DRAFT Comparative Effectiveness Review

Number xx

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

This information is distributed solely for the purposes of predissemination peer review. It has not been formally disseminated by the Agency for Healthcare Research and Quality. It does not represent and should not be construed to represent an Agency for Healthcare Research and Quality or Department of Health and Human Services (AHRQ) determination or policy.

Contract No. [REDACTED]

Prepared by:
[REDACTED]

Investigators:
[REDACTED]

AHRQ Publication No. xx-EHCxxx
<Month Year>

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

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

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

This document is in the public domain and may be used and reprinted without permission except those copyrighted materials that are clearly noted in the document. Further reproduction of those copyrighted materials is prohibited without the specific permission of copyright holders.

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

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

Suggested citation: [REDACTED]

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 http://www.effectivehealthcare.ahrq.gov/reference/purpose.cfm.

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

Transparency and stakeholder input from are essential to the Effective Health Care Program. Please visit the Web site (http://www.effectivehealthcare.ahrq.gov) to see draft research questions and reports or to join an e-mail 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 e-mail to epc@ahrq.hhs.gov.

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

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

Stephanie Chang, M.D., M.P.H.
Director
Evidence-based Practice Program
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

Investigator Affiliations

[REDACTED]

Acknowledgments

[REDACTED]

Key Informants

[REDACTED]

Technical Expert Panel

[REDACTED]

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 combination with traditional testing by electrophoresis and immunofixation, compared with traditional testing alone, in diagnosis and 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 to March 2011.

Review 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, outcomes, and study methodological issues. We considered studies of adults with undiagnosed and diagnosed PCDs, specifically monoclonal gammopathy of undetermined significance (MGUS) or multiple myeloma (MM), which includes light chain MM, nonsecretory MM, and AL amyloidosis (in which amyloid [A] proteins derived from immunoglobulin light chains [L] are deposited in tissue). The interventional comparison of interest was traditional tests for PCDs plus the SFLC assay versus traditional tests alone. The main outcomes of interest were measures of diagnostic accuracy, progression to MM, and treatment response.

Results: The literature search yielded 2819 citations, 13 of which met the inclusion criteria. Three fair-quality retrospective 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 six studies (three fair quality, three poor quality) of MM (five 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 low methodological quality and considerable clinical heterogeneity, the strength of evidence was rated as insufficient for the superiority of the combination of the SFLC assay and traditional testing over traditional testing alone: specifically, the SLFC assay's impact on diagnostic accuracy in undiagnosed patients, as well as whether it is a better predictor of outcome or progression of MGUS to MM, a better indicator for therapeutic decisionmaking, or an appropriate substitute for other interventions.

Conclusions: The role of the SFLC assay remains to be defined. The evidence was rated as insufficient to suggest that the SFLC assay may increase sensitivity when used as an adjunct to traditional testing for diagnosis of PCDs, 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**
Background ES-1
Objectives ES-2
Key Questions ES-3
Analytic Framework ES-3
Methods ES-4
Input from Stakeholders ES-4
Data Sources and Selection ES-4
Data Extraction and Quality Assessment ES-5
Data Synthesis and Analysis ES-6
Grading the Body of Evidence ES-6
Results ES-6
Literature Search ES-6
Study Quality Grade and Overall Strength of Evidence ES-7
Findings for KQs ES-8
Discussion ES-9
SFLC Assay and Diagnostic Testing (KQ1) ES-10
SFLC assay and treatment response and survival (KQ4) ES-11
SFLC assay in reducing the need for other interventions (KQ5) ES-11
Future Research ES-12
Conclusions ES-13
Introduction **1**
Key Questions 3
Methods **3**
AHRQ Task Order Officer 4
External Expert Input 4
Analytic Framework 4
Literature Search 5
Study Selection and Eligibility Criteria 5
Data Extraction and Data Management 6
Quality Assessment 6
Data Synthesis 7
Grading the Body of Evidence for Each KQ 7
Results **8**
Literature Search 8
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? 10
Results 10
Summary 11
KQ2: As compared with traditional tests, how well does the SFLC assay independently predict progression to MM in patients with MGUS? 14
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? 14
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)? 14
Results 14
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 interventions (e.g., bone marrow biopsy)? 24
Results 24
Discussion 26
SFLC Assay and Diagnostic Testing (KQ1) 26
SFLC Assay and Treatment Response and Survival (KQ4) 27
SFLC Assay in Outcome Prediction, Treatment Decisions, and Reducing Interventions (KQ2, KQ3, and KQ5) 28
Future Research 29
SFLC Assay in Diagnostic Testing 29
SFLC Assay in Risk Stratification and in Determining Prognosis 30
Reporting on the SFLC Assay 30
Conclusions 31
Acronyms 32
Glossary 33
References 35

Tables

Table 1. Distribution and quality of the 13 studies addressing a KQ 10
Table 2. Characteristics of studies addressing KQ1 12
Table 3. Characteristics of patients in studies addressing KQ1 12
Table 4. Results of studies addressing KQ1 12
Table 5. Characteristics of studies addressing KQ4 18
Table 6. Characteristics of patients in studies addressing KQ4 20
Table 7. Results of studies addressing KQ4 20
Table 8. Characteristics of studies addressing KQ5 25
Table 9. Characteristics of patients in studies addressing KQ5 25
Table 10. Results of studies addressing KQ5

25

Figures

Figure 1. Analytic Framework for SFLC analysis for the Diagnosis, Management, and Prognosis of PCDs 4
Figure 2. Summary of Search and Selection of Articles.....9

Appendices

Appendix A. Literature Search Strategy
Appendix B. Excluded Studies
Appendix C. Quality Criteria and Individual Study Grades

Executive Summary

Background

Plasma-cell dyscrasias (PCDs) are a group of clonal 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 produce 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 with organ dysfunction.

The International Myeloma Working Group (IMWG) recommend the following actions and tests for evaluation of a patient suspected of having a myeloma[7,8]: a complete history taking and physical examination; routine laboratory testing including serum protein electrophoresis (SPEP), serum immunofixation (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 (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 three percent of cases of NSMM, LCMM, or AL amyloidosis are not detected by traditional tests.[9] 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.

The SFLC assay (the Freelite™ Assay, The Binding Site Ltd., Birmingham, United Kingdom) was introduced in 2001 to measure the FLC component in particular.[10,11] The SFLC 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 (i.e., 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, with a sensitivity of less than 1 mg/dL in serum and less than 200 mg/day in urine[12] for detecting low concentrations of FLCs. The other main advantage is the ability to measure the ratio of kappa chains to lambda chains, for which the diagnostic range is 0.26–1.65.[13] This enables the identification of the FLC that is produced in excess 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. An abnormal ratio provides a useful index of clonality, as clonal disorders produce disproportionately high concentrations of a single type of light chain.

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 International Myeloma Working Group (IMWG) currently considers the SFLC assay to be an adjunct to traditional tests.[14] The assay could allow for quantitative monitoring of response and remission after treatment and provide prognostic information[15,16]; potentially reduce 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[14]; and 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,[13,14] although proper studies have yet to be undertaken to explore this potential benefit. 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[17]; 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 its short half-life (2 to 6 hours, vs. 21 days for complete immunoglobulins[18]), it may also help detect relapses and treatment failures earlier than by reliance on M-protein concentrations alone. These diagnostic applications need to be validated and standardized, however.[10] Thus, although the SFLC assay has been in use for a decade, it remains unclear how best to incorporate it into clinical practice.[19]

Objectives

The objective of this report is to summarize the existing literature regarding the role of SFLC testing in the diagnosis, management, and prognosis of patients with PCDs.

Key Questions

1. 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?

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

3. 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)?

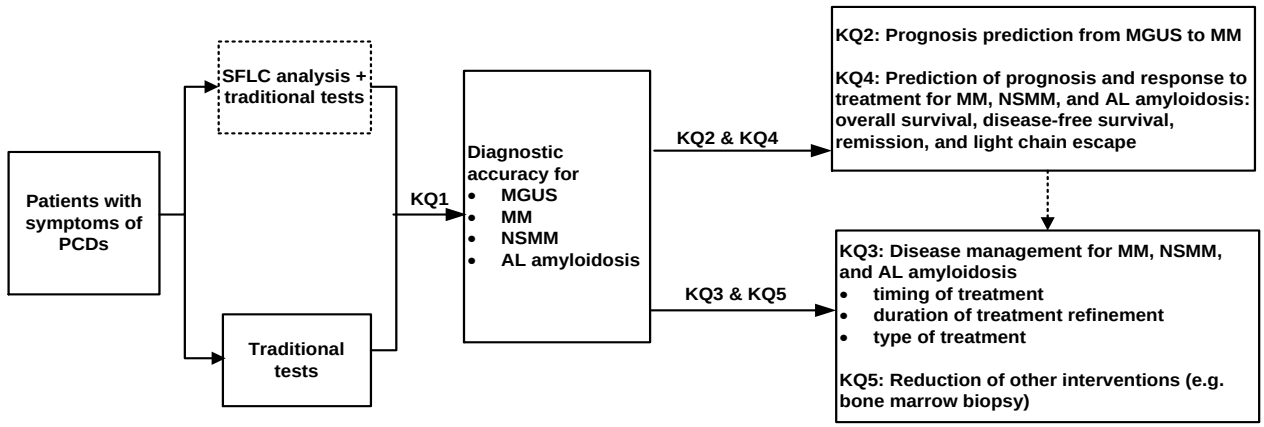
4. In patients with an existing diagnosis of PCD (MM, NSMM, or AL amyloidosis), is the SFLC assay better than traditional tests in indicating how the patient responds to treatment and of outcomes (overall survival, disease-free survival, remission, light chain escape, and quality of life)?

5. 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 interventions (e.g., bone marrow biopsy)?

Analytic Framework

To guide the development of the Key Questions (KQs), we developed an analytic framework (**Figure 1**) that maps the specific linkages associating the population (patients with PCD symptoms) to the additional intervention (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 interventions, 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 gammopathy of undetermined significance, MM = multiple myeloma, NSMM = nonsecretory multiple myeloma, PCD = plasma-cell dyscrasia, SFLC = serum free light chain.

Methods

Input from Stakeholders

This project had two phases: an initial phase of topic refinement, in which the KQs were proposed and refined, and a comparative effectiveness review phase, which we describe here. During the topic refinement phase, we formed a panel of Key Informants to help fine-tune the KQs. During the comparative effectiveness review phase, we formed a separate Technical Expert Panel (with some Key Informants as members) that provided clinical and methodological expertise in interpreting the KQs, identifying important issues, and defining parameters for the review of evidence. In addition, input from these experts was sought when questions arose during the review.

Data Sources and Selection

We conducted literature searches of studies published from January 1, 2000, to March 11, 2011, in MEDLINE®, the Cochrane Central Register of Controlled Trials, and the Cochrane Database of Systematic Reviews. The start year was chosen as 2000 because the SFLC assay was approved by the FDA in 2001; we considered that any reports of clinical use of the assay prior to 2000 would not be representative of the approved test. 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. We did not include unpublished data, non–English-language studies, abstracts, or conference proceedings. Abstracts were manually screened, using *Abstrackr*,[20] on the basis of the eligibility criteria and exclusions and were cross-checked by a second member of the team. Articles whose abstracts were relevant were retrieved in full text for detailed screening to determine eligibility. During full-text screening, equivocal articles were screened 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.

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

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

- KQ1: measures of diagnostic accuracy, such as sensitivity, specificity, predictive values, likelihood ratios, or area under the receiver operating characteristics curve;
- KQ2: progression to MM;
- KQ3: timing, duration, and type of treatment;
- KQ4: overall survival, disease-free survival, response to treatment or remission (categorized as partial, complete, or stringent complete on the basis of treatment-induced decline in M-protein or FLC concentrations[14,21]), light chain escape, or quality of life; and
- KQ5: clinic visits, hospital stays, bone marrow biopsies, or bone scans.

Studies could have any length of followup[14,21] or any setting (primary or specialty care, in-facility or home, inpatient or outpatient).

Data Extraction and Quality Assessment

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.

We assessed the methodological quality of the studies using standard AHRQ EPC methodology with a three-level grading system (A, B, or C). This system involves a generic grading scheme that is applicable to varying study designs including RCTs, 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).

Characteristics of a quality A study include having 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. Quality A studies are those judged to have the least likelihood of bias and are considered valid. Quality C studies have a substantial risk of bias and may not be valid. 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 and Analysis

We summarized all included studies in narrative form and in summary tables that succinctly describe the important study characteristics and their findings. We conducted mainly descriptive analyses.[22] 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

We followed the AHRQ *Methods Guide*[23] to grade the strength of the body of evidence for each KQ, with modifications, on the basis of on 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.

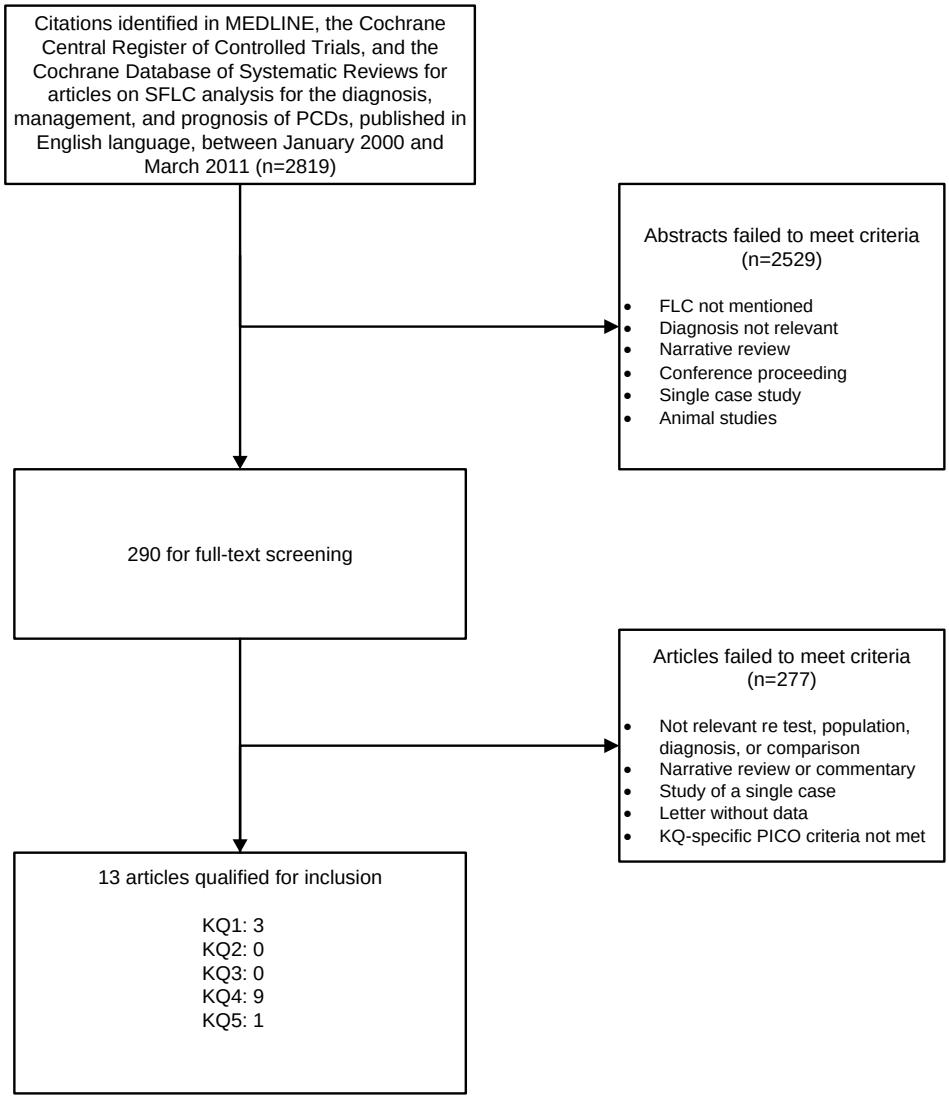
Results

Literature Search

The literature search yielded 2819 citations (**Figure 2**). From these, 290 articles were retrieved for full-text screening on the basis of abstracts and titles. Thirteen articles were judged to have met the inclusion criteria. A total of 277 articles were rejected on full-text screening, 41 of which underwent in-depth review before being rejected because they did not meet one or more of the PICO criteria for a particular KQ (see Appendix B for the list of rejected articles and the rationale for their rejection).

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], 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.

Figure 2. Summary of search and selection of articles



FLC = free light chain; KQ = Key Question; PCD = plasma-cell dyscrasia; PICO = population, intervention, comparator, and outcome; SFLC = serum free light chain.

Study Quality Grade and Overall Strength of Evidence

Table 1 summarizes the relevance and quality of the 13 studies reviewed in detail. The studies are organized by which KQ they addressed and the quality grade they were assigned.

Table 1. Distribution and quality of the 13 studies addressing a KQ

	KQ1	KQ2	KQ3	KQ4	KQ5	TOTAL
Quality A	0	0	0	0	0	0
Quality B	3	0	0	3	0	6
Quality C	0	0	0	6	1	7
Total no. of studies addressing KQ	**3**	**0**	**0**	**9**	**1**	**13**
Overall strength of the evidence	Insufficient	Insufficient	Insufficient	Insufficient	Insufficient	

KQ = Key Question

Findings for KQs

The main findings of this comparative effectiveness review are presented below for each KQ.

KQ1: Does adding the SFLC assay and the kappa/lambda ratio to traditional testing, compared with traditional testing alone, improve the diagnostic accuracy for PCDs in undiagnosed patients suspected of having a PCD?

Three studies[24-26] (all quality B) evaluated the addition of SFLC testing to traditional testing for the diagnosis of PCDs in undiagnosed patients suspected of having a PCD. All three were retrospective studies and compared test results to the diagnosis of disease verified by medical records on the basis of a panel of criteria. The addition of the SFLC assay to traditional testing increased the sensitivity for detection of PCDs by 6 to 36 percent. None of the three studies examined the statistical significance of the increase in sensitivity noted by the addition of the SFLC assay to the diagnostic panel. Given the heterogeneity in the designs, there was limited scope for data synthesis. We rate the evidence to evaluate the effect of adding the SFLC 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, 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?

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, is the SFLC assay better than traditional tests in indicating how the patient responds to treatment and of outcomes?

Nine studies (three quality B, six quality C) evaluated the SFLC assay and traditional testing in parallel and examined their relationship to clinical outcomes in PCDs.[16,27-34] The clinical outcomes addressed included treatment response and survival. Three studies of AL amyloidosis evaluated response to treatment as a predictor of outcomes[16,31,33]; the other six studies were of MM, evaluating either baseline concentrations or responses to treatment of SFLC and M-protein as predictors of clinical outcomes.[27-30,32,34] However, no direct comparisons between the SFLC assay and traditional tests (electrophoresis or IFE) were performed.

The three retrospective studies in AL amyloidosis (all quality C) showed that, relative to traditional testing, patients with greater reductions in abnormal SFLC concentrations (a >50 percent reduction[33] or >90 percent reduction,[16,31] vs. lesser reductions) after treatment with chemotherapy or stem-cell transplantation had better survival outcomes. The relationship

between quantitative reduction in M-protein and outcomes was inconsistent across studies, however. The presence of patients without "measurable disease," (i.e., patients with a PCD but without elevated SFLC concentrations before treatment, therefore precluding use of the SFLC assay as a marker of disease and treatment response) limited the utility of the SFLC assay in these studies.

Of the six studies that enrolled patients with MM (three quality B, three quality C), response to treatment was reported in five studies,[27-30,34] and two studies reported survival outcome.[32,34] Treatment given included induction therapy, chemotherapy, or stem-cell transplantation. The traditional-test comparators differed in each study (i.e., SPEP, UPEP, total kappa/lambda ratio measured by nephelometry, or standard response criteria [e.g., from IMWG]). Across the studies, an SFLC response occurred earlier than either an M-protein response or overall response and predicted an overall response. Two studies that addressed changes in SFLC or M-protein with treatment in relation to survival, showed conflicting results, one[28] showing no relationship and the other[34] showing that a greater percentage reduction in SFLC concentrations but not serum or urine M-protein after treatment was associated with better survival. Two studies examined the relationship of baseline SFLC concentrations and survival; one reported that both SFLC ratio and scores on Durie–Salmon staging and the International Staging System were independent predictors of survival,[32] but the other found that the SFLC ratio but not the serum or urine M-protein concentration significantly predicted survival.[34]

In summary, although SFLC response to therapy appeared to predict outcomes in patients with AL amyloidosis, no studies compared SFLC with traditional tests. Similarly, there was no evidence to ascertain whether SFLC response was a better predictor of outcomes than traditional tests in patients with MM. For both MM and AL amyloidosis, we rated the strength of evidence as insufficient for treatment response or baseline concentration of SFLC, compared with M-protein, as a better predictor of survival.

KQ5: In patients with an existing diagnosis of PCD, does the use of the SFLC assay reduce the need for other interventions?

We identified one retrospective cohort study (quality C) assessing the need for bone marrow examination if the SFLC assay was used to define the completeness of response to treatment: Chee 2009.[35] The patients had a negative IFE test after treatment of MM and a concomitant evaluable bone marrow biopsy. 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. Owing to the preliminary nature of the data, we rated the evidence as insufficient for addressing this question.

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, and relapse. In the present review, we assessed the role of the SFLC assay as an adjunct to traditional tests such as SPEP and SIFE for the diagnosis of PCDs in populations suspected of having the disease. We also ascertained the ability of the SFLC assay, 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 interventions.

Our results reveal that there continues to be a paucity of evidence to clarify the effectiveness and role of the SFLC assay for the diagnosis, management, and prognosis of PCDs. Thirteen studies, less than 1 percent of the 2819 studies identified in our literature search, met the inclusion criteria to address the KQs (see Appendix B for the excluded studies). Moreover, although in the 13 studies the SFLC assay and traditional testing were commonly conducted in parallel, they were not formally compared. Several articles evaluating the effectiveness and role of the SFLC assay were excluded because the populations were not of interest or there was a lack of data for traditional testing as a comparator. The study heterogeneity observed may also reflect the assay's uncertain role in research and clinical practice; across the studies, there was considerable clinical heterogeneity, with variation in study design and population as well as inconsistency in the comparisons being made. Moreover, the studies were of either poor or fair quality, not good quality. This, in combination with the study heterogeneity, limited the validity of the studies and the conclusions that could be drawn from them.

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, 4, and 5) and expand in detail the major needs of future studies.

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.[24-26] The studies were all retrospective, conducted in a hospital laboratory setting, and were of adults who were suspected to have a monoclonal gammopathy. They used archived laboratory samples that had been received for SPEP or UPEP. All three studies indicated that the addition of SFLC to traditional tests increased diagnostic sensitivity (by 6 to 36 percent), although the effect on diagnostic specificity was variable.

The limitations and potential biases in these studies limit their utility for informing clinical practice. We found that demographic details, including racial breakdown and comorbid conditions, were underreported. Overall 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.[14] Positive tests would be followed with more detailed testing, including IFE and bone marrow examination.[7,8] Therefore, ultimately the diagnosis 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 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. Given that PCDs are a heterogeneous group of conditions, the production of monoclonal light chains is 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. 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 diagnosis or test group to diminish the likelihood of misclassification bias. However, such safeguards were seldom emphasized in the studies reviewed.

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.

SFLC assay and treatment response and survival (KQ4)

Nine studies, three in patients with AL amyloidosis[16,31,33] and six in patients with MM,[27-30,32,34] evaluated SFLC testing compared with traditional tests 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. None directly compared the SFLC assay with traditional testing. Study quality was either B or C; 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.

The strength of evidence for this KQ was insufficient for both AL amyloidosis and MM for all outcomes examined. Nonetheless, there were several limitations in the literature reviewed. 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 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 limit the applicability of the findings to both the general PCD population and subgroups of interest.

SFLC assay in reducing the need for other interventions (KQ5)

A single study, Chee 2009,[35] explored whether use of the SFLC assay compared with traditional testing would reduce the need for other diagnostic interventions. The authors evaluated whether a negative IFE result or normalization of the SFLC ratio (or both) after treatment of MM would eliminate the need for bone marrow biopsy to stringently define remission. 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, however, this question requires more detailed and systematic evaluation in practice.

Future Research

It appears that there is considerable clinical uncertainty regarding the applications of the SFLC assay both within and beyond the 2009 IMWG consensus guidelines.[14] 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 in 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.

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, would yield 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.

Several limitations of the studies reviewed make it difficult to present clear conclusions regarding the comparative effectiveness of the SFLC assay and should be targeted in future studies. Diagnostic studies using only data from patients already diagnosed with PCDs were excluded from this review (see Appendix B). The studies of only patients with disease reflected the extreme end of the spectrum of disease severity, and the proportion of patients with a positive test was likely overestimated. Moreover, without a nondiseased population, true negatives could not be assessed. Studies could not avoid incorporation bias because the result from the reference test itself (e.g., SPEP or SIFE) was considered along with other factors, such as clinical information, to reach a diagnosis of PCD. 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. Most studies did not report whether data assessors were blinded to diagnosis or 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.

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 variability in SFLC concentrations, and the harms of testing, questions that were outside the scope of this review. In addition, the lack of a suitable gold 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.

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 for 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. Although it is well appreciated that MM is more common among African Americans than among whites, this source of variability was not routinely considered in the studies we reviewed. Future studies of SFLC 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.

References

1. Kyle RA, Rajkumar SV. Epidemiology of the plasma-cell disorders. Best Pract Res Clin Haematol. 2007;20:637-64

2. Ries LAG, Harkins D, Krapcho M, et al. SEER cancer statistics review, 1975-2003. NCI, Bethesda, MD; 2005. Available at: http://seer.cancer.gov/csr/1975_2003.

3. Jemal A, Siegel R, Ward E, et al. Cancer Statistics, 2007. CA: A Cancer Journal for Clinicians 2007;57(1):43-66. PMID: 10.3322/canjclin.57.1.43.

4. Katzel JA, Hari P, Vesole DH. Multiple Myeloma: Charging Toward a Bright Future. CA: A Cancer Journal for Clinicians 2007;57(5):301-18. PMID: 10.3322/CA.57.5.301.

5. Cook R. An economic perspective on treatment options in multiple myeloma. Managed Care Oncol 2007;2007(Spring):10-12

6. Messori A, Trippoli S, Santarlasci B. Pharmacotherapy of multiple myeloma: an economic perspective. Expert Opin Pharmacother 2003 Apr 1;4(4):515-24. PMID: doi: 10.1517/14656566.4.4.515.

7. Kyle RA, Rajkumar SV. Criteria for diagnosis, staging, risk stratification and response assessment of multiple myeloma. Leukemia 2009;23(1):3-9

8. Palumbo A, Sezer O, Kyle R, et al. International Myeloma Working Group guidelines for the management of multiple myeloma patients ineligible for standard high-dose chemotherapy with autologous stem cell transplantation. Leukemia 2009 Jun 4;23(10):1716-30

9. Rajkumar SV. Clinical features, laboratory manifestations, and diagnosis of multiple myeloma. In: Basow DS (eds). UpToDate. Waltham: UpToDate; 2011.

10. Bradwell AR, Carr-Smith HD, Mead GP, et al. Highly sensitive, automated immunoassay for immunoglobulin free light chains in serum and urine. Clin Chem 2001 Apr;47(4):673-80

11. Bradwell AR, Carr-Smith HD, Mead GP, et al. Highly Sensitive, Automated Immunoassay for Immunoglobulin Free Light Chains in Serum and Urine. Clin Chem 2001 Apr 1;47(4):673-80

12. Rajkumar SV. Recognition of monoclonal proteins. In: Kyle RA, Connor RF (eds). UpToDate. Waltham: UpToDate; 2011.

13. Katzmann JA. Serum free light chains: quantitation and clinical utility in assessing monoclonal gammopathies. Clin Lab News 2006 Jun;June:12-14

14. Dispenzieri A, Kyle R, Merlini G, et al. International Myeloma Working Group guidelines for serum-free light chain analysis in multiple myeloma and related disorders. [Review] [43 refs]. Leukemia 2009 Feb;23(2):215-24

15. Mead GP, Carr-Smith HD, Drayson MT, et al. Serum free light chains for monitoring multiple myeloma. British Journal of Haematology 2004 Aug;126(3):348-54

16. Sanchorawala V, Seldin DC, Magnani B, et al. Serum free light-chain responses after high-dose intravenous melphalan and autologous stem cell transplantation for AL (primary) amyloidosis. Bone Marrow Transplantation 2005 Oct;36(7):597-600

17. Kyle RA, Durie BGM, Rajkumar SV, et al. Monoclonal gammopathy of undetermined significance (MGUS) and smoldering (asymptomatic) multiple myeloma: IMWG consensus perspectives risk factors for progression and guidelines for monitoring and management. Leukemia 2010 Jun;24(6):1121-27

18. Bradwell AR. Serum free light chain measurements move to center stage. Clin Chem 2005 May;51(5):805-07

19. Whitlock EP, Lopez SA, Chang S, et al. AHRQ Series Paper 3: Identifying, selecting, and refining topics for comparative effectiveness systematic reviews: AHRQ and the Effective Health-Care program. J Clin Epidemiol. 2010;63:491-501

20. Wallace BC, Trikalinos TA, Lau J, et al. Semi-automated screening of biomedical citations for systematic reviews. BMC Bioinformatics 2010;11:55

21. Durie BG, Harousseau JL, Miguel JS, et al. International uniform response criteria for multiple myeloma.[Erratum appears in Leukemia. 2007 May;21(5):1134].[Erratum appears in Leukemia. 2006 Dec;20(12):2220]. Leukemia 2006 Sep;20(9):1467-73

22. 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

23. Agency for Healthcare Research and Quality. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. 2011

24. Abadie JM, Bankson DD. Assessment of serum free light chain assays for plasma cell disorder screening in a Veterans Affairs population. Annals of Clinical & Laboratory Science 2006;36(2):157-62

25. Piehler AP, Gulbrandsen N, Kierulf P, et al. Quantitation of serum free light chains in combination with protein electrophoresis and clinical information for diagnosing multiple myeloma in a general hospital population. Clin Chem 2008 Nov;54(11):1823-30

26. Vermeersch P, Van HL, Delforge M, et al. Diagnostic performance of serum free light chain measurement in patients suspected of a monoclonal B-cell disorder. British Journal of Haematology 2008 Nov;143(4):496-502

27. Dispenzieri A, Zhang L, Katzmann JA, et al. Appraisal of immunoglobulin free light chain as a marker of response. Blood 2008 May 15;111(10):4908-15

28. Giarin MM, Giaccone L, Sorasio R, et al. Serum free light chain ratio, total kappa/lambda ratio, and immunofixation results are not prognostic factors after stem cell transplantation for newly diagnosed multiple myeloma. Clin Chem 2009 Aug;55(8):1510-16

29. Khoriaty R, Hussein MA, Faiman B, et al. Prediction of response and progression in multiple myeloma with serum free light chains assay: corroboration of the serum free light chain response definitions. Clinical lymphoma, myeloma & leukemia 2010 Feb;10(1):E10-E13

30. Kroger N, Asenova S, Gerritzen A, et al. Questionable role of free light chain assay ratio to determine stringent complete remission in multiple myeloma patients. Blood ;115(16):3413-14

31. Kumar SK, Dispenzieri A, Lacy MQ, et al. Changes in serum-free light chain rather than intact monoclonal immunoglobulin levels predicts outcome following therapy in primary amyloidosis. American Journal of Hematology 2011 Mar;86(3):251-55

32. Kyrtsonis MC, Vassilakopoulos TP, Kafasi N, et al. Prognostic value of serum free light chain ratio at diagnosis in multiple myeloma. British Journal of Haematology 2007 May;137(3):240-43

33. Lachmann HJ, Gallimore R, Gillmore JD, et al. Outcome in systemic AL amyloidosis in relation to changes in concentration of circulating free immunoglobulin light chains following chemotherapy. British Journal of Haematology 2003 Jul;122(1):78-84

34. van RF, Bolejack V, Hollmig K, et al. High serum-free light chain levels and their rapid reduction in response to therapy define an aggressive multiple myeloma subtype with poor prognosis. Blood 2007 Aug 1;110(3):827-32

35. Chee CE, Kumar S, Larson DR, et al. The importance of bone marrow examination in determining complete response to therapy in patients with multiple myeloma. Blood 2009 Sep 24;114(13):2617-18

36. Reitsma JB, Rutjes AWS, Whiting P, et al. Chapter 9: Assessing methodological quality. In: Deeks JJ, Bossuyt PM, Gastonis C (eds). Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy (v.1.0.0). The Cochrane Collaboration; 2009. p. 1-24

Acronyms

AAAC	American Association for Clinical Chemistry
AHRQ	Agency for Healthcare Research and Quality
AL amyloidosis	Systemic, or primary, amyloidosis in which amyloid [A] proteins derived from immunoglobulin light chains [L] are deposited in tissue (also called light-chain amyloidosis)
CER	Comparative Effectiveness Review
CI	confidence interval
EPC	Evidence-based Practice Center
FDA	Food and Drug Administration
FLC	Free light chain
IFE	Immunofixation electrophoresis
IMWG	International Myeloma Working Group
LCMM	Light chain multiple myeloma
MGUS	Monoclonal gammopathy of undetermined significance
MM	Multiple myeloma
M-protein	Monoclonal protein (also called paraprotein)
NSMM	Nonsecretory multiple myeloma
PCD	Plasma-cell dyscrasia
KQ	Key Question
PICOTS	Populations, interventions, comparators, outcomes, timing, and settings
RCT	randomized controlled trial
SFLC	Serum free light chain
SIFE	Serum immunofixation electrophoresis
SMM	Smoldering multiple myeloma
SPEP	Serum protein electrophoresis
TEP	Technical Expert Panel
TOO	Task Order Officer
UIFE	Urine immunofixation electrophoresis
UPEP	urine protein electrophoresis

Glossary

Term	Definition	Source, if applicable
Cast nephropathy (or “myeloma kidney”)	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	
Differential verification	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.	Reitsma 2009[36]
Disease progression or recovery bias	Bias from an inappropriately long interval (or any interval) between conduct of reference test and conduct of index test.	Reitsma 2009[36]
Incorporation bias	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).	Reitsma 2009[36]
Involved FLC or SFLC	The free light chain or serum free light chain that is produced in excess and is causing disease (either kappa or lambda)	
Light chain escape	A type of PCD remission in which, for unclear reasons, a subclone of malignant plasma cells	Dispenzieri 2009[14]

	expands that is incapable of producing significant amounts of immunoglobulin heavy chain but retains the ability to make light chains.	
Measurable disease	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	
M-protein or paraprotein	Intact immunoglobulins or FLCs of a single type produced in excess by an abnormally expanded clone of malignant plasma cells (biomarkers of PCDs)	
Monoclonal gammopathy	Disease class comprising PCDs as well as other conditions such as hematological disorders associated with a monoclonal band	
Oligosecretory MM	MM in which very small amounts of M-protein are produced by the malignant plasma cells	
Polyclonal gammopathy	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	
Selection bias (also called partial verification bias, workup bias, or sequential ordering bias)	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.	Reitsma 2009[36]
SFLC ratio	The ratio of kappa chains to lambda chains, for which the normal range is 0.26–1.65.	Katzmann 2006[13]
Spectrum bias or effect	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.)	Reitsma 2009[36]
Verification bias	Bias in appraising performance of a diagnostic index test introduced through the inaccuracy of the reference standard.	Reitsma 2009[36]

Introduction

Plasma-cell dyscrasias (PCDs) are a group of clonal 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 produce immunoglobulins that constitute the body's normal humoral immune response. Immunoglobulins consist of two heavy chains of the same type (gamma, alpha, delta, epsilon, or mu) and two light chains of the same type (either kappa or lambda). The intact molecules are named according to which types of chains they contain (e.g., IgG-kappa is an immunoglobulin [Ig] consisting of two gamma [G] heavy chains and two kappa light chains). Immunoglobulins are normally either secreted into the circulation as free antibodies or are found bound to the plasma-cell membrane, serving as antigen receptors. In addition, 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 diagnosis of 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 with organ dysfunction.

The International Myeloma Working Group (IMWG) recommend the following actions and tests for evaluation of a patient suspected of having a myeloma[7,8]: a complete history taking and physical examination; routine laboratory testing including serum protein electrophoresis (SPEP), serum immunofixation (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 (UIFE), plus immunoglobulin quantification. The SFLC assay, a relatively recent development, is also now used to measure the FLC component in particular.

In general, for diagnosis, SPEP is estimated to detect an immunoglobulin peak in 82 percent of patients with MM.[9] The addition of SIFE increases the sensitivity to 93 to 95 percent,[4,5] which is further increased to 97 percent by performing UPEP and UIFE.[9] The 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 three percent of cases of NSMM, LCMM, or AL amyloidosis are not detected by traditional tests.[9] 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.

The SFLC assay (the Freelite™ Assay, The Binding Site Ltd., Birmingham, United Kingdom) was introduced in 2001 to measure the FLC component in particular.[10,11] The SFLC 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 (i.e., 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, with a sensitivity of less than 1 mg/dL in serum and less than 200 mg/day in urine[12] for detecting low concentrations of FLCs. The other main advantage is the ability to measure the ratio of kappa chains to lambda chain, for which the diagnostic range is 0.26–1.65.[13] This enables the identification of the FLC that is produced in excess 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. An abnormal ratio provides a useful index of clonality, as clonal disorders produce disproportionately high concentrations of a single type of light chain.

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 International Myeloma Working Group (IMWG) currently considers the SFLC assay to be an adjunct to traditional tests.[14] The assay could allow for quantitative monitoring of response and remission after treatment and provide prognostic information[15,16]; potentially reduce 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[14]; and 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,[13,14] although proper studies have yet to be undertaken to explore this potential benefit. 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[17]; 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 its short half-life (2 to 6 hours, vs. 21 days for complete immunoglobulins[18]), it may also help detect relapses and treatment failures earlier than by reliance on M-protein concentrations alone. These diagnostic applications need to be validated and standardized, however.[10]

Recommendations from professional societies now usually include the use of the SFLC assay in some capacity but the breadth of recommended use varies widely. For example, the National Comprehensive Cancer Network Clinical Practice Guidelines in Oncology—Multiple Myeloma were updated in January 2011 in part to change the description of the SFLC assay from "useful under some circumstances" to recommended for use in "initial diagnostic workup."[19] In addition, the European Group for Blood and Bone Marrow Transplant (EBMT)/International Bone Marrow Transplant Registry (IBMTR)/American Bone Marrow Transplant Registry (ABMTR) criteria from 1998 have been updated by the IMWG, as of 2009, to include use of the SFLC assay for the diagnosis of MM and related PCDs in several circumstances.[10,13] However, these guidelines are not based on systematic review of the available evidence.

Thus, although the SFLC assay has been in use for a decade, it remains unclear how best to implement it into clinical practice.[20]

Key Questions

The objective of this report is to summarize the existing literature regarding the role of SFLC testing in the diagnosis, management, and prognosis of patients with PCDs, specifically in response to the following key questions. The Key Questions (KQs) were formulated in consultation with AACC and the Agency for Healthcare Research and Quality (AHRQ).

1. 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?

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

3. 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)?

4. 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)?

5. 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 interventions (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*[21] and *Methods Guide for Medical Test Reviews*.[22]

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 [REDACTED] 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

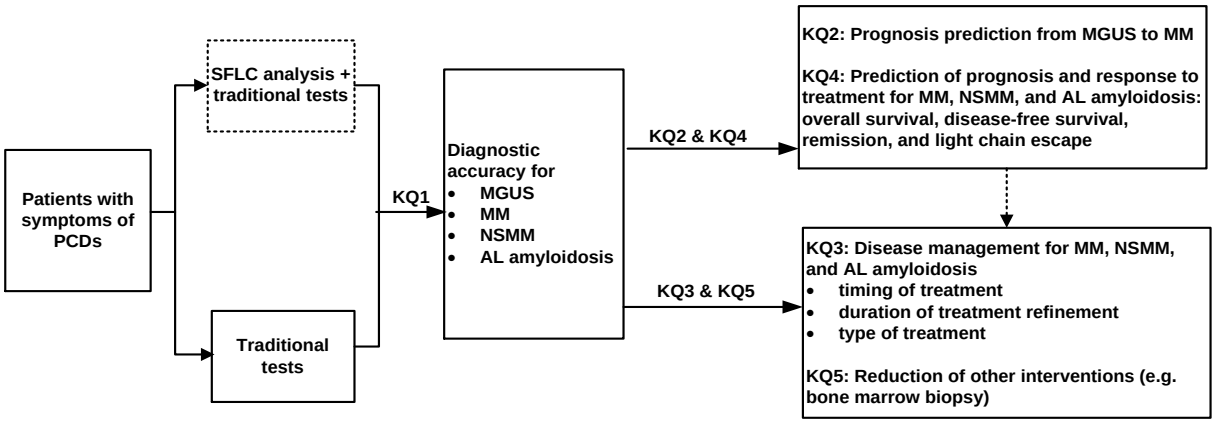
During a topic refinement phase, the initial questions were refined with input from a panel of Key Informants: topic nominators from AAAC; experts in renal amyloidosis, clinical chemistry, and general internal medicine and geriatrics; patient advocates; and representatives from the Centers for Medicare 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 multiple myeloma and/or amyloidosis, clinical chemistry, and general medicine to help refine KQs, identify important issues, define parameters for the review, and provide expert input to technical questions arose during the review. The Preface lists the Key Informants and members of the TEP; the title page lists the local domain expert, who provided names of experts and gave input into framing of the KQs. Discussions among the [REDACTED] EPC and the TOO, the Key Informants, and the TEP occurred during a series of teleconferences and via email.

Analytic Framework

KQs take into account the patient populations, interventions (or diagnostic tests), comparators, outcomes, timing, and settings (PICOTS) that are clinically relevant to the use of the SFLC analysis. Five KQs are addressed in the present report. These pertain to the diagnosis of PCDs, prognosis (i.e., progression from MGUS to MM and 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 interventions (e.g., bone marrow biopsy), respectively.

To guide the development of the KQs, we developed an analytic framework (**Figure 1**) that maps the specific linkages associating the population (patients with PCD symptoms) to the additional intervention (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 interventions, 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 gammopathy 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, to March 11, 2011, in MEDLINE®, the Cochrane Central Register of Controlled Trials, and the Cochrane Database of Systematic Reviews. The start year was chosen as 2000 because the SFLC assay was approved by the FDA in 2001; we considered that 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 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*,[23] on the basis of the eligibility criteria and exclusions and were cross-checked 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 screening to determine eligibility. During full-text screening, equivocal articles were screened 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.

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

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

- KQ1: measures of diagnostic accuracy, such as sensitivity, specificity, predictive values, likelihood ratios, or area under the receiver operating characteristics curve;
- KQ2: progression to MM;
- KQ3: timing, duration, and type of treatment;
- KQ4: overall survival, disease-free survival, response to treatment or remission (categorized as partial, complete, or stringent complete on the basis of treatment-induced decline in M-protein or FLC concentrations[14,24]), light chain escape, or quality of life; and
- KQ5: clinic visits, hospital stays, bone marrow biopsies, or bone scans.

Studies could have any length of followup[14,24] 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 experienced 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.

Quality Assessment

For quality assessment, we used predefined methods for evaluating study quality that are common within the EPC Program.[21,25,26] 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 RCTs, 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, 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; 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, interventions, outcomes, results, and study quality. We included diagnostic performance parameters, risk estimates, and their 95 percent CI and P value where applicable. Results are presented in separate summary tables for each KQ.

We conducted mainly descriptive analyses[27] 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 for each KQ, with modifications, on the basis of on our level of confidence that the evidence reflected the true effect for the major comparisons of interest. The strength of evidence (mostly a measure of risk

of bias) 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 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.

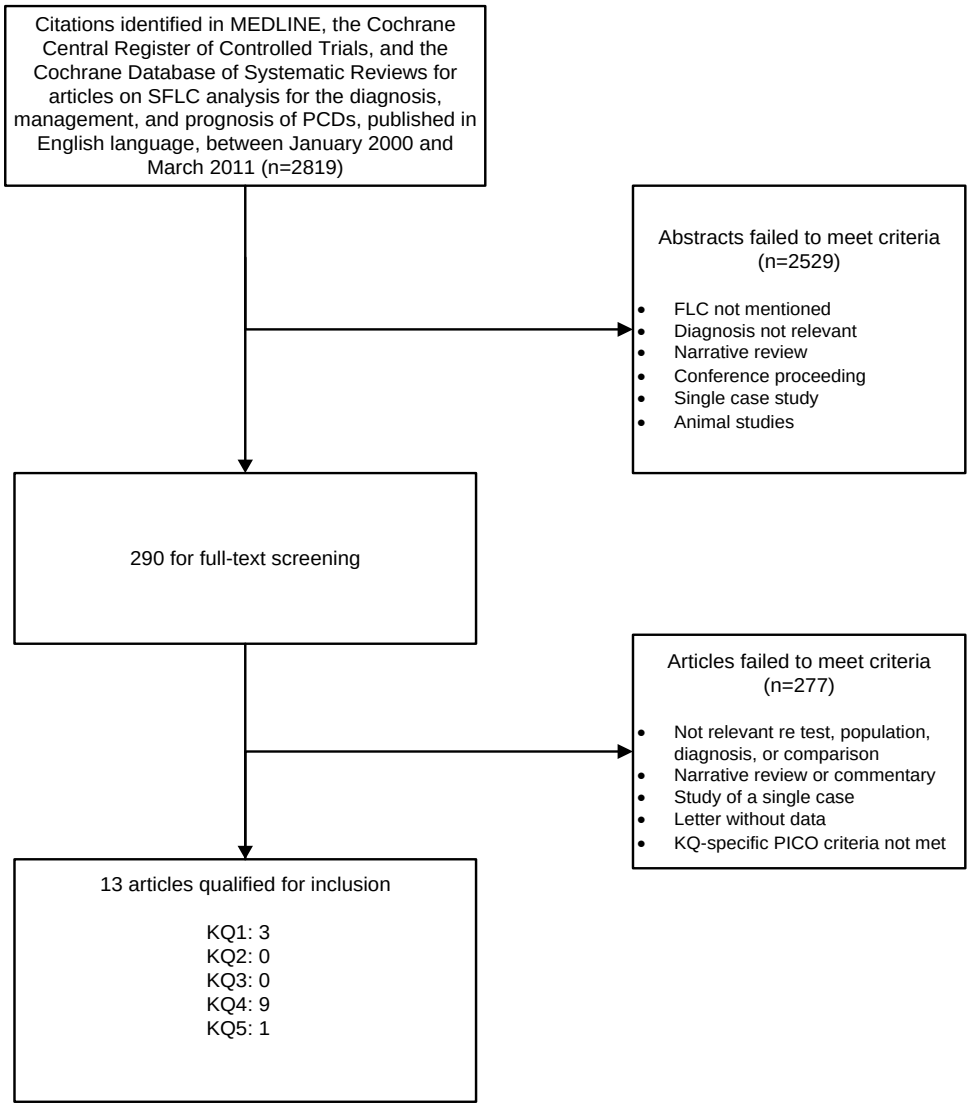
Results

Literature Search

The literature search yielded 2819 citations (**Figure 2**). From these, 290 articles were retrieved for full-text screening on the basis of abstracts and titles. Thirteen articles were judged to have met the inclusion criteria. A total of 277 articles were rejected on full-text screening, 41 of which underwent in-depth review before being rejected because they did not meet one or more of the PICO criteria for a particular KQ (see Appendix B for the list of rejected articles and the rationale for their rejection).

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.

Figure 2. Summary of search and selection of articles



FLC = free light chain; KQ = Key Question; PCD = plasma-cell dyscrasia; PICO = population, intervention, comparator, and outcome; SFLC = serum free light chain.

Study Quality Grade and Overall Strength of Evidence

Table 1 summarizes the relevance and quality of the 13 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 1. Distribution and quality of the 13 studies addressing a KQ

	KQ1	KQ2	KQ3	KQ4	KQ5	TOTAL
Quality A	0	0	0	0	0	0
Quality B	3	0	0	3	0	6
Quality C	0	0	0	6	1	7
Total no. of studies addressing KQ	**3**	**0**	**0**	**9**	**1**	**13**
Overall strength of the evidence	Insufficient	Insufficient	Insufficient	Insufficient	Insufficient	

KQ = Key Question

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[28-30] 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.

All three studies were retrospective and compared test results to the diagnosis of disease verified by medical records on the basis of a panel of criteria. Each study was rated B quality because of the presence of biases and lack of statistical testing to quantify uncertainty. One study reported industry-associated resources and also was the only study of the three to report the demographic characteristics of the study population [28]

Abadie 2006[28] 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 SFLC increased the sensitivity to 1.00 and the specificity to 0.99, with four false negatives (two for MM and two for "potential MM").

Piehler 2008[29] 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[30] 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 studies addressed KQ1, evaluating the SFLC assay in combination with traditional tests in undiagnosed patients suspected of having a PCD. The addition of the SFLC assay to traditional testing increased the sensitivity for detection of PCDs by 6 to 36 percent. None of the three studies examined the statistical significance of the increase in sensitivity noted by the addition of the SFLC assay to the diagnostic panel. Given the heterogeneity in the designs, there was limited scope for data synthesis. We rate the evidence to evaluate the effect of adding the SFLC to traditional testing on diagnostic performance as insufficient.

Table 2. Characteristics of studies addressing KQ1

Author Year [PMID]	Index test/ Comparator test	Sample size	Funding	Enrollment period	Prospective study?	Diagnosis documented in medical records	Quality grade and issues
Abadie 2006[28] [16682511]	SFLC+SPEP SPEP	312	Kit/reagents provided by industry	2004–2005	No	Medical records	B (no measure of statistical uncertainty, consecutive sampling, no major biases)
Piehler 2008[29] [18801937]	SFLC+SPEP SPEP	489	nd	2005–2006	No	Medical records	B (no measure of statistical uncertainty, consecutive recruitment, no major biases)
Vermeersch 2008[30] [18729849]	SFLC+SIFE, SFLC+SIFE+SPEP SIFE, SIFE+SPEP	833	None	2004–2006	No	Medical records	B (no measure of statistical uncertainty, well-described sample, no major biases)

SFLC = serum free light chain [note this can refer to the light chain itself or the assay], KQ = Key Question, SPEP = serum protein electrophoresis, UPEP = urine protein electrophoresis, IFE = immunofixation electrophoresis, UIFE = urine immunofixation electrophoresis, SIFE = serum immunofixation electrophoresis

Table 3. Characteristics of patients in studies addressing KQ1

Author Year [PMID]	Sample size	Enrollment method	Diagnosed before study?	PCD prevalence	Age (yr)	Percent Male	Treated?
Abadie 2006[28] [16682511]	312	Consecutive	No	Malignant PCD, 15%	67 (mean)	97	nd
Piehler 2008[29] [18801937]	489	Selection if SPEP performed	No	Any PCD, 27% LCMM, 2.1% MM, 6.6% AL amyloidosis, 0.6% MGUS, 13.6%	nd	nd	nd
Vermeersch 2008[30] [18729849]	833	Consecutive	No	Malignant PCD, 3%; MGUS, 19%	nd	nd	NA

Table 4. Results of studies addressing KQ1

Author Year [PMID]	Sample size	Diagnosis	Index test	Sensitivity	Specificity	Other results
Abadie 2006[28]	312	PCD	SFLC	0.88	0.98	Considered MGUS as false positive

[16682511]			SPEP	0.64	0.81	
			SFLC+SPEP	1.00	0.99	
Piehler 2008[29] [18801937]	332	PCD	SFLC	0.66	0.78	Specificity was affected by SFLC positivity in patients with other hematological diagnosis or decreased kidney function
			SPEP	0.87	0.98	
			SFLC+SPEP	0.96	0.78	
Vermeersch 2008[30] [18729849]	833	Monoclonal gammopathy*	SFLC	0.37	0.97	Missed 3 MM, 1 plasmacytoma, 112 MGUS cases
			SIFE	0.92	1.00	Missed 2 MGUS cases
			SFLC+SIFE	0.94	0.97	Missed1 MGUS cases
			SPEP (+SIFE for confirmation)	0.79	1.00	Missed 1 MM, 1 AL-amyloidosis, 1 plasmacytoma, 26 MGUS cases
			SFLC+ SPEP (+SIFE for confirmation)	0.82	0.97	Missed 1 plasmacytoma, 23 MGUS cases
			UIFE+ SPEP (+SIFE for confirmation)	0.82	1.00	Missed 24 MGUS cases
			SIFE+UIFE	0.92	1.00	Missed 2 MGUS cases

SIFE = serum immunofixation, xx

*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

Nine studies evaluated the SFLC assay and traditional testing in parallel and examined their relationship to clinical outcomes in PCDs.[16,31-38] Two studies[35,37] reported industry-associated funding or authorship. Eight studies were retrospective[16,31-33,35-38]; the remaining study[34] lacked enough detail to determine the study design. Followup times varied from 3 months to 13 years, with sample sizes of 52 to 443 patients. Among studies reporting patient characteristics, median age ranged from 54 to 64 years and the study populations were 45 to 65 percent male.

All nine studies provided outcomes data on SFLC testing and traditional testing conducted in parallel—no direct comparisons between the SFLC assay and traditional tests were performed. Three studies were conducted in patients with AL amyloidosis[16,35,37] and six in patients with MM.[31-34,36,38] Three studies were graded as quality B and the remaining six as quality C, owing to limitations in study design, including selection bias and small sample sizes.

Patients with AL Amyloidosis

We identified three studies addressing KQ4 that examined the SFLC assay in patients with AL amyloidoisis and reported the use of SFLC assay in evaluating treatment response and predicting prognosis: Kumar 2011,[16,35,37] Lachmann 2003,[16,35,37] and Sanchorawala 2005.[16,35,37] All three studies measured both SFLC responses and paraprotein responses to treatment by traditional testing (electrophoresis or IFE) and examined their relationship to outcomes. Paraprotein reduction was usually reported in the context of whether a hematologically complete response (which incorporates paraprotein response) was achieved.[39]

All three studies were retrospective, with sample sizes of 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, there were 61 percent and 63 percent men. All three studies were rated as quality C; none of the three studies performed direct statistical comparisons of relative strength of prediction, providing unadjusted estimates only for each predictor.

All three studies that, relative to traditional testing, patients with greater reductions in abnormal SFLC concentrations (a >50 percent reduction[37] or >90 percent reduction,[16,35] vs. lesser reductions) after treatment (either chemotherapy or stem-cell transplantation) had better survival outcomes.

Although Kumar 2011[35] did not find quantitative paraprotein concentrations (in contrast with SFLC concentrations) a good predictor, Lachmann 2003[37] found the concentration to be significantly related to survival; however, the relationship seemed to be weaker than that of SFLC reduction to survival. In Kumar 2011 in particular, the presence of patients without "measurable disease," (i.e., patients with a PCD but without elevated SFLC concentrations before treatment, therefore precluding use of the SFLC assay as a marker of disease and treatment response) limited the utility of the SFLC assay. Sanchorwala 2005[16] found that both an SFLC reduction by more than 90 percent and the achievement of a complete response were predictive of a lower mortality and were complementary, both providing independent predictive information.

Summary for AL Amyloidosis

Overall, the three studies addressing KQ4 in patients with AL amyloidosis suggest that the SFLC assay may aid in assessing treatment response and monitoring outcomes in such patients—particularly because of the sensitivity of the assay to detect SFLC at low concentrations. The effectiveness of the SFLC assay compared with traditional tests (electrophoresis or IFE) for assessment of treatment response and outcome is inconclusive, as no direct comparisons were performed. The strength of evidence underlying this comparison was therefore rated as insufficient. Results from these studies should be interpreted with the caveat that all three were graded as quality C, owing to limitations in study design, including selection/spectrum bias as well as (in one study) small sample size.

Patients with MM

We identified six studies[31-34,36,38] that 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).

Five of the six—Dispenzieri 2008,[31] Giarin 2009,[32] Khoriaty 2010,[33] van Rhee 2007,[38] and Kyrtsonis 2007[36]—were retrospective analyses of cohorts; study design was not specified in the remaining study, Kroger 2010.[34] Sample size ranged from 43 to 303, and median followup duration was 3 months to 13 years. Study quality was graded as B in three of the six studies,[31,32,38] and C in the other three.[33,34,36] Response to treatment was reported in five studies,[31-34,38] and two studies reported survival outcome.[36,38]

Three studies addressed the relationship of outcomes to SFLC response to treatment.[31,33,34] They examined whether the magnitude of reduction of the SFLC concentrations after induction therapy, chemotherapy, or stem-cell transplantation affected the achievement of complete response or event-free or overall survival. The traditional test comparators differed in each study (i.e., SPEP, UPEP, total kappa/lambda ratio measured by nephelometry, or standard response criteria [e.g., from IMWG]). Across the studies, an SFLC response occurred earlier than either an M-protein response or an overall response and predicted an overall response. One small study[34] reported on a subgroup of 10 patients in whom 9 had an abnormal SFLC (kappa/lambda) ratio before relapse and a positive IFE test. The quality of this study was rated C, because of the limited information about study design, SFLC response definitions, and results. Two reasonably large studies, both graded B, gave discordant results for the relationship between SFLC results after treatment with stem-cell transplantation and survival. One study[32] did not find a significant relationship between a normal SFLC kappa/lambda ratio after treatment and overall or event-free survival but found that an abnormal (vs. normal) total kappa/lambda ratio by nephelometry was predictive of poorer event-free survival. However, contrary results were noted in the other study [38]: patients with a percent reduction of 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 percent reduction (after adjustment for serum lactate dehydrogenase concentration and cytogenetic abnormalities), despite a paradoxically better response to induction therapy. The authors did not find any relationship between the tertiles of reductions in serum and urine M-protein values and overall or event-free survival, however.

Two studies examined the relationship of baseline SFLC concentrations and survival; one included concomitant evaluation of the predictive ability of traditional testing (in the form of measurement of baseline concentrations of serum and urine M-protein),[38] and the other incorporated clinical staging systems (Durie–Salmon staging and the International Staging System [ISS]).[36] In the former study, the top tertile of SFLC concentrations (>75 mg/dL) were considered the risk category,[38] whereas in the latter study, patients were stratified according to whether the SFLC ratio was above or below the median (with the ratio calculated using the involved FLC in the numerator, for a monotonic distribution).[36] In both studies, higher concentrations of either SFLCs or the SFLC ratio were predictive of poorer overall or disease-specific survival compared with lower values. Again, whereas the former study did not find serum or urine M-protein concentrations to be predictive of survival, the latter study found that the disease stage was an independent predictor. This study followed 94 patients for 33 months and reported that while Durie–Salmon and ISS staging were independent predictors (both $p<0.0001$), the SFLC ratio was also significantly associated with 3- and 5-year disease-specific survival rates ($p=0.0001$).[36]

Summary for MM

Six 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. All of these studies were retrospective cohort studies of quality B or C. The studies were heterogeneous with respect to population, intervention, and comparator and the degree of adjustment for confounders. All these factors taken together, conclusions about the definitive use of the SFLC assay in prognosis prediction are limited, and the strength of evidence is rated as insufficient for comparisons with traditional testing in patients with MM.

Table 5. Characteristics of studies addressing KQ4

Author Year [PMID]	Index test/ Comparator test	Sample size	Funding	Study design	Enrollment period	Followup duration	Diagnostic criteria	Quality grade and issues
AL amyloidosis: SFLC response to therapy and relationship to outcomes								
Kumar, 2011[35] [21328431]	Post-treatment dFLC Post-treatment quantitative M-protein concentrations	443 Cohort I: 347 Cohort II: 96	Government, industry	Retrospective	nd	72 mo	Biopsy-proven AL amyloidosis	C (retrospective, extreme selection/spectrum bias)
Lachmann, 2003[37] [12823348]	Post-treatment SFLC concentration Post-treatment quantitative paraprotein concentrations	262	Government, author(s) employed by industry	Retrospective	1992–2002	21–29 mo	Immunohistochemically confirmed AL amyloidosis	C (retrospective, selection/spectrum bias, sample not uniformly treated)
Sanchorawala, 2005[16] [16044137]	Post-treatment SFLC concentration Hematological complete response (defined by EBMT[40]; includes M-protein response)	66	Government, academic	Retrospective	1994–2003	5 yr	Histological diagnosis of AL amyloidosis with evidence of PCD and eligibility for high-dose melphalan SCT treatment in clinical protocols	C (retrospective, small sample size)
MM SFLC response to therapy and relationship to outcomes								
Dispenzieri, 2008[31] [18364469]	SFLC response SPEP, UPEP	399	Government	Retrospective	1988–1992	13 yr	M-protein ≥10 g/L or urine monoclonal FLC >200 mg in 24 hr or serially measurable soft tissue plasmacytoma or bone marrow plasmacytosis ≥20%	B (retrospective without adjustment)
Giarin,2009[32] [19520760]	SFLC ratio Total kappa/lambda ratio, SIFE	203	Government	Retrospective	1995–2006	37 mo	nd	B (retrospective without adjustment)
Khoriaty, 2010[33] [20223721]	SFLC concentrations	89 (of which 43 with	nd	Retrospective	2004–2006	40 mo	nd	C (small sample

	IMWG criteria	evaluable disease)*						size, retrospective without adjustment)
Kroger, 2010[34] [2043663]	SFLC response	52	nd	Unclear	2003–2008	3 mo	nd	C (letter to the editor with limited information, small sample size, few details about SFLC response criteria and study design, limited data available)
	SIFE or UIFE							
van Rhee, 2007[38] [17416735]	SFLC response tertiles	303	Government	Retrospective	nd	21 mo	nd	B (retrospective with adjustment)
	Percent reduction of serum and urine M-protein concentrations							
Baseline SFLC concentrations and relationship to survival								
Kyrtsonis, 2007[36] [17408464]	SFLC ratio	94	Nonprofit foundation	Retrospective	nd	33 mo	nd	C (limited information about patient recruitment and study design, small sample size)
	ISS stages 1–3; Durie–Salmon stages I–III**							
van Rhee, 2007[38] [17416735]	Baseline SFLC concentrations	303	Government	Retrospective	nd	21 mo	nd	B (retrospective with adjustment)
	Baseline concentrations of serum and urine M-protein							

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, LCMM = light chain myeloma, MM = multiple myeloma, mo = months, nd = not described, NPV = negative predictive value, PCD = plasma-cell dyscrasia, SCT = stem cell transplantation, SFLC = serum free light chain, SIFE = serum immunofixation, SPEP = serum protein electrophoresis, UPEP = urine protein electrophoresis, yr = years

Table 6. Characteristics of patients in studies addressing KQ4

Author Year [PMID]	Enrollment method	Median age (yr)	Percent male	Population description and inclusion criteria
AL amyloidosis				
Kumar, 2011[35] [21328431]	nd	Cohort I: 58 Cohort II: 64	Cohort I: 57 Cohort II: 64	AL amyloidosis, with 347 patients receiving autologous SCT and 96 melphalan and dexamethasone
Lachmann, 2003[37] [12823348]	Referred patients	54–64	nd	Systemic AL amyloidosis not receiving prior chemotherapy, excluding those with concurrent MM or other malignant B-cell dyscrasia
Sanchorawala, 2005[16] [16044137]	nd	60	63	Receipt of high-dose intravenous melphalan and autologous SCT
MM				
SFLC response to therapy and relationship to outcomes				
Dispenzieri, 2008[31] [18364469]	Patients enrolled in a previous treatment trial (E9486)	63	65	Enrollment in a previously published treatment trial, diagnosis of MM, measurable disease in absence of treatment, pre- and post-treatment serum samples available
Giarin, 2009[32] [19520760]	nd	56	55	Newly diagnosed MM between July 1995 and February 2006, receipt of autologous or autologous + allogeneic SCT
Khoriaty, 2010[33] [20223721]	nd	61	65	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)
Kroger, 2010[34] [2043663]	nd	nd	nd	Diagnosed of MM, complete response between January 2003 and December 2008 for at least 3 mo, negative SIFE or UIFE test
Van Rhee, 2007[38] [17416735]	Patients enrolled in a previous trial (Total Therapy 3)	nd	64	Newly diagnosed MM, participation in a tandem autotransplantation trial
Baseline SFLC concentrations and relationship to survival				
Kyrtsonis, 2007[36] [17408464]	nd	32% >65 yr	45	MM, with or without treatment
Van Rhee, 2007[38] [17416735]	Patients enrolled in a previous trial (Total Therapy 3)	nd	64	Newly diagnosed MM, participation in a tandem autotransplantation trial

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, MM = multiple myeloma, SCT = stem cell transplantation, SFLC = serum free light chain, SIFE = serum immunofixation, UIFE = urine immunofixation

Table 7. Results of studies addressing KQ4

Author Year [PMID]	Sample size	Index test/ Comparator test	Results
AL amyloidosis			
Kumar, 2011[35]	443	Post-treatment dFLC	dFLC (vs. SPEP) significantly affected overall survival (p<0.0001)

Author Year [PMID]	Sample size	Index test/ Comparator test	Results
[21328431]	Cohort I: 347 Cohort II: 96	Post-treatment Quantitative M-protein concentrations	≤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 >90% decrease (p <0.001)
Lachmann, 2003[37] [12823348]	262	Post-treatment SFLC concentration Post-treatment quantitative paraprotein concentrations	86 patients with abnormal FLC concentration falling >50%after chemotherapy had 88% 5-year survival vs. only 39% among those with lesser reduction (p <0.0001) Amyloidogenic FLC reduction >50% associated with survival benefit, regardless of type of chemotherapy Amyloid load correlated with changes in SFLC concentration (p <0.0001). Among 73 patients with serially quantifiable serum paraprotein, survival better in those whose concentration fell by >50% vs. those in whom it fell by ≤50% (p <0.05).
Sanchorawala, 2005[16] [16044137]	66	Post-treatment SFLC concentration Hematological complete response (defined by EBMT criteria,[40] includes M-protein response)	Deaths: % (number/total number) Complete vs. noncomplete response: 4% (1/27) vs. 18% (7/39) FLC response >90% vs. ≤90%: 6% (2/35) vs. 19% (6/31) Clinical improvement: % (number/total number) Complete vs. noncomplete response: 96% (26/27) vs. 67% (26/39) FLC response >90% vs ≤90%: 97% (34/35) vs. 58% (18/31) FLC response and measures of hematological response complementary
MM			
SFLC response to therapy and relationship to outcomes			
Dispenzieri, 2008[31] [18364469]	139	SFLC response SPEP, UPEP	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<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 <0.001
Giarin,2009[32] [19520760]	203	SFLC ratio Total kappa/lambda ratio, SIFE	3 mo after SCT, overall and event-free survival did not differ significantly : between patients with and those without normal SFLC ratio between patients with and those without normal (negative) SIFE test 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)
Khoriaty, 2010[33] [20223721]	43 (those with evaluable disease)	SFLC ratio IMWG criteria[24]	For SFLC predicting 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%) For SFLC predicting progression (95% CI): Sensitivity: 93% (68 to 98%) Specificity: 80% (62 to 91%)

Author Year [PMID]	Sample size	Index test/ Comparator test	Results
			PPV: 72% (49 to 87%) NPV: 95% (78 to 99%)
Kroger, 2010[34] [2043663]	52	SFLC SIFE or UIFE	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
Van Rhee, 2007[38] [17416735]	303	SFLC response tertiles Percent reduction of serum and urine M-protein concentrations	Rate of near-complete response to induction therapy higher among patients with baseline SFLC >75 mg/dL than patients with baseline SFLC ≤75 mg/dL (37% vs. 20%, p = 0.002) Adjusted HR (95% CI) for overall survival: Top tertile (vs. lower two tertiles) in % SFLC reduction after cycle 2: 2.15 (1.03 to 4.47), p = 0.041 Top tertile (vs. lower two tertiles) in % SFLC reduction after transplantation: 2.24 (1.03 to 4.87), p = 0.042 Baseline SFLC >75 (vs. ≤75) mg/dL: 2.43 (1.18 to 5.01), p = 0.016 Adjusted HR (95% CI) for event-free survival: Top tertile(vs. lower two tertiles) in % SFLC reduction after cycle 2: 1.96 (1.03 to 3.74), p = 0.041 Top tertile (vs. lower two tertiles) in % SFLC reduction after transplantation: 2.01 (1.02 to 3.97), p = 0.045 Reductions in serum and urine M-protein values not significantly associated with overall or event-free survival
Baseline SFLC concentrations and relationship to survival			
Kyrtsonis, 2007[36] [17408464]	94	SFLC ratio ISS stages 1–3, Durie–Salmon stages I–III**	3- and 5-year disease-specific survival rates, 94% and 82%, respectively, with SFLC ratio below median (vs. 58% and 30%, respectively, SFLC concentrations above the median; p = 0.0001) Durie–Salmon and ISS staging were independent predictors of survival (p <0.0001 for both)
Van Rhee, 2007[38] [17416735]	303	Baseline SFLC concentrations Baseline concentrations of serum and urine M-protein	Rate of near-complete response to induction therapy higher among patients with baseline SFLC >75 mg/dL than patients with baseline SFLC ≤75 mg/dL (37% vs. 20%, p = 0.002). Adjusted HR (95% CI) for overall survival: Baseline SFLC >75 (vs. ≤75) mg/dL: 2.43 (1.18 to 5.01), p = 0.016 Adjusted HR (95% CI) for event-free survival: Baseline SFLC >75 (vs. ≤75) mg/dL: 2.40 (1.26 to 4.57), p = 0.008 Baseline concentrations of standard serum and urine M-protein did not identify prognostic subgroups

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, HR = hazard ratio, IFE = immunofixation, IMWG = International Myeloma Working Group, ISS = International Staging System, KQ = Key Question, LCMM = light chain myeloma, MM = multiple myeloma, mo = months, nd = not described, NPV = negative predictive value, PCD = plasma-cell dyscrasia, PPV = positive predictive value, SCT = stem cell transplantation, SFLC = serum free light chain, SIFE = serum immunofixation, SPEP = serum protein electrophoresis, UPEP = urine protein electrophoresis, yr = years

*Standard ECOG) response criteria are as follows: 50% decrease in serum M-protein or, in patients lacking a serum M-protein measurement, a 90% decrease in 24-hour urine M-protein.[31]

**ISS classification incorporates concentrations of serum albumin and β2 microglobulin.[41] The Durie–Salmon staging system classification incorporates concentrations of serum and urinary paraproteins.[42]

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 interventions (e.g., bone marrow biopsy)?

Results

We identified one retrospective cohort study, which we classified as being of quality C, assessing the need for bone marrow examination if the SFLC assay was used to define the completeness of response to treatment: Chee 2009[43] (**Tables 8–10**). As currently defined in the EBMT 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.[24,39]

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 interventions such as bone marrow examination. 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

Author Year [PMID]	Index test	Funding	Study design	Enrollment period	Followup duration	Diagnostic criteria	Quality grade and issues
Chee 2009,[43] [19641191]	SFLC ratio	Government	Retrospective cohort	nd	1995–??	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)	C (retrospective, small, convenience sample)

KQ = Key Question, MM = multiple myeloma, nd = not described, SFLC = serum free light chain, SIFE = serum immunofixation electrophoresis, UIFE = urine immunofixation electrophoresis

Table 9. Characteristics of patients in studies addressing KQ5

Author Year [PMID]	Enrollment method	Sample size	Median age (yr)	Sex	Exclusion criteria
Chee 2009,[43] [19641191]	Selected patients	92 With negative IFE, including 29 with normalized SFLC ratio Treatment: • Bone marrow transplantation, 51 • Chemotherapy, 26 • Second-line therapy, 10 • Unknown, 5	59 (median)	nd	Not specified

IFE = immunofixation electrophoresis, KQ = Key Question, nd = not described, SFLC = serum free light chain, yr = years

Table 10. Results of studies addressing KQ5

Author Year [PMID]	Index test	Comparator test and definition	Sample size	Results
Chee 2009,[43] [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, and relapse. In the present review, we assessed the role of the SFLC assay as an adjunct to traditional tests such as SPEP and SIFE for the diagnosis of PCDs in populations suspected of having the disease. We also ascertained the ability of the SFLC assay, 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 interventions.

Our results reveal that there continues to be a paucity of evidence to clarify the effectiveness and role of the SFLC assay for the diagnosis, management, and prognosis of PCDs. Thirteen studies, less than 1 percent of the 2819 studies identified in our literature search, met the inclusion criteria to address the KQs (see Appendix B for the excluded studies). Moreover, although in the 13 studies the SFLC assay and traditional testing were commonly conducted in parallel, they were not formally compared. Several articles evaluating the effectiveness and role of the SFLC assay were excluded because the populations were not of interest or there was a lack of data for traditional testing as a comparator. The study heterogeneity observed may also reflect the assay's uncertain role in research and clinical practice; across the studies, there was considerable clinical heterogeneity, with variation in study design and population as well as inconsistency in the comparisons being made. Moreover, the studies were of either poor or fair quality, not good quality. This, in combination with the study heterogeneity, limited the validity of the studies and the conclusions that could be drawn from them.

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 expand in detail the major needs of future studies.

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.[28-30] The studies were all retrospective, conducted in a hospital laboratory setting, and were of adults who were suspected to have a monoclonal gammopathy. They used archived laboratory samples that had been received for SPEP or UPEP. All three studies indicated that the addition of SFLC to traditional tests increased diagnostic sensitivity (by 6 to 36 percent), although the effect on diagnostic specificity was variable.

The limitations and potential biases in these studies limit their utility for informing clinical practice. We found that demographic details, including racial breakdown and comorbid conditions, were underreported. Overall 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.[14] Positive tests would be followed with more detailed testing, including IFE and bone marrow examination.[7,8] Therefore, ultimately the diagnosis 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 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. Given that PCDs are a heterogeneous group of conditions, the production of monoclonal light chains is 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. 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 diagnosis or test group to diminish the likelihood of misclassification bias. However, such safeguards were seldom emphasized in the studies reviewed.

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[15,44-62] or have examined the SFLC assay as an adjunct but only in populations already diagnosed with a monoclonal gammopathy or AL amyloidosis.

SFLC Assay and Treatment Response and Survival (KQ4)

Nine studies, three in patients with AL amyloidosis[16,35,37] and six in patients with MM,[31-34,36,38] evaluated SFLC testing compared with traditional tests 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, which were retrospective cohort studies, 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 studies, 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.

Six studies were reviewed in patients with MM[31-34,36,38]; five addressed the relationship between outcomes and SFLC response to treatment (induction therapy, chemotherapy, or stem-cell transplantation) (the sixth was descriptive). Outcomes reported were the achievement of response to treatment or event-free or overall survival. The traditional test comparators reported were quantitative changes in M-protein concentrations. Across the studies, an SFLC response

occurred earlier than an M-protein response or an overall response, and was also predictive of an overall response. Of two of these six studies that addressed SFLC or M-protein changes in relation to treatment, one did not show a relationship between normalization of the SFLC ratio and survival whereas the other showed that patients in the top tertile of SFLC reduction after treatment had a poorer survival. Two other studies that evaluated baseline SFLC concentrations as a predictor of survival showed that higher concentrations predicted poorer survival, independent of M-protein concentrations or disease stage but did not compare SFLC concentrations and traditional test results directly. Three studies were graded B, and three were graded C. 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 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. Nonetheless, there were several limitations in the literature reviewed. 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 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 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 Interventions (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.[63-65] 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 in defining the need for adjuvant therapy in patients with AL amyloidosis[66] 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.[67]

A single study, Chee 2009,[43] explored whether use the SFLC assay compared with traditional testing would reduce the need for other diagnostic interventions (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,[24,39] 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, however, this question requires more detailed and systematic evaluation in practice.

Future Research

It appears that there is considerable clinical uncertainty regarding the applications of the SFLC assay both within and beyond the 2009 IMWG consensus guidelines.[14] 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 in 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.

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, would yield 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.

Several limitations of the studies reviewed make it difficult to present clear conclusions regarding the comparative effectiveness of the SFLC assay and should be targeted in future studies. Diagnostic studies using only data from patients already diagnosed with PCDs were excluded from this review (see Appendix B). The studies of only patients with disease reflected the extreme end of the spectrum of disease severity, and the proportion of patients with a positive test was likely overestimated. Moreover, without a nondiseased population, true negatives could not be assessed. Studies could not avoid incorporation bias because the result from the reference test itself (e.g., SPEP or SIFE) was considered along with other factors, such as clinical information, to reach a diagnosis of PCD. 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. Most studies did not report whether data assessors were blinded to diagnosis or 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.

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 variability in SFLC concentrations, and the harms of testing, questions that were outside the scope of this review. In addition, the lack of a suitable gold 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.

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 for 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. Although it is well appreciated that MM is more common among African Americans than among whites, this source of variability was not routinely considered in

the studies we reviewed. Future studies of SFLC 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

AAAC	American Association for Clinical Chemistry
AHRQ	Agency for Healthcare Research and Quality
AL amyloidosis	Systemic, or primary, amyloidosis in which amyloid [A] proteins derived from immunoglobulin light chains [L] are deposited in tissue (also called light-chain amyloidosis)
CER	Comparative Effectiveness Review
CI	confidence interval
EPC	Evidence-based Practice Center
FDA	Food and Drug Administration
FLC	Free light chain
IFE	Immunofixation electrophoresis
IMWG	International Myeloma Working Group
LCMM	Light chain multiple myeloma
MGUS	Monoclonal gammopathy of undetermined significance
MM	Multiple myeloma
M-protein	Monoclonal protein (also called paraprotein)
NSMM	Nonsecretory multiple myeloma
PCD	Plasma-cell dyscrasia
KQ	Key Question
PICOTS	Populations, interventions, comparators, outcomes, timing, and settings
RCT	randomized controlled trial
SFLC	Serum free light chain
SIFE	Serum immunofixation electrophoresis
SMM	Smoldering multiple myeloma
SPEP	Serum protein electrophoresis
TEP	Technical Expert Panel
TOO	Task Order Officer
UIFE	Urine immunofixation electrophoresis
UPEP	urine protein electrophoresis

Glossary

Term	Definition	Source, if applicable
Cast nephropathy (or "myeloma kidney")	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	
Differential verification	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.	Reitsma 2009[68]
Disease progression or recovery bias	Bias from an inappropriately long interval (or any interval) between conduct of reference test and conduct of index test.	Reitsma 2009[68]
Incorporation bias	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).	Reitsma 2009[68]
Involved FLC or SFLC	The free light chain or serum free light chain that is produced in excess and is causing disease (either kappa or lambda)	
Light chain escape	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.	Dispenzieri 2009[14]
Measurable disease	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	
M-protein or paraprotein	Intact immunoglobulins or FLCs of a single type produced in excess by an abnormally expanded clone of malignant plasma cells (biomarkers of PCDs)	
Monoclonal gammopathy	Disease class comprising PCDs as well as other conditions such as hematological disorders associated with a monoclonal band	
Oligosecretory MM	MM in which very small amounts of M-protein are produced by the malignant plasma cells	
Polyclonal gammopathy	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	
Selection bias (also called partial verification bias, workup bias, or sequential ordering bias)	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.	Reitsma 2009[68]
SFLC ratio	The ratio of kappa chains to lambda chains, for which the normal range is 0.26–1.65.	Katzmann 2006[13]
Spectrum bias or effect	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.)	Reitsma 2009[68]

Verification bias	Bias in appraising performance of a diagnostic index test introduced through the inaccuracy of the reference standard.	Reitsma 2009[68]

References

1. Kyle RA, Rajkumar SV. Epidemiology of the plasma-cell disorders. Best Pract Res Clin Haematol. 2007;20:637-64

2. Ries LAG, Harkins D, Krapcho M, et al. SEER cancer statistics review, 1975-2003. NCI, Bethesda, MD; 2005. Available at: http://seer.cancer.gov/csr/1975_2003.

3. Jemal A, Siegel R, Ward E, et al. Cancer Statistics, 2007. CA: A Cancer Journal for Clinicians 2007;57(1):43-66. PMID: 10.3322/canjclin.57.1.43.

4. Katzel JA, Hari P, Vesole DH. Multiple Myeloma: Charging Toward a Bright Future. CA: A Cancer Journal for Clinicians 2007;57(5):301-18. PMID: 10.3322/CA.57.5.301.

5. Cook R. An economic perspective on treatment options in multiple myeloma. Managed Care Oncol 2007;2007(Spring):10-12

6. Messori A, Trippoli S, Santarlasci B. Pharmacotherapy of multiple myeloma: an economic perspective. Expert Opin Pharmacother 2003 Apr 1;4(4):515-24. PMID: doi: 10.1517/14656566.4.4.515.

7. Kyle RA, Rajkumar SV. Criteria for diagnosis, staging, risk stratification and response assessment of multiple myeloma. Leukemia 2009;23(1):3-9

8. Palumbo A, Sezer O, Kyle R, et al. International Myeloma Working Group guidelines for the management of multiple myeloma patients ineligible for standard high-dose chemotherapy with autologous stem cell transplantation. Leukemia 2009 Jun 4;23(10):1716-30

9. Rajkumar SV. Clinical features, laboratory manifestations, and diagnosis of multiple myeloma. In: Basow DS (eds). UpToDate. Waltham: UpToDate; 2011.

10. Bradwell AR, Carr-Smith HD, Mead GP, et al. Highly sensitive, automated immunoassay for immunoglobulin free light chains in serum and urine. Clin Chem 2001 Apr;47(4):673-80

11. Bradwell AR, Carr-Smith HD, Mead GP, et al. Highly Sensitive, Automated Immunoassay for Immunoglobulin Free Light Chains in Serum and Urine. Clin Chem 2001 Apr 1;47(4):673-80

12. Rajkumar SV. Recognition of monoclonal proteins. In: Kyle RA, Connor RF (eds). UpToDate. Waltham: UpToDate; 2011.

13. Katzmann JA. Serum free light chains: quantitation and clinical utility in assessing monoclonal gammopathies. Clin Lab News 2006 Jun;June:12-14

14. Dispenzieri A, Kyle R, Merlini G, et al. International Myeloma Working Group guidelines for serum-free light chain analysis in multiple myeloma and related disorders. [Review] [43 refs]. Leukemia 2009 Feb;23(2):215-24

15. Mead GP, Carr-Smith HD, Drayson MT, et al. Serum free light chains for monitoring multiple myeloma. British Journal of Haematology 2004 Aug;126(3):348-54

16. Sanchorawala V, Seldin DC, Magnani B, et al. Serum free light-chain responses after high-dose intravenous melphalan and autologous stem cell transplantation for AL (primary) amyloidosis. Bone Marrow Transplantation 2005 Oct;36(7):597-600

17. Kyle RA, Durie BGM, Rajkumar SV, et al. Monoclonal gammopathy of undetermined significance (MGUS) and smoldering (asymptomatic) multiple myeloma: IMWG consensus perspectives risk factors for progression and guidelines for monitoring and management. Leukemia 2010 Jun;24(6):1121-27

18. Bradwell AR. Serum free light chain measurements move to center stage. Clin Chem 2005 May;51(5):805-07

19. National Comprehensive Cancer Network. NCCN clinical practice guidelines in oncology: multiple myeloma. 2011. Available at: http://www.nccn.org/professionals/physician_gls/f_guidelines.asp#site. Accessed on February 21, 2011

20. Whitlock EP, Lopez SA, Chang S, et al. AHRQ Series Paper 3: Identifying, selecting, and refining topics for comparative effectiveness systematic reviews: AHRQ and the Effective Health-Care program. J Clin Epidemiol. 2010;63:491-501

21. Agency for Healthcare Research and Quality. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. 2011

22. AHRQ. Methods Guide for Medical Test Reviews. Agency for Healthcare Research and Quality, Rockville, MD; 2010. Available at: http://www.effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?productid=558&pageaction=displayproduct.

23. Wallace BC, Trikalinos TA, Lau J, et al. Semi-automated screening of biomedical citations for systematic reviews. BMC Bioinformatics 2010;11:55

24. Durie BG, Harousseau JL, Miguel JS, et al. International uniform response criteria for multiple myeloma.[Erratum appears in Leukemia. 2007 May;21(5):1134].[Erratum appears in Leukemia. 2006 Dec;20(12):2220]. Leukemia 2006 Sep;20(9):1467-73

25. Atkins D, Eccles M, Flottorp S, et al. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches The GRADE Working Group. BMC Health Services Research 4(1):38, 2004 Dec 22

26. West S, King V, Carey TS, et al. Systems to Rate the Strength of Scientific Evidence. Evidence Report/Technology Assessment No. 47 (Prepared by the Research Triangle Institute-University of North Carolina Evidence-based Practice Center under Contract No. 290-97-0011). 2002

27. 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

28. Abadie JM, Bankson DD. Assessment of serum free light chain assays for plasma cell disorder screening in a Veterans Affairs population. Annals of Clinical & Laboratory Science 2006;36(2):157-62

29. Piehler AP, Gulbrandsen N, Kierulf P, et al. Quantitation of serum free light chains in combination with protein electrophoresis and clinical information for diagnosing multiple myeloma in a general hospital population. Clin Chem 2008 Nov;54(11):1823-30

30. Vermeersch P, Van HL, Delforge M, et al. Diagnostic performance of serum free light chain measurement in patients suspected of a monoclonal B-cell disorder. British Journal of Haematology 2008 Nov;143(4):496-502

31. Dispenzieri A, Zhang L, Katzmann JA, et al. Appraisal of immunoglobulin free light chain as a marker of response. Blood 2008 May 15;111(10):4908-15

32. Giarin MM, Giaccone L, Sorasio R, et al. Serum free light chain ratio, total kappa/lambda ratio, and immunofixation results are not prognostic factors after stem cell transplantation for newly diagnosed multiple myeloma. Clin Chem 2009 Aug;55(8):1510-16

33. Khoriaty R, Hussein MA, Faiman B, et al. Prediction of response and progression in multiple myeloma with serum free light chains assay: corroboration of the serum free light chain response definitions. Clinical lymphoma, myeloma & leukemia 2010 Feb;10(1):E10-E13

34. Kroger N, Asenova S, Gerritzen A, et al. Questionable role of free light chain assay ratio to determine stringent complete remission in multiple myeloma patients. Blood ;115(16):3413-14

35. Kumar SK, Dispenzieri A, Lacy MQ, et al. Changes in serum-free light chain rather than intact monoclonal immunoglobulin levels predicts outcome following therapy in primary amyloidosis. American Journal of Hematology 2011 Mar;86(3):251-55

36. Kyrtsonis MC, Vassilakopoulos TP, Kafasi N, et al. Prognostic value of serum free light chain ratio at diagnosis in multiple myeloma. British Journal of Haematology 2007 May;137(3):240-43

37. Lachmann HJ, Gallimore R, Gillmore JD, et al. Outcome in systemic AL amyloidosis in relation to changes in concentration of circulating free immunoglobulin light chains following chemotherapy. British Journal of Haematology 2003 Jul;122(1):78-84

38. van RF, Bolejack V, Hollmig K, et al. High serum-free light chain levels and their rapid reduction in response to therapy define an aggressive multiple myeloma subtype with poor prognosis. Blood 2007 Aug 1;110(3):827-32

39. Gertz MA, Comenzo R, Falk RH, et al. Definition of organ involvement and treatment response in immunoglobulin light chain amyloidosis (AL): a consensus opinion from the 10th International Symposium on Amyloid and Amyloidosis, Tours, France, 18-22 April 2004. [Review] [100 refs]. American Journal of Hematology 2005 Aug;79(4):319-28

40. Bladë J, Samson D, Reece D, et al. Criteria for evaluating disease response and progression in patients with multiple myeloma treated by high-dose therapy and haemopoietic stem cell transplantation. British Journal of Haematology 1998;102(5):1115-23. PMID: 10.1046/j.1365-2141.1998.00930.x.

41. Greipp PR, Miguel JS, Durie BGM, et al. International Staging System for Multiple Myeloma. Journal of Clinical Oncology 2005 May 20;23(15):3412-20

42. Durie BG, Salmon SE. A clinical staging system for multiple myeloma. Correlation of measured myeloma cell mass with presenting clinical features, response to treatment, and survival. Cancer. 1975;36:842-54

43. Chee CE, Kumar S, Larson DR, et al. The importance of bone marrow examination in determining complete response to therapy in patients with multiple myeloma. Blood 2009 Sep 24;114(13):2617-18

44. Abadie JM, van Hoeven KH, Wells JM. Are renal reference intervals required when screening for plasma cell disorders with serum free light chains and serum protein electrophoresis? American Journal of Clinical Pathology 2009 Feb;131(2):166-71

45. Abraham RS, Katzmann JA, Clark RJ, et al. Quantitative analysis of serum free light chains. A new marker for the diagnostic evaluation of primary systemic amyloidosis. American Journal of Clinical Pathology 2003 Feb;119(2):274-78

46. Akar H, Seldin DC, Magnani B, et al. Quantitative serum free light chain assay in the diagnostic evaluation of AL amyloidosis. Amyloid 2005 Dec;12(4):210-15

47. Beetham R, Wassell J, Wallage MJ, et al. Can serum free light chains replace urine electrophoresis in the detection of monoclonal gammopathies? Annals of Clinical Biochemistry 2007 Nov;44(Pt:6):6-22

48. Bochtler T, Hegenbart U, Heiss C, et al. Evaluation of the serum-free light chain test in untreated patients with AL amyloidosis. Haematologica 2008 Mar;93(3):459-62

49. Boer K, Deufel T. Quantitation of serum free light chains does not compensate for serum immunofixation only when screening for monoclonal gammopathies. Clinical Chemistry & Laboratory Medicine 2009;47(9):1109-15

50. Fulton RB, Fernando SL. Serum free light chain assay reduces the need for serum and urine immunofixation electrophoresis in the evaluation of monoclonal gammopathy. Annals of Clinical Biochemistry 2009 Sep;46(Pt:5):5-12

51. Hofmann W, Garbrecht M, Bradwell AR, et al. A new concept for detection of Bence Jones proteinuria in patients with monoclonal gammopathy. Clinical Laboratory 2004;50(3-4):181-85

52. Holding S, Spradbery D, Hoole R, et al. Use of serum free light chain analysis and urine protein electrophoresis for detection of monoclonal gammopathies. Clinical Chemistry & Laboratory Medicine 2011 Jan;49(1):83-88

53. Hutchison CA, Plant T, Drayson M, et al. Serum free light chain measurement aids the diagnosis of myeloma in patients with severe renal failure. BMC Nephrology 2008;9:11

54. Jaskowski TD, Litwin CM, Hill HR. Detection of kappa and lambda light chain monoclonal proteins in human serum: automated immunoassay versus immunofixation electrophoresis. Clinical & Vaccine Immunology: CVI 2006 Feb;13(2):277-80

55. Katzmann JA, Dispenzieri A, Kyle RA, et al. Elimination of the need for urine studies in the screening algorithm for monoclonal gammopathies by using serum immunofixation and free light chain assays. Mayo Clinic Proceedings 2006 Dec;81(12):1575-78

56. Katzmann JA, Kyle RA, Benson J, et al. Screening panels for detection of monoclonal gammopathies. Clin Chem 2009 Aug;55(8):1517-22

57. Marien G, Oris E, Bradwell AR, et al. Detection of monoclonal proteins in sera by capillary zone electrophoresis and free light chain measurements. Clin Chem 2002 Sep;48(9):1600-01

58. Nowrousian MR, Brandhorst D, Sammet C, et al. Serum free light chain analysis and urine immunofixation electrophoresis in patients with multiple myeloma. Clinical Cancer Research 2005 Dec 15;11(24:Pt 1):t-14

59. Palladini G, Russo P, Bosoni T, et al. Identification of amyloidogenic light chains requires the combination of serum-free light chain assay with immunofixation of serum and urine. Clin Chem 2009 Mar;55(3):499-504

60. Ramasamy I. Serum free light chain analysis in B-cell dyscrasias. Annals of Clinical & Laboratory Science 2007;37(3):291-94

61. Sthaneshwar P, Nadarajan V, Maniam JA, et al. Serum free light chains: diagnostic and prognostic value in multiple myeloma. Clinical Chemistry & Laboratory Medicine 2009;47(9):1101-07

62. Wood PB, McElroy YG, Stone MJ. Comparison of serum immunofixation electrophoresis and free light chain assays in the detection of monoclonal gammopathies. Clinical lymphoma, myeloma & leukemia 2010 Aug 1;10(4):278-80

63. Dispenzieri A, Katzmann JA, Kyle RA, et al. Prevalence and risk of progression of light-chain monoclonal gammopathy of undetermined significance: a retrospective population-based cohort study.[Erratum appears in Lancet. 2010 Jul 31;376(9738):332]. Lancet 2010 May 15;375(9727):1721-28

64. Rajkumar SV, Kyle RA, Therneau TM, et al. Serum free light chain ratio is an independent risk factor for progression in monoclonal gammopathy of undetermined significance. Blood 2005 Aug 1;106(3):812-17

65. Rajkumar SV, Kyle RA, Therneau TM, et al. Presence of monoclonal free light chains in the serum predicts risk of progression in monoclonal gammopathy of undetermined significance. British Journal of Haematology 2004 Nov;127(3):308-10

66. Cohen AD, Zhou P, Chou J, et al. Risk-adapted autologous stem cell transplantation with adjuvant dexamethasone +/- thalidomide for systemic light-chain amyloidosis: results of a phase II trial. British Journal of Haematology 2007 Oct;139(2):224-33

67. Hutchison CA, Bradwell AR, Cook M, et al. Treatment of acute renal failure secondary to multiple myeloma with chemotherapy and extended high cut-off hemodialysis. Clinical Journal of The American Society of Nephrology: CJASN 2009 Apr;4(4):745-54

68. Reitsma JB, Rutjes AWS, Whiting P, et al. Chapter 9: Assessing methodological quality. In: Deeks JJ, Bossuyt PM, Gastonis C (eds). Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy (v.1.0.0). The Cochrane Collaboration; 2009. p. 1-24