFLC+SPEP</td>
<td>1.00 (nd)</td>
<td>0.99 (nd)</td>
<td></td>
</tr>
<tr>
<td>Piehler 2008(^\text{32}) [18801937]</td>
<td>332</td>
<td>PCD</td>
<td>SFLC</td>
<td>0.66 (nd)</td>
<td>0.78 (nd)</td>
<td>Specificity was affected by SFLC assay positivity in patients with other hematological diagnosis or decreased kidney function</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SPEP</td>
<td>0.87 (nd)</td>
<td>0.98 (nd)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SFLC+SPEP</td>
<td>0.96 (nd)</td>
<td>0.78 (nd)</td>
<td></td>
</tr>
<tr>
<td>Vermeersch 2008(^\text{33}) [18729849]</td>
<td>833</td>
<td>Monoclonal gammopathy(^\ast)</td>
<td>SFLC</td>
<td>0.37 (nd)</td>
<td>0.97 (nd)</td>
<td>Missed 3 MM, 1 plasmacytoma, 112 MGUS cases</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SIFE</td>
<td>0.92 (nd)</td>
<td>1.00 (nd)</td>
<td>Missed 2 MGUS cases</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SFLC+SIFE</td>
<td>0.94 (nd)</td>
<td>0.97 (nd)</td>
<td>Missed 1 MGUS cases</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SPEP (+SIFE for confirmation)</td>
<td>0.79 (nd)</td>
<td>1.00 (nd)</td>
<td>Missed 1 MM, 1 AL amyloidosis, 1 plasmacytoma, 26 MGUS cases</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SFLC+ SPEP (+SIFE for confirmation)</td>
<td>0.82 (nd)</td>
<td>0.97 (nd)</td>
<td>Missed 1 plasmacytoma, 23 MGUS cases</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>UIFE+ SPEP (+SIFE for confirmation)</td>
<td>0.82 (nd)</td>
<td>1.00 (nd)</td>
<td>Missed 24 MGUS cases</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>UIFE</td>
<td>0.92 (nd)</td>
<td>1.00 (nd)</td>
<td>Missed 2 MGUS cases</td>
</tr>
</tbody>
</table>

**Notes:**
- AL amyloidosis=systemic amyloidosis in which amyloid [A] proteins derived from immunoglobulin light chains [L] are deposited in tissue; CI=confidence interval; KQ=Key Question, LCMM=light chain myeloma, MGUS=monoclonal gammopathy of undetermined significance, MM=multiple myeloma, nd=no data, SFLC=serum free light chain, SIFE=serum immunofixation, SPEP=serum protein electrophoresis, UIFE=urine immunofixation electrophoresis.
- *Monoclonal gammopathy includes PCDs as well as other conditions such as hematological disorders associated with a monoclonal band.
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?

- 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)?

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)?

Results

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 in patients with AL amyloidosis and eight in patients with 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 443 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 SFLC assay in patients with AL amyloidosis and reported the use of SFLC assay in evaluating treatment response and predicting prognosis: Kumar 2011, Lachmann 2003, and Sanchoiarawa 2005. 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.

The sample sizes were 66, 262, and 443 patients (Tables 5–7). Followup times were 21 months to 5 years. Kumar 2011 and Lachmann 2003 reported industry-associated funding or authorship. All three studies reported explicit diagnostic criteria. Lachmann 2003 reported enrolling referred patients; the other two studies did not describe the enrollment method. The
median age of study participants was 54 to 64 years and, in the two studies with data on patient sex, one had 61 percent men and the other, 63 percent men. All three studies were rated as quality C; none of the three studies performed direct statistical comparisons of the relative strength of prediction, providing only unadjusted estimates for each predictor.

All three studies showed that patients with greater reductions in abnormal SFLC concentrations (a >50 percent reduction\textsuperscript{40} or >90 percent reduction,\textsuperscript{13,38} vs. lesser reductions) after treatment (either chemotherapy or stem-cell transplantation) had better survival outcomes. Although Kumar 2011\textsuperscript{38} did not find quantitative paraprotein concentrations to be a good predictor (unlike SFLC concentrations), Lachmann 2003\textsuperscript{40} found the paraprotein concentration to be significantly related to survival; however, the relationship seemed to be weaker than that of SFLC reduction to survival. In Kumar 2011, some patients with PCD did not have “measurable disease,” (i.e., they did not have elevated SFLC concentrations before treatment), which precluded use of the SFLC assay as a marker of disease and treatment response, limiting the assay’s utility. Sanchorwala 2005\textsuperscript{13} found that a reduction in SFLC concentration by more than 90 percent and achievement of a complete response were both predictive of a lower mortality and both provided independent predictive information.

Summary for AL Amyloidosis

Although the three studies reported that 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. All three studies were rated as quality C, owing to 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 compared with traditional tests for assessment of treatment response and outcome is inconclusive. The strength of evidence underlying this comparison was therefore rated as insufficient.

Patients With MM

Eight studies\textsuperscript{34-37,39,41,42,43} enrolled patients with MM and compared the use of SFLC assay and other traditional tests in evaluating treatment response and predicting prognosis (Tables 5–7). Six of the eight—Dispenzieri 2008,\textsuperscript{34} Giarin 2009,\textsuperscript{35} Khoriaty 2010,\textsuperscript{36} van Rhee 2007,\textsuperscript{41} Kyrtsonis 2007,\textsuperscript{39} and Paiva 2011\textsuperscript{42}—were retrospective analyses of cohorts; one study, Dytfeld 2011,\textsuperscript{43} was prospective; and study design was not specified in the remaining study, Kroger 2010.\textsuperscript{37} Sample size ranged from 40 to 303, and median followup duration was 3 months to 13 years. Study quality was graded as B in three of the eight studies, owing to retrospective designs without adjustments for potential confounders,\textsuperscript{34,35,41} and C in the other five studies, owing to small sample sizes, limited information about study design, and/or potential selection bias.\textsuperscript{36,37,39,42,43} 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\textsuperscript{34,36,37,42} addressed the use of SFLC assay in the assessment of treatment response and one study\textsuperscript{43} addressed the prediction of treatment response. The traditional test comparators that were also 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]).

Of the four studies that used SFLC 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.42 Another study, of B quality, found that after 2 months of therapy, treatment response was achieved by 23 percent of 139 patients using the paraprotein definition, compared with 62 percent using the SFLC definition.34 In a C-quality study, the majority (27 of 43 patients) achieved treatment response as defined by both M protein criteria and SFLC criteria at the same time; SFLC response occurred earlier than M protein response in eight other patients.36 A fourth study37 of unclear design reported an abnormal SFLC ratio before relapse and a positive IFE test in 9 of a subgroup of 10 patients. The quality of this study was rated as C because of the limited information about study design, SFLC response definitions, and results.

Only 1 study, of C quality, reported data on prediction of treatment response.43 Patients received VDD (bortezomib, pegylated liposomal doxorubicin, and dexamethasone) treatment for newly diagnosed, histologically confirmed MM. An SFLC and M protein–based prognostic model predicted that either a 90 percent or greater reduction in serum M protein level or involved SFLC level, or 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 VDD treatment. Sensitivity increased to 96 percent after three cycles of VDD treatment. Taking into account the heterogeneity of MM and its spectrum of M protein presentations, measurements of both the involved SFLC and M protein were needed to fully monitor response to treatment. Neither the rate of decline in M protein or 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**

Two studies examined the relationship of baseline SFLC concentrations and survival; one followed 303 patients for 21 months and included concomitant evaluation of the predictive ability of traditional testing (in the form of measurement of baseline concentrations of serum and urine M protein),41 whereas the other followed 94 patients for 33 months and incorporated the clinical Durie–Salmon staging system and the International Staging System (ISS).39 In the former study, of B quality, the top tertile of SFLC concentrations (>75 mg/dL) were considered the risk category,41 whereas in the latter study, of C quality, patients were stratified according to whether the SFLC ratio was above or below the median (with the ratio calculated using the involved SFLC in the numerator, for a monotonic distribution).39 In both studies, patients with higher SFLC concentrations or ratio had significantly lower survival rates than did patients with lower SFLC concentrations or ratio. The former study did not find serum or urine M protein concentrations to be predictive of survival and reported significantly poorer overall and event-free survival rates among patients with a baseline SFLC level of greater than 75 mg/dL (vs. ≤75 mg/dL; p=0.016 and p=0.008, respectively).41 The latter study reported that while Durie–Salmon and ISS staging were independent predictors (both p<0.0001), an abnormal SFLC ratio was also significantly associated with 3- and 5-year disease-specific survival rates (p=0.0001).39
**Relationship Between Post-Treatment SFLC Measurements and Survival**

Three studies examined the relationship between post-treatment SFLC ratios and survival.\(^3\,4\,4\) One study\(^4\) of C quality analyzed the SFLC ratios after induction therapy among a subset of 102 patients enrolled in a previous trial. 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.\(^4\)

A second study,\(^3\) of B quality, analyzed immunofixation results and SFLC ratios after stem-cell transplantation among 202 patients. 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.\(^3\) However, this study also reported that 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 \(p=\text{NS}\).

In a third study of 303 patients,\(^4\) also of B quality, patients with a percent reduction in SFLC concentration in the top tertile after transplantation had nearly twice the risk of death—that is, hazard ratios greater than 2 for overall or event-free survival—than patients with less of a percent reduction (after adjustment for serum lactate dehydrogenase concentration and cytogenetic abnormalities), despite a paradoxically better response to induction therapy. However, there was no significant relationship between the tertiles of percent 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 (75 percent) were retrospective cohort studies, and only 3 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 the strength of evidence was rated as insufficient for comparisons with traditional testing in patients with MM.
Table 5. Characteristics of studies addressing KQ4

<table>
<thead>
<tr>
<th>Author Year [PMID]</th>
<th>Index Test/ Comparator Test</th>
<th>Sample Size</th>
<th>Funding</th>
<th>Study Design</th>
<th>Enrollment Period</th>
<th>Followup Duration</th>
<th>Diagnostic Criteria</th>
<th>Quality Grade and Issues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kumar, 2011 [21328431]</td>
<td>Post-treatment dFLC</td>
<td>443</td>
<td>Government, industry</td>
<td>Retrospective</td>
<td>nd</td>
<td>72 mo</td>
<td>Biopsy-proven AL amyloidosis</td>
<td>C (retrospective, extreme selection/spectrum bias)</td>
</tr>
<tr>
<td></td>
<td>Post-treatment quantitative M protein concentrations</td>
<td>Cohort I: 347</td>
<td>Cohort II: 96</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Post-treatment quantitative paraprotein concentrations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hematological complete response (defined by EBMT, includes M protein response)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Author Year [PMID]</td>
<td>Index Test/ Comparator Test</td>
<td>Sample Size</td>
<td>Funding</td>
<td>Study Design</td>
<td>Enrollment Period</td>
<td>Followup Duration</td>
<td>Diagnostic Criteria</td>
<td>Quality Grade and Issues</td>
</tr>
<tr>
<td>-------------------------</td>
<td>-----------------------------</td>
<td>-------------</td>
<td>------------</td>
<td>--------------</td>
<td>-------------------</td>
<td>-------------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>Dispenzieri, 2008[34]</td>
<td>SFLC response</td>
<td>399</td>
<td>Government</td>
<td>Retrospective</td>
<td>1988–1992</td>
<td>13 yr</td>
<td>M protein ≥10 g/L or urine monoclonal FLC &gt;200 mg in 24 hr or serially measurable soft tissue plasmacytoma or bone marrow plasmacytosis ≥20%</td>
<td>B (retrospective without adjustment)</td>
</tr>
<tr>
<td></td>
<td>SPEP, UPEP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dytfeld, 2011[43]</td>
<td>Percent reduction in involved FLC concentrations</td>
<td>40</td>
<td>Industry, author(s) employed by industry, manuscript reviewed by industry</td>
<td>Prospective</td>
<td>2005–2007</td>
<td>45 mo</td>
<td>Histologically confirmed diagnosis of MM</td>
<td>C (small sample size, sample not uniformly treated)</td>
</tr>
<tr>
<td></td>
<td>Normalization of FLC ratio</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Percent reduction in serum and urine M protein concentrations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Khoriaty, 2010[36]</td>
<td>SFLC concentrations</td>
<td>89 (43 with evaluable disease)</td>
<td>nd</td>
<td>Retrospective</td>
<td>2004–2006</td>
<td>40 mo</td>
<td>nd</td>
<td>C (small sample size, retrospective without adjustment)</td>
</tr>
<tr>
<td></td>
<td>IMWG criteria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Author</td>
<td>Year</td>
<td>PMID</td>
<td>Index Test/ Comparator Test</td>
<td>Sample Size</td>
<td>Funding</td>
<td>Study Design</td>
<td>Enrollment Period</td>
<td>Followup Duration</td>
</tr>
<tr>
<td>-----------------</td>
<td>------</td>
<td>--------------</td>
<td>-------------------------------------------------</td>
<td>-------------</td>
<td>--------------</td>
<td>--------------</td>
<td>-------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Kroger, 2010</td>
<td>37</td>
<td>[2043663]</td>
<td>SFLC response</td>
<td>52</td>
<td>nd</td>
<td>Unclear</td>
<td>2003–2008</td>
<td>3 mo</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SIFE or UIFE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paiva, 2011</td>
<td>42</td>
<td>[21402611]</td>
<td>SFLC ratio normalization (stringent complete response)</td>
<td>102</td>
<td>Nonprofit foundation</td>
<td>Retrospective</td>
<td>nd</td>
<td>32 mo</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Immunophenotypic response</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kyrtsonis, 2007</td>
<td>39</td>
<td>[17408464]</td>
<td>SFLC ratio</td>
<td>94</td>
<td>Nonprofit foundation</td>
<td>Retrospective</td>
<td>nd</td>
<td>33 mo</td>
</tr>
<tr>
<td>van Rhee, 2007</td>
<td>41</td>
<td>[17416735]</td>
<td>Baseline SFLC concentrations</td>
<td>303</td>
<td>Government</td>
<td>Retrospective</td>
<td>nd</td>
<td>21 mo</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Baseline concentrations of serum and urine M protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5. Characteristics of studies addressing KQ4 (continued)

<table>
<thead>
<tr>
<th>Author Year</th>
<th>Index Test/ Comparator Test</th>
<th>Sample Size</th>
<th>Funding</th>
<th>Study Design</th>
<th>Enrollment Period</th>
<th>Followup Duration</th>
<th>Diagnostic Criteria</th>
<th>Quality Grade and Issues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paiva, 2011 [21402611]</td>
<td>SFLC ratio normalization (stringent complete response) Immunophenotypic response</td>
<td>102</td>
<td>Nonprofit foundation</td>
<td>Retrospective</td>
<td>nd</td>
<td>32 mo</td>
<td>nd</td>
<td>C (retropective without adjustment, potential selection bias because inclusion was based on availability of serum samples)</td>
</tr>
<tr>
<td>van Rhee, 2007 [17416735]</td>
<td>SFLC response tertiles Percent reduction of serum and urine M protein concentrations</td>
<td>303</td>
<td>Government</td>
<td>Retrospective</td>
<td>nd</td>
<td>21 mo</td>
<td>nd</td>
<td>B (retropective with adjustment)</td>
</tr>
</tbody>
</table>

AL amyloidosis = systemic amyloidosis in which amyloid [A] proteins derived from immunoglobulin light chains [L] are deposited in tissue, dFLC = difference in the levels of the involved free light chain (FLC, either kappa or gamma) and the other (uninvolved FLC), EBMT = European Group for Blood and Bone Marrow Transplant, IMWG = International Myeloma Working Group, ISS = International Staging System, KQ = Key Question, MM = multiple myeloma, mo = months, nd = no data, PCD = plasma cell dyscrasia, SCT = stem cell transplantation, SFLC = serum free light chain, SIFE = serum immunofixation electrophoresis, SPEP = serum protein electrophoresis, UIFE = urine protein electrophoresis, UPEP = urine protein electrophoresis, yr = years.

*ISS classification incorporates concentrations of serum albumin and β2 microglobulin. The Durie–Salmon staging system classification incorporates concentrations of serum and urinary paraproteins.
<table>
<thead>
<tr>
<th>Author Year [PMID]</th>
<th>Enrollment Method</th>
<th>Median Age (yr)</th>
<th>Percent Male</th>
<th>Population Description and Inclusion Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kumar, 2011 [21328431]</td>
<td>nd</td>
<td>Cohort I: 58, Cohort II: 64</td>
<td>nd</td>
<td>AL amyloidosis, with 347 patients receiving autologous SCT and 96 receiving melphalan and dexamethasone</td>
</tr>
<tr>
<td>Lachmann, 2003 [12823348]</td>
<td>Referred patients</td>
<td>54–64</td>
<td>nd</td>
<td>Systemic AL amyloidosis, no prior chemotherapy, excluding those with concurrent MM or other malignant B-cell dyscrasias</td>
</tr>
<tr>
<td>Sanchorawala, 2005 [16044137]</td>
<td>nd</td>
<td>60</td>
<td>63</td>
<td>Receipt of high-dose intravenous melphalan and autologous SCT</td>
</tr>
<tr>
<td>Dispenzieri, 2008 [18364469]</td>
<td>Patients enrolled in a previous treatment trial (E9486)</td>
<td>63</td>
<td>65</td>
<td>Diagnosed with MM, enrollment in a previously published treatment trial, measurable disease in absence of treatment, pre- and post-treatment serum samples available</td>
</tr>
<tr>
<td>Dytfeld, 2011 [21699382]</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>Diagnosed with MM, receiving VDD treatment for newly diagnosed, histologically confirmed MM</td>
</tr>
<tr>
<td>Khoriaty, 2010 [20223721]</td>
<td>nd</td>
<td>61</td>
<td>65</td>
<td>Diagnosed with MM (relapsed or newly diagnosed), treatment at the Cleveland Clinic Taussig Cancer Institute, enrolled in other trials, with SFLC measurement every 4 weeks from April 2004 to December 2006 (Only 43 patients [48%] had evaluable disease)</td>
</tr>
<tr>
<td>Kroger, 2010 [2043663]</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>Diagnosed with MM, complete response between January 2003 and December 2008 for at least 3 mo, negative SIFE or UIFE test</td>
</tr>
<tr>
<td>Paiva, 2011 [21402611]</td>
<td>Patients enrolled in a previous trial (GEM05&gt;65y PETHEMA/GEM trial)</td>
<td>72</td>
<td>44</td>
<td>Diagnosed with MM, enrolled in a previous treatment trial, who achieved at least a partial response with 70% reduction in M protein after the six planned induction cycles; patients with available serum samples</td>
</tr>
<tr>
<td>Kyrtsonis, 2007 [17408464]</td>
<td>nd</td>
<td>32% &gt;65 yr</td>
<td>45</td>
<td>Diagnosed with MM, with or without treatment</td>
</tr>
<tr>
<td>van Rhee, 2007 [17416735]</td>
<td>Patients enrolled in a previous trial (Total Therapy 3)</td>
<td>nd</td>
<td>64</td>
<td>Newly diagnosed MM, participation in a tandem autotransplantation trial</td>
</tr>
</tbody>
</table>
Table 6. Characteristics of patients in studies addressing KQ4 (continued)

<table>
<thead>
<tr>
<th>Author Year [PMID]</th>
<th>Enrollment Method</th>
<th>Median Age (yr)</th>
<th>Percent Male</th>
<th>Population Description and Inclusion Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Giarin, 200935 [19520760]</td>
<td>nd</td>
<td>56</td>
<td>55</td>
<td>Newly diagnosed MM between July 1995 and February 2006, receipt of autologous or autologous and allogeneic SCT</td>
</tr>
<tr>
<td>Paiva, 201142 [21402611]</td>
<td>Patients enrolled in a previous trial (GEM05&gt;65y PETHEMA/GEM trial)</td>
<td>72</td>
<td>44</td>
<td>Diagnosed with MM, enrolled in a previous treatment trial, achievement of at least a partial response with 70% reduction in M protein after the six planned induction cycles; patients with available serum samples</td>
</tr>
<tr>
<td>van Rhee, 200741 [17416735]</td>
<td>Patients enrolled in a previous trial (Total Therapy 3)</td>
<td>nd</td>
<td>64</td>
<td>Newly diagnosed MM, participation in a tandem autotransplantation trial</td>
</tr>
</tbody>
</table>

AL amyloidosis=systemic amyloidosis in which amyloid [A] proteins derived from immunoglobulin light chains [L] are deposited in tissue, KQ=Key Question, MM=multiple myeloma, nd=no data, SCT=stem cell transplantation, SFLC=serum free light chain, SIFE=serum immunofixation electrophoresis,.UIFE=urine immunofixation electrophoresis, VDD=bortezomib, pegylated liposomal doxorubicin, and dexamethasone, yr=years.
<table>
<thead>
<tr>
<th>Author Year (PMID)</th>
<th>Sample Size</th>
<th>Index Test/Comparator Test</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kumar, 2011 [21328431]</td>
<td>443 Cohort I: 347 Cohort II: 96</td>
<td>Post-treatment dFLC, Post-treatment quantitative M protein concentrations</td>
<td>dFLC (vs. SPEP) significantly affected overall survival ($p&lt;0.0001$) ≤90% reduction in dFLC best predicted survival at 3 or 5 yr; median overall survival was not reached among those with a ≤90% reduction but was 37.4 months with &gt;90% decrease ($p&lt;0.001$)</td>
</tr>
<tr>
<td>Lachmann, 2003 [12823348]</td>
<td>262</td>
<td>Post-treatment SFLC concentrations, Post-treatment quantitative paraprotein concentrations</td>
<td>86 patients with abnormal FLC concentration falling &gt;50% after chemotherapy had 88% 5-year survival vs. only 39% among those with lesser reduction ($p&lt;0.0001$) Amyloidogenic FLC reduction &gt;50% associated with survival benefit, regardless of type of chemotherapy Amyloid load correlated with changes in SFLC concentration ($p&lt;0.0001$). Among 73 patients with serially quantifiable serum paraprotein, survival was better in those whose concentration fell by &gt;50% vs. those whose fell by ≤50% ($p&lt;0.05$).</td>
</tr>
<tr>
<td>Sanchorawala, 2005 [16044137]</td>
<td>66</td>
<td>Post-treatment SFLC concentrations</td>
<td>Death: % (number/total number) Complete vs. noncomplete response: 4% (1/27) vs. 18% (7/39) ($p$-value not available) FLC response &gt;90% vs. ≤90%: 6% (2/35) vs. 19% (6/31) ($p$ value not available) Clinical improvement: % (number/total number) Complete vs. noncomplete response: 96% (26/27) vs. 67% (26/39) ($p=0.047$) FLC response &gt;90% vs. ≤90%: 97% (34/35) vs. 58% (18/31) ($p$ value not available) FLC response and measures of hematological response complementary</td>
</tr>
<tr>
<td>Dispenzieri, 2008 [18364469]</td>
<td>139</td>
<td>SFLC response</td>
<td>After 2 months of therapy, 23% had achieved a paraprotein response in SPEP and/or UPEP compared with 62% who achieved an FLC response 85% of FLC responders developed overall objective response vs. 51% of FLC nonresponders ($p&lt;0.001$) Prediction of ECOG overall objective response status*: 2-mo FLC response: sensitivity 69%, specificity 73%, risk 0.3; 2-mo paraprotein response: sensitivity 34%, specificity 98%, risk 0.5; $p&lt;0.001$</td>
</tr>
<tr>
<td>Dytfeld, 2011 [21699382]</td>
<td>40</td>
<td>Percent reduction in involved FLC concentrations, Normalization of SFLC ratio</td>
<td>A novel FLC and M protein–based prognostic model predicts that ≥90% reduction of serum M protein or ≥90% reduction of involved FLC or normalization of SFLC ratio predicted ≥VGPR with 92% sensitivity and 93% specificity after two cycles of treatment with VDD, with sensitivity increasing to 96% after three cycles of treatment. Neither the rate of M protein decline nor the decline of involved FLC independently predicted VGPR at the end of six cycles of VDD (at 90% sensitivity and specificity). When the involved was replaced by urine M protein in the predictive model, sensitivity, specificity, and predictive value were all &lt;90%.</td>
</tr>
</tbody>
</table>

**MM**

Assessment and Prediction of Treatment Response

<table>
<thead>
<tr>
<th>Author Year (PMID)</th>
<th>Sample Size</th>
<th>Index Test/Comparator Test</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dispenzieri, 2008 [18364469]</td>
<td>139</td>
<td>SPEP, UPEP</td>
<td>After 2 months of therapy, 23% had achieved a paraprotein response in SPEP and/or UPEP compared with 62% who achieved an FLC response 85% of FLC responders developed overall objective response vs. 51% of FLC nonresponders ($p&lt;0.001$) Prediction of ECOG overall objective response status*: 2-mo FLC response: sensitivity 69%, specificity 73%, risk 0.3; 2-mo paraprotein response: sensitivity 34%, specificity 98%, risk 0.5; $p&lt;0.001$</td>
</tr>
</tbody>
</table>

---

**AL Amyloidosis**

Table 7. Results of studies addressing KQ4
Table 7. Results of studies addressing KQ4 (continued)

<table>
<thead>
<tr>
<th>Author Year [PMID]</th>
<th>Sample Size</th>
<th>Index Test/Comparator Test</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>MM (continued)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Assessment and Prediction of Treatment Response (continued)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Khoriaty, 2010[36] [20223721] | 43 (those with evaluable disease) | SFLC ratio | For SFLC assay prediction of response to treatment (95% CI):  
Sensitivity: 81% (51 to 94%)  
Specificity: 83% (65 to 92%)  
PPV: 64% (38 to 83%)  
NPV: 92% (68 to 98%) |
| | | IMWG criteria[27] | For SFLC assay prediction of progression (95% CI):  
Sensitivity: 93% (68 to 98%)  
Specificity: 80% (62 to 91%)  
PPV: 72% (49 to 87%)  
NPV: 95% (78 to 99%) |
| Kroger, 2010[37] [2043663] | 52 | SFLC | 51/52 (98%) patients had normal SFLC ratio  
In the subgroup of 10 patients who relapsed, 9 had abnormal SFLC ratio before having a positive IFE test |
| | | SIFE or UIFE |  |
| Paiva, 2011[42] [21402611] | 102 | SFLC ratio | 22 patients had discordant results of treatment response between SFLC ratio definition and immunophenotypic response definition:  
6 had abnormal SFLC ratio but achieved immunophenotypic response  
5 had normal SFLC ratio but no complete response because immunofixation was positive  
11 had normal SFLC ratio and negative immunofixation but no immunophenotypic response |
| | | Immunophenotypic response |  |
| **Relationship Between Baseline SFLC Ratios and Survival** | | |  |
| Kyrtonis, 2007[39] [17408464] | 94 | SFLC ratio | 3- and 5-year disease-specific survival rates, 94% and 82%, respectively, with SFLC ratio below median (vs. 58% and 30%, respectively, with SFLC ratio above the median; p=0.0001)  
Durie–Salmon and ISS stages 1–3, Durie–Salmon stages I–III** |
### Table 7. Results of studies addressing KQ4 (continued)

<table>
<thead>
<tr>
<th>Author Year [PMID]</th>
<th>Sample Size</th>
<th>Index Test/ Comparator Test</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Relationship Between Baseline SFLC Ratios and Survival (continued)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>van Rhee, 2007[41] [17416735]</td>
<td>303</td>
<td>Baseline SFLC concentrations</td>
<td>Rate of near-complete response to induction therapy higher among patients with baseline SFLC &gt;75 mg/dL than patients with baseline SFLC ≤75 mg/dL (37% vs. 20%, p=0.002).</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Baseline concentrations of serum and urine M protein</td>
<td>Adjusted HR (95% CI) for overall survival: Baseline SFLC &gt;75 (vs. ≤75) mg/dL: 2.43 (1.18 to 5.01), p=0.016</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Adjusted HR (95% CI) for event-free survival: Baseline SFLC &gt;75 (vs. ≤75) mg/dL: 2.40 (1.26 to 4.57), p=0.008</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Baseline concentrations of standard serum and urine M protein did not identify prognostic subgroups</td>
</tr>
<tr>
<td><strong>Relationship Between Post-Treatment SFLC Ratios and Survival</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Giarin, 2009[35] [19520760]</td>
<td>203</td>
<td>SFLC ratio</td>
<td>3 mo after SCT, overall and event-free survival did not differ significantly between patients with and those without normal SFLC ratio or between patients with and those without normal (negative) SIFE test.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total kappa/lambda ratio, SIFE</td>
<td>Longer event-free but not overall survival significantly associated with normal SFLC ratio at 3 mo post SCT (HR, 0.68; 95% CI, 0.50 to 0.93, p=0.02)</td>
</tr>
<tr>
<td>Paiva, 2011[42] [21402611]</td>
<td>102</td>
<td>SFLC ratio normalization (stringent complete response)</td>
<td>Among 44 patients with negative immunofixation (conventional complete response), rate of 3-year progression-free survival did not differ between patients with normal SFLC ratio and patients with abnormal SFLC ratio (69% vs. 64%, p=0.4). Similarly, time to progression and overall survival did not differ between groups (p=0.2 and p=0.9, respectively).</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Immunophenotypic response</td>
<td>Among 78 patients with positive immunofixation, rate of 3-year progression-free survival, time to progression, and overall survival did not significantly differ between patients with normal and abnormal SFLC ratios (p=0.2, p=0.1, p=0.3, respectively).</td>
</tr>
</tbody>
</table>
### Table 7. Results of studies addressing KQ4 (continued)

<table>
<thead>
<tr>
<th>Author Year [PMID]</th>
<th>Sample Size</th>
<th>Index Test/ Comparator Test</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>van Rhee, 2007[41] [17416735]</td>
<td>303</td>
<td>SFLC response tertiles, Percent reduction of serum and urine M protein concentrations</td>
<td>Rate of near-complete response to induction therapy higher among patients with baseline SFLC &gt;75 mg/dL than patients with baseline SFLC ≤75 mg/dL (37% vs. 20%, p=0.002).</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Adjusted HR (95% CI) for overall survival:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Top tertile (vs. lower two tertiles) in percent SFLC reduction after cycle 2: 2.15 (1.03 to 4.47), p=0.041</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Top tertile (vs. lower two tertiles) in percent SFLC reduction after transplantation: 2.24 (1.03 to 4.87), p=0.042</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Baseline SFLC &gt;75 (vs. ≤75) mg/dL: 2.43 (1.18 to 5.01), p=0.016</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Adjusted HR (95% CI) for event-free survival poorer with higher percent reduction in SFLC level:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Top tertile (vs. lower two tertiles) in percent SFLC reduction after cycle 2: 1.96 (1.03 to 3.74), p=0.041</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Top tertile (vs. lower two tertiles) in percent SFLC reduction after transplantation: 2.01 (1.02 to 3.97), p=0.045</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reductions in serum and urine M protein values not significantly associated with overall or event-free survival</td>
</tr>
</tbody>
</table>

AL amyloidosis = systemic amyloidosis in which amyloid [A] proteins derived from immunoglobulin light chains [L] are deposited in tissue, CI = confidence interval, dFLC = difference in the levels of the involved free light chain (FLC, either kappa or gamma) and the other (uninvolved FLC), EBMT = European Group for Blood and Bone Marrow Transplant, ECOG = Eastern Cooperative Oncology Group, FLC = free light chain, HR = hazard ratio, IFE = immunofixation electrophoresis, IMWG = International Myeloma Working Group, ISS = International Staging System, KQ = Key Question, MM = multiple myeloma, mo = months, NPV = negative predictive value, PPV = positive predictive value, SCT = stem cell transplantation, SFLC = serum free light chain, SIFE = serum immunofixation electrophoresis, SPEP = serum protein electrophoresis, UIFE = urine immunofixation electrophoresis, UPEP = urine protein electrophoresis, VDD = bortezomib, pegylated liposomal doxorubicin, VGPR = very good partial response, yr = years.

* Standard ECOG response criteria are as follows: 50 percent decrease in serum M protein or, in patients lacking a serum M protein measurement, a 90 percent decrease in 24-hour urine M protein.

** ISS classification incorporates concentrations of serum albumin and β2 microglobulin. The Durie–Salmon staging system classification incorporates concentrations of serum and urinary paraproteins.
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)?

Results

We identified one C-quality retrospective cohort study assessing the need for bone marrow examination, with the SFLC assay used to define the completeness of response to treatment: Chee 200947 (Tables 8–10). 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.22,27

Chee 2009 enrolled 92 patients with MM who achieved negative SIFE and UIFE tests after therapy and had a bone marrow aspirate or biopsy performed within 30 days before or after those tests. A subgroup of 29 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. Fourteen percent of patients with a negative IFE test had more than 5 percent plasma cells in bone marrow, as did 10 percent of patients with a normal SFLC ratio. Among patients with IFE-negative status, those with less than 5 percent plasma cells in the marrow had improved overall survival compared with those with 5 percent or more plasma cells (6.2 years vs. 2.3 years, respectively; p <0.01).

Summary

A single study was found that addressed whether IFE or SFLC testing would reduce the need for other diagnostic tests such as bone marrow examination; the authors concluded that it was not possible to eliminate such tests. Owing to the preliminary nature of the data, we rated the strength of evidence as insufficient for addressing this question.
Table 8. Characteristics of studies addressing KQ5

<table>
<thead>
<tr>
<th>Author Year [PMID]</th>
<th>Index Test</th>
<th>Funding</th>
<th>Study Design</th>
<th>Enrollment Period</th>
<th>Followup Duration</th>
<th>Diagnostic Criteria</th>
<th>Quality Grade and Issues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chee 2009,47 [19641191]</td>
<td>SFLC ratio</td>
<td>Government</td>
<td>Retrospective cohort</td>
<td>nd</td>
<td>1995–??</td>
<td>MM, measurable M protein concentrations at baseline (serum M protein ≤1 g/dL or urine M protein ≤0.2 g/day), and since start of study, negative SIFE and UIFE with concomitant bone marrow aspirate or biopsy and normal SFLC ratio (with all tests performed within 30 days of each other)</td>
<td>C (retrospective, small convenience sample)</td>
</tr>
</tbody>
</table>

KQ=Key Question, MM=multiple myeloma, nd=no data, SFLC=serum free light chain, SIFE=serum immunofixation electrophoresis, UIFE=urine immunofixation electrophoresis.

Table 9. Characteristics of patients in studies addressing KQ5

<table>
<thead>
<tr>
<th>Author Year [PMID]</th>
<th>Enrollment Method</th>
<th>Sample Size</th>
<th>Median Age (yr)</th>
<th>Sex</th>
<th>Exclusion Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chee 2009,47 [19641191]</td>
<td>Selected patients</td>
<td>92 with negative IFE, including 29 with normalized SFLC ratio</td>
<td>59</td>
<td>nd</td>
<td>Not specified</td>
</tr>
</tbody>
</table>

Treatment:  
Bone marrow transplantation, 51  
Chemotherapy, 26  
Second-line therapy, 10  
Unknown, 5

IFE=immunofixation electrophoresis, KQ=Key Question, nd=no data, SFLC=serum free light chain, yr=years.

Table 10. Results of studies addressing KQ5

<table>
<thead>
<tr>
<th>Author Year [PMID]</th>
<th>Index Test</th>
<th>Comparator Test and Definition</th>
<th>Sample Size</th>
<th>Results</th>
</tr>
</thead>
</table>
| Chee 2009,47 [19641191] | Normal SFLC ratio | IFE test followed by bone marrow aspirate or biopsy, performed within 30 days of SFLC assay | 92 with negative IFE, including 29 with normalized SFLC ratio | 14% of patients with negative IFE had ≥5% plasma cells in bone marrow.  
10% of patients with normal SFLC ratio had >5% plasma cells in bone marrow. Addition of normal SFLC ratio to negative serum and urine IFE appears insufficient to confirm complete response accurately in the absence of a bone marrow aspirate or biopsy using standard EBMT/IMWG criteria. SFLC ratio does not eliminate the need for bone marrow for quantifying plasma cells for assessment of response in MM. |

EBMT=European Group for Blood and Bone Marrow Transplant, IFE=immunofixation electrophoresis, IMWG=International Myeloma Working Group, KQ=Key Question, MM=multiple myeloma, SFLC=serum free light chain.
Discussion

Since its introduction in 2001, the SFLC assay has been used in various clinical contexts: screening and diagnosis of PCDs, baseline measurement of SFLCs for disease prognostication, and quantitative monitoring of patients treated for PCDs in order to document treatment response, disease remission, or relapse. 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; its utility in prognostication for malignant PCDs; its role in determining treatment decisions; and whether its use could eliminate the need for other diagnostic tests. Table 11 summarizes the main findings addressing the five KQs of this CER.

Our results reveal that there is a paucity of evidence to clarify the comparative effectiveness of the role of the SFLC assay for the diagnosis, management, and prognosis of PCDs. Only 15 studies were identified in our literature search, having met all the inclusion criteria to address the KQs and being comparative in nature (see Appendix B for the excluded studies). Many articles evaluating the effectiveness and role of the SFLC assay were excluded because the populations did not fit into the specified eligibility criteria or there was a lack of data for traditional testing as a comparator. Across the included studies, there was considerable clinical heterogeneity with regard to variation in type or stage of disease and phase of treatment. In addition, 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 the uncertainties associated with the role of the assay in research and clinical practice. Finally, the majority of studies were of poor quality. All these factors limited the validity of the studies and the conclusions that could be drawn from them.

The role of the assay also remains uncertain in certain PCDs such as NSMM, LCMM, and AL amyloidosis, particularly in addressing comparative effectiveness. The insufficient evidence in these disease subgroups 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.

To synthesize our overall findings in more detail, below we present specific summaries of the state of the evidence for each KQ for which we found relevant publications (i.e., KQ1, KQ4, and KQ5) and describe the major needs of future studies.
Table 11. Summary of findings for KQs 1–5

<table>
<thead>
<tr>
<th>KQ</th>
<th>Strength of Evidence</th>
<th>Summary, Comments, and Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>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?</td>
<td>Insufficient (favoring use of the SFLC assay and ratio)</td>
<td>Three retrospective studies (all quality B) directly evaluated the SFLC assay in the context of diagnosing PCDs. All 3 compared test results with the diagnosis of disease verified by medical records. Although these studies showed an increase in sensitivity with the addition of the SFLC assay, owing to the heterogeneity in design, patient selection, and comparators used, meta-analysis could not be performed. The effect on specificity was inconsistent. <strong>Conclusions:</strong> 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.</td>
</tr>
<tr>
<td>KQ2: As compared with traditional tests, how well does the SFLC assay independently predict progression to MM in patients with MGUS?</td>
<td>Insufficient</td>
<td>No studies directly compared the use of the SFLC assay with traditional tests to determine whether it provided better prediction of progression to MM <strong>Conclusions:</strong> Owing to the lack of directly applicable data, we rated the evidence as insufficient.</td>
</tr>
<tr>
<td>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?</td>
<td>Insufficient</td>
<td>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. <strong>Conclusions:</strong> Owing to the lack of directly applicable data, we rated the evidence as insufficient.</td>
</tr>
<tr>
<td>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?</td>
<td>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</td>
<td>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 were in patients with AL amyloidosis and evaluated response to treatment as a predictor of outcomes; the other 8 studies were in 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 &gt;50% or &gt;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 use of the SFLC assay, precluding its utility in patients without elevated levels before treatment. Five of the 8 studies that enrolled patients with MM addressed the use of 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. <strong>Conclusions:</strong> Although SFLC response to therapy appeared to be a consistent predictor of outcomes in AL amyloidosis, there was no evidence that the SFLC assay is 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.</td>
</tr>
</tbody>
</table>
Table 11. Summary of findings for KQs 1–5 (continued)

<table>
<thead>
<tr>
<th>KQ</th>
<th>Strength of Evidence</th>
<th>Summary, Comments, and Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>KQ5: In PCD patients, does the use of the SFLC assay reduce the need for other diagnostic tests (e.g., bone marrow biopsy)?</td>
<td>Insufficient to support that use of the SFLC assay reduces the need for other diagnostic tests</td>
<td>One study (quality C) addressed this question. The study is 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 ≥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.</td>
</tr>
</tbody>
</table>

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.

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, conducted in a hospital laboratory setting, and were of 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 the addition of 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, the specific comparisons addressed, and whether patients with MGUS were included.

The presence of symptoms or laboratory abnormalities suggestive of a PCD usually triggers screening tests. Traditionally SPEP and UPEP would be performed; current recommendations include the SFLC assay as well. Positive tests would be followed with more detailed testing, including IFE and bone marrow examination. Ultimately, then, the diagnosis is based on a set of criteria including the results of the screening tests. 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 is often difficult to eliminate because the result from the reference test itself (e.g., SPEP or SIFE) is usually considered along with other factors, such as clinical information, 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. Another important caveat is that the diagnostic performance of the SFLC assay varies depending on the type and distribution of PCDs in the study sample. The SFLC assay detects polyclonal, not monoclonal, light chains and is only useful for PCDs associated with light chain production.
A related source of bias in this context is spectrum bias, in which the reported sensitivity and specificity may be exaggerated in populations with increased disease severity. Some measures recommended to maximize the quality of test interpretation include repeat testing and targeted followup of false positives, as well as blinding of data assessors to the diagnosis or test group to diminish the likelihood of misclassification bias. However, such safeguards were seldom emphasized in the studies reviewed. 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.

Patients without a pre-existing diagnosis of PCD were selected as the relevant target population for KQ1, with input from the Technical Expert Panel. This approach was taken to allow for a comparison of test accuracy among patients that were not preselected as having disease. 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, where 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.

Although there is a large body of literature relating to the effectiveness of the SFLC assay in diagnosis of various PCDs, there is limited information on its comparative effectiveness. Most studies assessing comparative effectiveness have either compared the SFLC assay alone (not as an adjunct) versus one or more traditional tests, in either undiagnosed or diagnosed populations or have examined the SFLC assay as an adjunct but only in populations already diagnosed with a monoclonal gammopathy or AL amyloidosis. Several studies examined the issue of test accuracy in patients diagnosed with disease. These studies did not meet our population eligibility criterion, as they could not address test performance in patients who did not have disease. Included in these studies was one large trial: Katzmann 2009, that tested 1877 patients with a diagnosis of PCD by the SFLC assay, SPEP, UPEP, SIFE, or UIFE. The authors examined the diagnostic accuracy of these tests singly and in combination. Other studies compared the SFLC assay as a standalone test (not in combination with traditional testing) with traditional tests. Only two of these studies were carried out in undiagnosed patients, comparing SFLC testing alone with traditional testing, one for the detection of monoclonal protein (n=691) and one for the diagnosis of monoclonal gammopathy (n=753).

**SFLC Assay and Treatment Response and Survival (KQ4)**

Eleven studies, three in patients with AL amyloidosis and eight in patients with MM, evaluated SFLC testing compared with traditional testing for assessing treatment response and in relation to 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 test used as a comparator for the SFLC assay.

The three studies of AL amyloidosis examined the relationship of SFLC response to treatment and outcomes, in addition to measuring quantitative M protein responses and independently evaluating the ability of each to predict outcomes. In all three, a reduction in the SFLC concentration after treatment was associated with improved survival. Despite this finding, 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 studies 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.

Eight studies were reviewed in patients with MM.34-37,39,41-43 Most (75 percent) were retrospective cohort studies, and only three were of quality B. Five addressed the use of SFLC assay in assessing or predicting response to treatment. The traditional test comparators reported varied in each study. Discordance of the SFLC response and the response as assessed by traditional testing was found in all the studies, but four reported achievement of an SFLC response prior to a response on traditional tests. Two studies examined the relationship between baseline SFLC concentrations and survival, and three 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. The strength of evidence for SFLC response being a better predictor of survival than traditional testing was rated as insufficient. Consideration of the B quality studies only did not qualitatively change the pattern of observations outlined above or the grading of the strength of evidence. In the literature search, we found other studies of SFLC concentrations as a prognostic indicator in MM with regard to survival outcomes, renal outcomes, and light chain escape, but none were comparative in nature.

The strength of evidence for this KQ was insufficient for both AL amyloidosis and MM for all outcomes examined. Limitations in the literature reviewed were several. Demographic details, including distributions of races or ethnic groups and comorbid conditions, were not consistently reported. Information was limited 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 versus 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 limit the applicability of the findings to both the general PCD population and subgroups of interest.

**SFLC Assay in Outcome Prediction, Treatment Decisions, and Reducing Other 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). The literature reviewed in relation to this KQ consisted of two retrospective cohort studies and one case–control study that compared rates of progression among patients with different baseline SFLC ratios but not in comparison to traditional testing.17,18,69 There is a growing awareness that patients with MGUS who have elevated SFLC concentrations may have a different disease biology than patients with MGUS whose SFLC concentrations are normal, and some incorporate the SFLC ratio into risk-scoring systems for MGUS progression.

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. Two noncomparative studies reported results of treatment protocols determined by SFLC testing, one to define the need for adjuvant therapy in patients with AL amyloidosis\(^7^0\) and the other to determine the need for high-cut-off hemodialysis in combination with chemotherapy for the removal of SFLCs in patients with cast nephropathy.\(^7^1\) More information in this context is anticipated from the results of the BMT CTN 0702 trial (Center for International Blood and Marrow Transplant Research [www.cibmtr.org/Studies/ClinicalTrials/BMT_CTN/Protocols/Pages/0702.aspx]), which will be prospectively collecting serum samples for light chain analysis, along with flow-cytometry measurement of bone marrow and traditional tests for M protein. This should provide useful information regarding the role of the SFLC assay in monitoring MM patients.

A single study, Chee 2009,\(^4^7\) explored whether the use of the SFLC assay compared with traditional testing would reduce the need for other diagnostic tests (re KQ5). The authors evaluated whether a negative IFE result or normalization of the SFLC ratio (or both) after treatment of MM is sufficient to characterize a hematological response,\(^2^2,2^7\) such that the need for bone marrow examination to evaluate the percentages of plasma cells (to stringently define remission) could potentially be eliminated. Bone marrow examinations can be cumbersome in clinical practice and uncomfortable for patients, causing considerable noncompliance among physicians. Ten percent of patients with such an achievement still had 5 percent or more of plasma cells in marrow, and the authors concluded that bone marrow examination should not be eliminated for the assessment of treatment response. Since this conclusion is based on one study only, this question requires more detailed and systematic evaluation.

Limitations

As discussed above, 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, interventions (diagnostic test/disease monitoring), 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 do not include newer advances such as positron emission tomography. Finally, most studies were underpowered with respect to PCDs where 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,\(^7^2\) but no studies that were included in our review addressed whether race modifies 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.\(^7^3\)

We had to exclude the majority of studies of diagnostic accuracy of SFLC testing we found because they were carried out in populations with preexisting diagnosis of disease. These findings cannot readily be generalized to undiagnosed populations, which is the population of interest. Although such studies were excluded from our review, we found that the included studies also have potential biases (selection, spectrum, incorporation, and other types of bias) that limited generalizability.
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.

**Context of Findings**

Current clinical uses of the SFLC assay in MM and related disorders focus on three main areas: the diagnostic, therapeutic, and monitoring approach to PCDs. Here, we discuss the applicability of the evidence for comparative effectiveness for current practice.

In the setting of diagnosis, the SFLC assay has been used primarily in patients suspected of having a PCD. The SFLC assay in combination with SPEP and SIFE is highly sensitive and its use potentially negates the need for 24-hour urine studies for diagnoses other than AL amyloidosis. In this CER, we identified only three studies that assessed the added value of the FLC assay in undiagnosed populations compared with traditional testing. Given the practical difficulties associated with obtaining a 24-hour urine sample, SFLC assay would be of tremendous value if its effectiveness is confirmed. Although the comparative diagnostic efficacy of the SFLC assay versus UPEP or UIFE has been shown in patients with preexisting disease, it has not yet been shown in undiagnosed populations, where the danger of false negatives for the SFLC assay has not been thoroughly vetted. On the other hand, if an abnormal SFLC ratio is the only test in a diagnostic panel that signals a PCD (e.g., light chain MGUS), it will be difficult to further evaluate positive test results that may be erroneous. This conundrum exemplifies the challenges surrounding evaluation of a test in monoclonal disorders, given their heterogeneity and need for a multiplicity of tests to define a full diagnosis. It is likely that studies based on diagnostic samples from patients with confirmed disease will yield inflated estimates of test accuracy.

The baseline measurement of FLCs has been found to have major prognostic value for virtually every PCD. Another important group is the oligosecretory PCDs (including AL amyloidosis, NSMM, and LCMM), for which the SFLC assay can be useful for quantitative monitoring of patients. This assay has been used as a clinical tool in both settings. However, we found no evidence to assess its comparative value against traditional testing or bone marrow examination.

The SFLC kappa/lambda ratio has also been used to define a stringent complete treatment response. We did not find sufficient evidence that a complete response, defined with or without the SFLC ratio criteria, provided differential prognoses for progression-free survival or overall survival or that stringent complete response correlated with bone marrow response. The recognition of light chain escape by periodic SFLC measurements is another relevant indication for the use of the assay in therapeutic monitoring, given the changing disease behavior in response to chemotherapy. However, we found very few studies addressing light chain escape.

In summary, this CER demonstrates a paucity of evidence to determine the benefits of the use of SFLC assay instead of or as an adjunct to traditional testing. While the clinical effectiveness of the test in various settings was not the focus of the CER, its end users—clinicians, consumers, and policymakers—should be aware that there remains uncertainty regarding SFLC testing in a comparative context. There are clear evidence gaps for the clinician
who is using the test, and these lend themselves to defining the particular research gaps that we focus on in the next section.

**Future Research**

Uncertainties remain regarding the applications of the SFLC assay both within and beyond the 2009 IMWG consensus guidelines. 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 therapeutic decisions and monitoring, recognition of 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 clinical impression of its effectiveness, the role of future research into the assay’s comparative effectiveness should be targeted toward populations and settings that will potentially maximize its utility.

**SFLC Assay in Diagnostic Testing**

Prospectively designed single-cohort studies consisting of both diseased and nondiseased people, representative of the clinically relevant population where a PCD may be suspected, are needed to provide a more accurate assessment of the effect of adding SFLC to traditional tests used to diagnose PCDs. Studies should have *a priori* calculation of the sample size needed for determination of the desired precision and should include inferences based on formal statistical testing of estimates of diagnostic accuracy. Although it has been repeatedly suggested that serum SFLC measurement can replace the 24-hour urine collection for UPEP or UIFE in diagnostic panels, these studies have only been performed in patients with disease, so evidence for replacement is still lacking.

There are practical difficulties associated with obtaining a urine sample, and much of the SFLC assay’s value is that it does not require urine collection. While the comparative diagnostic performance of the SFLC assay and UPEP or UIFE has been shown in patients with preexisting disease, this is not true of undiagnosed populations, where the danger of false negatives for the FLC assay has not been thoroughly vetted. More study is needed in this regard.

Inherent challenges exist in carrying out diagnostic-testing studies for PCDs, which should be addressed to facilitate further study. The potentially increased sensitivity of the SFLC assay has the downside of increasing the number of false positive results, but 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. 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 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.
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 (e.g., MM, NSMM, LCMM, AL amyloidosis, and light chain deposition disease). Definitions of FLC response are largely empirical in the current guidelines for AL amyloidosis (Consensus Opinion from the 10th International Symposium on Amyloid and Amyloidosis) and MM (International Uniform Response Criteria) 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 either 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 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 and 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 or whether it would help prognosticate the disease course. Its precise role and optimal use across the spectrum of PCDs and clinical settings still needs 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.
### Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AACC</td>
<td>American Association for Clinical Chemistry</td>
</tr>
<tr>
<td>AHRQ</td>
<td>Agency for Healthcare Research and Quality</td>
</tr>
<tr>
<td>AL amyloidosis</td>
<td>Systemic, or primary, amyloidosis in which amyloid [A] proteins derived from immunoglobulin light chains [L] are deposited in tissue (also called light chain amyloidosis)</td>
</tr>
<tr>
<td>CER</td>
<td>Comparative Effectiveness Review</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence interval</td>
</tr>
<tr>
<td>EBMT</td>
<td>European Group for Blood and Bone Marrow Transplant</td>
</tr>
<tr>
<td>ECOG</td>
<td>Eastern Cooperative Oncology Group</td>
</tr>
<tr>
<td>EPC</td>
<td>Evidence-based Practice Center</td>
</tr>
<tr>
<td>FDA</td>
<td>Food and Drug Administration</td>
</tr>
<tr>
<td>FLC</td>
<td>Free light chain</td>
</tr>
<tr>
<td>IFE</td>
<td>Immunofixation electrophoresis</td>
</tr>
<tr>
<td>IMWG</td>
<td>International Myeloma Working Group</td>
</tr>
<tr>
<td>ISS</td>
<td>International Staging System</td>
</tr>
<tr>
<td>LCMM</td>
<td>Light chain multiple myeloma</td>
</tr>
<tr>
<td>MGUS</td>
<td>Monoclonal gammopathy of undetermined significance</td>
</tr>
<tr>
<td>MM</td>
<td>Multiple myeloma</td>
</tr>
<tr>
<td>M protein</td>
<td>Monoclonal protein (also called paraprotein)</td>
</tr>
<tr>
<td>NSMM</td>
<td>Nonsecretory multiple myeloma</td>
</tr>
<tr>
<td>PCD</td>
<td>Plasma-cell dyscrasia</td>
</tr>
<tr>
<td>KQ</td>
<td>Key Question</td>
</tr>
<tr>
<td>PICO (also PICOTS)</td>
<td>Populations, interventions (diagnostic tests/disease monitoring), comparators, outcomes (and timing and settings)</td>
</tr>
<tr>
<td>SCT</td>
<td>Stem-cell transplantation</td>
</tr>
<tr>
<td>SFLC</td>
<td>Serum free light chain</td>
</tr>
<tr>
<td>SIFE</td>
<td>Serum immunofixation electrophoresis</td>
</tr>
<tr>
<td>SPEP</td>
<td>Serum protein electrophoresis</td>
</tr>
<tr>
<td>TEP</td>
<td>Technical Expert Panel</td>
</tr>
<tr>
<td>TOO</td>
<td>Task Order Officer</td>
</tr>
<tr>
<td>UIFE</td>
<td>Urine immunofixation electrophoresis</td>
</tr>
<tr>
<td>UPEP</td>
<td>Urine protein electrophoresis</td>
</tr>
<tr>
<td>VDD</td>
<td>Bortezomib, pegylated liposomal doxorubicin, and dexamethasone</td>
</tr>
<tr>
<td>VGPR</td>
<td>Very good partial response</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>---------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Cast nephropathy (or “myeloma kidney”)</td>
<td>Disorder in which monoclonal urinary immunoglobulin light chains (Bence Jones proteins) lead to acute or chronic renal failure through intratubular cast formation and direct tubular toxicity</td>
</tr>
<tr>
<td>Differential verification</td>
<td>Verification of test result or disease status of each patient using one of a variety of standards rather than one reference standard across the whole study population, which is problematic if the tests vary in accuracy</td>
</tr>
<tr>
<td>Disease progression or recovery bias</td>
<td>Bias from an inappropriately long interval (or any interval) between conduct of reference test and conduct of index test</td>
</tr>
<tr>
<td>Incorporation bias</td>
<td>Bias caused by use of a reference test consisting of a suite of investigations, including the index test results (and thereby overestimating the diagnostic accuracy of the index test)</td>
</tr>
<tr>
<td>Involved FLC or SFLC</td>
<td>The free light chain or serum free light chain that is produced in excess and is causing disease (either kappa or lambda)</td>
</tr>
<tr>
<td>Light chain escape</td>
<td>A type of plasma-cell-dyscrasia remission in which, for unclear reasons, a subclone of malignant plasma cells expands that is incapable of producing significant amounts of immunoglobulin heavy chain but retains the ability to make light chains</td>
</tr>
<tr>
<td>Measurable disease</td>
<td>Presence of a plasma cell dyscrasia but absence of elevated SFLC concentrations before treatment, therefore precluding use of the SFLC assay as a marker of disease and treatment response</td>
</tr>
<tr>
<td>M protein or paraprotein</td>
<td>Intact immunoglobulins or FLCs of a single type produced in excess by an abnormally expanded clone of malignant plasma cells (biomarkers of PCDs)</td>
</tr>
<tr>
<td>Monoclonal gammopathy</td>
<td>Disease class comprising PCDs as well as other conditions such as hematological disorders associated with a monoclonal band</td>
</tr>
<tr>
<td>Oligosecretory MM</td>
<td>MM in which very small amounts of M protein are produced by the malignant plasma cells</td>
</tr>
<tr>
<td>Polyclonal gammopathy</td>
<td>Disease similar to monoclonal gammopathy (or PCD) except that the clonal expansion occurs across various B-cell populations that produce more than one kind of immunoglobulin</td>
</tr>
<tr>
<td>Term</td>
<td>Description</td>
</tr>
<tr>
<td>------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Selection bias (also called partial verification bias, workup bias, or sequential ordering bias)</td>
<td>Bias resulting from failure to verify the disease status or test result of all, or a random selection of, enrolled patients with the use of the reference standard</td>
</tr>
<tr>
<td>SFLC ratio</td>
<td>The ratio of kappa chains to lambda chains, for which the normal range is 0.26–1.65</td>
</tr>
<tr>
<td>Spectrum bias or effect</td>
<td>Bias caused by use of sampling methods unlikely to capture a representative sample (for purposes of this review; term also can refer to bias from representation of inappropriate patient population)</td>
</tr>
<tr>
<td>Verification bias</td>
<td>Incomplete verification of index test results</td>
</tr>
</tbody>
</table>
References


30. 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.


43. Dytfeld D, Griffith KA, Friedman J, et al. Superior overall survival of patients with myeloma achieving very good partial response or better to initial treatment with bortezomib, pegylated liposomal doxorubicin, and dexamethasone, predicted after two cycles by a free light chain- and M-protein-based model: extended follow-up of a phase II trial. Leukemia & Lymphoma 2011 Jul;52(7):1271-80.


Appendix A. Literature Search Strategy

Databases: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations <January 31, 2012>, EBM Reviews - Cochrane Database of Systematic Reviews <2005 to January 31, 2012>, EBM Reviews - Cochrane Central Register of Controlled Trials <1st Quarter 2012>, Ovid MEDLINE(R) without Revisions <1996 to January Week 4 2012>

Last run 1/31/2012

1 Immunoglobulin Light Chain*.mp. or exp Immunoglobulin Light Chains/
2 monoclonal light chain*.mp.
3 serum free light chain*.mp.
4 immunoglobulin-free light chain*.mp.
5 Bence Jones protein.mp. or exp Bence Jones Protein/
6 1 or 2 or 3 or 4 or 5
7 limit 6 to english language [Limit not valid in CDSR,CCTR; records were retained]
8 limit 7 to yr="2000 -Current"
9 remove duplicates from 8
Appendix B. Excluded Studies

Of the 325 articles obtained for full-text review, 15 were included and 310 were excluded; most failed to meet KQ inclusion criteria regarding the test, population, diagnosis, or comparison of interest. The other main reasons for exclusion were that the article was a narrative review, commentary, or letter without sufficient data or that it described a single case series.

All 310 excluded references are presented below, in alphabetic order of first author’s surname, along with the reason for exclusion for each.

Studies Excluded after Full-Text Review (n=310)


Not relevant re test, population, diagnosis, or comparison.


Not relevant re test, population, diagnosis, or comparison.


Not relevant re test, population, diagnosis, or comparison.


Narrative review or commentary.


Not relevant re test, population, diagnosis, or comparison.


Letter without data.


Not relevant re test, population, diagnosis, or comparison.


Study of single case series.


Not relevant re test, population, diagnosis, or comparison.


Not relevant re test, population, diagnosis, or comparison.


Not relevant re test, population, diagnosis, or comparison.


Narrative review or commentary.


Not relevant re test, population, diagnosis, or comparison.


Not relevant re test, population, diagnosis, or comparison.


Not relevant re test, population, diagnosis, or comparison.


41. Ching AK, Li PS, Chan WY, et al. Strand bias in Ig somatic hypermutation is determined by signal sequence within the variable region. International Immunology 2000 Sep;12(9):1245-53. Not relevant re test, population, diagnosis, or comparison.


Not relevant re test, population, diagnosis, or comparison.


Study of single case series.

Not relevant re test, population, diagnosis, or comparison.

Not relevant re test, population, diagnosis, or comparison.

Narrative review or commentary.

Not relevant re test, population, diagnosis, or comparison.

Not relevant re test, population, diagnosis, or comparison.

Not relevant re test, population, diagnosis, or comparison.

Not relevant re test, population, diagnosis, or comparison.

Not relevant re test, population, diagnosis, or comparison.

Not relevant re test, population, diagnosis, or comparison.

Not relevant re test, population, diagnosis, or comparison.

Narrative review or commentary.

Not relevant re test, population, diagnosis, or comparison.

Not relevant re test, population, diagnosis, or comparison.

Not relevant re test, population, diagnosis, or comparison.
   Not relevant re test, population, diagnosis, or comparison.

   Not relevant re test, population, diagnosis, or comparison.

   Not relevant re test, population, diagnosis, or comparison.

   Not relevant re test, population, diagnosis, or comparison.

   Not relevant re test, population, diagnosis, or comparison.

   Not relevant re test, population, diagnosis, or comparison.

   Not relevant re test, population, diagnosis, or comparison.

   Letter without data.

   Not relevant re test, population, diagnosis, or comparison.

   Not relevant re test, population, diagnosis, or comparison.

   Not relevant re test, population, diagnosis, or comparison.

   Not relevant re test, population, diagnosis, or comparison.

   Not relevant re test, population, diagnosis, or comparison.

   Narrative review or commentary.

   Narrative review or commentary.


Not relevant re test, population, diagnosis, or comparison.


Not relevant re test, population, diagnosis, or comparison.

Not relevant re test, population, diagnosis, or comparison.

206. Nemazee D. Receptor editing in B cells. [Review] [231 refs]. Advances in Immunology 2000;74:89-126.
Not relevant re test, population, diagnosis, or comparison.

Not relevant re test, population, diagnosis, or comparison.

Not relevant re test, population, diagnosis, or comparison.

Not relevant re test, population, diagnosis, or comparison.

Not relevant re test, population, diagnosis, or comparison.

Not relevant re test, population, diagnosis, or comparison.

Narrative review or commentary.

Not relevant re test, population, diagnosis, or comparison.

Not relevant re test, population, diagnosis, or comparison.

Narrative review or commentary.

Not relevant re test, population, diagnosis, or comparison.

Not relevant re test, population, diagnosis, or comparison.

Not relevant re test, population, diagnosis, or comparison.

Not relevant re test, population, diagnosis, or comparison.


223. Pika T, Minarik J, Schneiderka P, et al. The correlation of serum immunoglobulin free light chain levels and selected biological markers in multiple myeloma. Biomedical Papers of the Medical Faculty of Palacky University in Olomouc, Czech Republic 2008 Jun;152(1):61-64. Not relevant re test, population, diagnosis, or comparison.


234. Rajkumar SV, Kyle RA, Therneau TM, et al. Serum free light chain ratio is an independent


Not relevant re test, population, diagnosis, or comparison.

Not relevant re test, population, diagnosis, or comparison.

Not relevant re test, population, diagnosis, or comparison.

Not relevant re test, population, diagnosis, or comparison.

Not relevant re test, population, diagnosis, or comparison.

Not relevant re test, population, diagnosis, or comparison.

Not relevant re test, population, diagnosis, or comparison.

Not relevant re test, population, diagnosis, or comparison.

Not relevant re test, population, diagnosis, or comparison.

Not relevant re test, population, diagnosis, or comparison.

Not relevant re test, population, diagnosis, or comparison.

Narrative review or commentary.

Not relevant re test, population, diagnosis, or comparison.

Not relevant re test, population, diagnosis, or comparison.


306. Xu J, Davis MM. Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities. Immunity 2000 Jul;13(1):37-45. *Not relevant re test, population, diagnosis, or comparison.*

307. Yang CY. Using a heavy chain-loss hybridoma 26.4.1LL for studying the structural basis of immunoglobulin chain association. Proceedings
Not relevant re test, population, diagnosis, or comparison.

Not relevant re test, population, diagnosis, or comparison.

Not relevant re test, population, diagnosis, or comparison.

Narrative review or commentary.
# Appendix C. Quality Criteria and Individual Study Grades

## Table for Key Question 1

<table>
<thead>
<tr>
<th>Author Year [PMID]</th>
<th>Prospective/ Retrospective</th>
<th>Selection/ spectrum bias</th>
<th>Case–control design</th>
<th>Consecutive patient selection</th>
<th>Lack of verification bias</th>
<th>Blinded index-test readers</th>
<th>Proper analysis if repeated sampling</th>
<th>Time interval between index and reference test reported</th>
<th>Statistical test used to quantify uncertainty</th>
<th>Quality Grade</th>
<th>Summary of grade rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abadie 2006¹ [16682511]</td>
<td>R</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>B</td>
<td>No measure of statistical uncertainty, no major biases, clear description of population</td>
</tr>
<tr>
<td>Piehler 2008² [18801937]</td>
<td>P</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>B</td>
<td>No measure of statistical uncertainty, no major biases, consecutive recruitment</td>
</tr>
<tr>
<td>Vermeersch 2008³ [18729849]</td>
<td>R</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>ND</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>B</td>
<td>No measure of statistical uncertainty, described sample, no major biases</td>
</tr>
</tbody>
</table>

Y = Yes, N = No, ND = not described, P = Prospective study design, R = Retrospective study design.

Types of bias are defined in Glossary and also described at the end of each row under “Summary of grade rationale.”

Criteria are derived from STARD (www.stard-statement.org) and STROBE (www.strobe-statement.org).
### Table for Key Questions 4–5

<table>
<thead>
<tr>
<th>Author Year [PMID]</th>
<th>Prospective/ Retrospective</th>
<th>Outcomes clearly defined</th>
<th>Bias present</th>
<th>Confounders clearly defined/ analyzed</th>
<th>Loss to follow up explained</th>
<th>Population clearly described</th>
<th>Data lost/not analyzed/ missing</th>
<th>Inclusion/ exclusion criteria defined</th>
<th>Quality grade</th>
<th>Summary of grade rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Key Question 4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dispenzieri, 2008</td>
<td>R</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>ND</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>B</td>
<td>Retrospective without adjustment</td>
</tr>
<tr>
<td>[18364469]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dyfield, 2011</td>
<td>P</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>ND</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>C</td>
<td>Small sample size, sample not uniformly treated</td>
</tr>
<tr>
<td>[21699382]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Giarin, 2009</td>
<td>R</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>ND</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>B</td>
<td>Retrospective without adjustment</td>
</tr>
<tr>
<td>[19520760]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Khoriaty, 2010</td>
<td>R</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>ND</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>C</td>
<td>Small sample size, retrospective without adjustment</td>
</tr>
<tr>
<td>[20223721]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kroger, 2010</td>
<td>?</td>
<td>N</td>
<td>?</td>
<td>N</td>
<td>ND</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>C</td>
<td>Letter to the editor with limited information, small sample size, SFLC response definitions not described, few details about study design, limited data</td>
</tr>
<tr>
<td>[2043663]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kumar 2011</td>
<td>R</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>ND</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>C</td>
<td>Retrospective, extreme selection/spectrum bias</td>
</tr>
<tr>
<td>[21328431]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kyrtsonis, 2007</td>
<td>P</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>ND</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>C</td>
<td>Limited information about patient recruitment and study design, small sample size</td>
</tr>
<tr>
<td>[17408464]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lachmann 2003</td>
<td>P</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>ND</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>C</td>
<td>Retrospective, selection/spectrum bias, sample not uniformly treated</td>
</tr>
<tr>
<td>[12823348]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paiva, 2011</td>
<td>R</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>ND</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>C</td>
<td>Retrospective without adjustment, potential selection bias because inclusion was based on availability of serum samples</td>
</tr>
<tr>
<td>[21402611]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sanchorawala 2005</td>
<td>R</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>ND</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>C</td>
<td>Retrospective, small sample size</td>
</tr>
<tr>
<td>[16044137]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Van Rhee, 2007</td>
<td>P</td>
<td>Y</td>
<td>?</td>
<td>N</td>
<td>ND</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>B</td>
<td>Retrospective with adjustment</td>
</tr>
<tr>
<td>[17416735]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Key Question 5</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chee 2009</td>
<td>R</td>
<td>n</td>
<td>?</td>
<td>N</td>
<td>ND</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>C</td>
<td>Small sample, selection/spectrum bias</td>
</tr>
<tr>
<td>[19641191]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Y = Yes, N = No, ND = not described, P = Prospective study design, R = Retrospective study design, ? = unclear.

Types of bias are defined in Glossary and also described at the end of each row under “Summary of grade rationale.”

Criteria are derived from STARD (www.stard-statement.org) and STROBE (www.strobe-statement.org).
Appendix C References


5. Dytfeld D, Griffith KA, Friedman J, et al. Superior overall survival of patients with myeloma achieving very good partial response or better to initial treatment with bortezomib, pegylated liposomal doxorubicin, and dexamethasone, predicted after two cycles by a free light chain- and M-protein-based model: extended follow-up of a phase II trial. Leukemia & Lymphoma 2011 Jul;52(7):1271-80.


