



Published in final edited form as:

Clin Chem. 2011 December ; 57(12): 1687–1692. doi:10.1373/clinchem.2011.171314.

Long-Term Biological Variation of Serum Protein Electrophoresis M-Spike, Urine M-Spike, and Monoclonal Serum Free Light Chain Quantification: Implications for Monitoring Monoclonal Gammopathies

Jerry A. Katzmann^{1,2,*}, Melissa R. Snyder¹, S. Vincent Rajkumar^{1,2}, Robert A. Kyle^{1,2}, Terry M. Therneau³, Joanne T. Benson³, and Angela Dispenzieri^{1,2}

¹Division of Clinical Biochemistry & Immunology, Department of Laboratory Medicine & Pathology; Mayo Clinic, Rochester, MN

²Division of Hematology, Department of Internal Medicine; Mayo Clinic, Rochester, MN

³Department of Biomedical Statistics & Informatics, Mayo Clinic, Rochester, MN

Abstract

BACKGROUND—We analyzed serial data in patients with clinically stable monoclonal gammopathy to determine the total variation of serum M-spikes [measured with serum protein electrophoresis (SPEP)], urine M-spikes [measured with urine protein electrophoresis (UPEP)], and monoclonal serum free light chain (FLC) concentrations measured with immunoassay.

METHODS—Patients to be studied were identified by (a) no treatment during the study interval, (b) no change in diagnosis and <5 g/L change in serum M-spike over the course of observation; (c) performance of all 3 tests (SPEP, UPEP, FLC immunoassay) in at least 3 serial samples that were obtained 9 months to 5 years apart; (d) serum M-spike 10 g/L, urine M-spike 200 mg/24 h, or clonal FLC 100 mg/L. The total CV was calculated for each method.

RESULTS—Among the cohort of 158 patients, 90 had measurable serum M-spikes, 25 had urine M-spikes, and 52 had measurable serum FLC abnormalities. The CVs were calculated for serial SPEP M-spikes (8.1%), UPEP M-spikes (35.8%), and serum FLC concentrations (28.4%). Combining these CVs and the interassay analytical CVs, we calculated the biological CV for the serum M-spike (7.8%), urine M-spike (35.5%), and serum FLC concentration (27.8%).

© 2011 American Association for Clinical Chemistry

*Address correspondence to this author at: Mayo Clinic, 200 First St. SW, Rochester, MN 55905. Fax 507-266-4088; katzmann.jerry@mayo.edu.

Author Contributions: All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.

Authors' Disclosures or Potential Conflicts of Interest: Upon manuscript submission, all authors completed the Disclosures of Potential Conflict of Interest form. Potential conflicts of interest:

Employment or Leadership: None declared.

Consultant or Advisory Role: None declared.

Stock Ownership: None declared.

Honoraria: J.A. Katzmann, Binding Site; R.A. Kyle, Binding Site.

Research Funding: None declared.

Expert Testimony: None declared.

Other Remuneration: A. Dispenzieri, Binding Site paid for travel to a meeting.

CONCLUSIONS—The variations in urine M-spike and serum FLC measurements during patient monitoring are similar and are larger than those for serum M-spikes. In addition, in this group of stable patients, a measurable serum FLC concentration was available twice as often as a measurable urine M-spike.

Plasma cell proliferative disorders are commonly associated with the synthesis and secretion of a monoclonal immunoglobulin. These abnormal proteins may be monitored by a variety of methods. Serum protein electrophoresis (SPEP)⁴ and/or urine protein electrophoresis (UPEP) M-spike quantifications can distinguish polyclonal and monoclonal immunoglobulins and are used in multiple myeloma (MM) to detect response to treatment or relapse. In addition, these measures are commonly assessed in patients with monoclonal gammopathy of undetermined significance (MGUS) and smoldering MM (SMM) as indicators of progression. Guidelines for MM disease monitoring recommend the use of the serum M-spike if it is ≥ 10 g/L (i.e., measurable) and urine M-spike if ≥ 200 mg/24 h (1). In serum, reductions in the M-spike of at least 25% and 50% are considered minimal and partial responses, respectively (1, 2). The urine M-spike, however, requires at least a 50% and 90% decrease for minimal and partial responses (1, 2). A complete response requires the absence of a monoclonal protein detected by immunofixation electrophoreses.

Serum immunoglobulin concentrations, often assessed as a quality check for changes in the SPEP M-spike, may be especially useful when the M-spike is >20 – 30 g/L (3) or if the electrophoretic migration of a monoclonal IgA, or rarely IgM, is obscured within the β fraction. In addition, in the absence of a measurable serum or urine M-spike, the International Myeloma Working Group (IMWG) has recommended that serum free light chain (FLC) concentration be used to monitor disease if the monoclonal (involved) FLC (iFLC) concentration is ≥ 100 mg/L in the presence of an abnormal FLC κ/λ ratio (rFLC). Analogous to the SPEP M-spike, a 50% decrease has been suggested as a partial response criterion (1, 4).

There is a large body of work regarding within-person variation and the meaning of differences between sequential laboratory results (5, 6). These studies have recognized the variation to be the sum of preanalytic and analytical variability, as well as intraindividual biological variability. Traditionally, these studies have focused on healthy individuals with results within the working range of an assay. Serum immunoglobulins can be quantified in patients with monoclonal immunoglobulins and variations can be compared with reference intervals for serum immunoglobulins, but there are no normal counterparts to serum and urine M-spikes. We have analyzed serial samples in clinically stable patients to assess the total variability (analytical plus biological) of these monitoring tests. Intrinsic to this approach is that the biological variability also may contain disease variability despite our restricting the patient cohort to clinically defined stable patients. We have undertaken these studies to evaluate disease-monitoring recommendations, with particular emphasis on the recommendations for serum FLC.

Methods

To identify patients with clinically stable disease, we queried our database for patients who met the following criteria: (a) no treatment during the study interval, (b) no change in clinical diagnosis, (c) <5 g/L change in serum M-spike over the course of the observation compared to the first sample in the series, (d) at least 3 serial samples obtained within 5

⁴Nonstandard abbreviations: SPEP, serum protein electrophoresis; UPEP, urine protein electrophoresis; MM, multiple myeloma; MGUS, monoclonal gammopathy of undetermined significance; SMM, smoldering MM; IMWG, International Myeloma Working Group; FLC, free light chain; iFLC, involved (monoclonal) FLC; rFLC, ratio of κ to λ FLC; uFLC, uninvolved FLC.

years, and (e) availability of all test results in the clinical history. For a patient's results to be analyzed for variation, at least 1 method had to fulfill the definition of a measurable response criterion (3): (1) SPEP M-spike ≥ 10 g/L, (2) UPEP M-spike ≥ 200 mg/24 h, or (3) iFLC ≥ 100 mg/L in the presence of an rFLC that was not within reference intervals. All data were retrospective and obtained from the medical record, and all queries to the database were done under a protocol approved by the Mayo Clinic Institutional Review Board.

We measured immunoglobulins by immunonephelometry using a Siemens BNII and Siemens reagent sets. We measured M-spikes using Helena SPIFE SPE reagent sets (Helena Labs) in combination with serum total protein determined with Biuret reagent (Siemens) on an Advia 1200 (Siemens) and urine total protein determined with pyrogallol red (Wako) on a Cobas c501 analyzer (Roche Diagnostics). We measured κ and λ FLC using a Siemens BNII nephelometer and Freelite reagent sets from The Binding Site. The FLC data were analyzed as the iFLC concentration, the uninvolved FLC (uFLC) concentration, and the rFLC.

Visual inspection of the data revealed that a few individuals displayed a slight trend over time. The variation within an individual was therefore defined as the SD about the fitted line for each person. We computed the total CV for each test as the mean square error using ANOVA on the logarithm of the laboratory value. A single fit was done including all individuals, with separate slope and intercept parameters per person. Regression on the log(response) is appropriate for data with constant CV, and the resulting estimate of variance from the model is a direct estimate of the CV of the data on the original scale (7).

For the analytical CVs of the methods, we used the clinical laboratory interassay CVs for patient-derived control samples. All samples were collected in a single institution, and we have assumed minimal preanalytic variance.

Results

We selected 158 patients for evaluation: there were a minimum of 3 observations per patient and a maximum of 16 observations (median 4) (Table 1). All patients had monoclonal proteins of the IgG isotype and had a diagnosis of MGUS, SMM, or MM. Although not all assays had results that were measurable according to international consensus criteria, all patients had a serum M-spike and serum FLC quantification. The mean (range) for serum M-spike concentration was 14.3 g/L (1–51 g/L). The rFLC ranged from 0.001 to 1860. Only 25 patients had a urine M-spike; 148 had quantification of serum IgG.

We calculated the CVs from the patient sample sets and graphed the distribution of the CVs in relation to the mean concentration of the sample set (Fig. 1). Current recommendations for “measurable” values are indicated by the dashed lines.

The total CVs for serial measurements of serum M-spike, IgG quantification, urine M-spike, iFLC, uFLC, and rFLC are tabulated by diagnosis in Table 2. There were 90 patients whose serum M-spike was ≥ 10 g/L; the mean CVs for serum M-spike and IgG quantification were 8.1% and 13.0%, respectively. The larger CV for the IgG quantification is presumably due to the inclusion of monoclonal and polyclonal IgG in the nephelometric quantification. Although SMM and MGUS are both premalignant disorders, SMM patients are monitored at more frequent intervals than those with MGUS—the SMM samples were collected within 15 months, whereas the MGUS samples were collected within 5 years.

There were 25 patients whose urine M-spike was ≥ 200 mg/24 h; only 1 of these had MGUS. The urine M-spike CV was 35.8% (Table 2). There were 52 patients whose serum iFLC was

100 mg/L; the serum iFLC CV was 28.4%. These 2 variabilities were 3–4 times larger than the serum M-spike, which quantifies intact monoclonal immunoglobulin in these patients.

The CVs for all 158 iFLC and 157 uFLC measurements were 34.9% and 45.2%, respectively. The larger CV for the uFLC is presumably due to the lower concentrations of uFLC. The CV of the rFLC (47.5%) is a combination of the iFLC and uFLC and is even larger.

The percentage decreases needed in each assay to achieve 50%, 80%, 90%, or 95% probability of statistical significance between consecutive results in an individual are listed in Table 3. Each of these values indicates the decrease needed at each of the different probabilities to minimize the chances of falsely reporting a positive response in a stable patient.

The total variation inherent in these patient sample sets is a composite of the interassay analytical variation and the long-term biological variation. The analytical variations of the assays are known from the interassay CV of our clinical laboratory assay validation documentation. The analytical CVs for the SPEP M-spike, UPEP M-spike, and iFLC are 2.1%, 4.5%, and 5.8%, respectively (Table 4). The within-person biological variation is used here as an aggregate variable including preanalytical variation, disease variation, and within-person biological variation. Using the total average CVs and the interassay analytical CVs, the calculated biological variations are 7.8% (serum M-spike), 35.5% (urine M-spike), and 27.8% (iFLC).

Discussion

The guidelines for monitoring monoclonal gammopathies propose that a 25% decline in serum M-spike defines a minimal response and a 50% decline is considered a partial response; in urine, the corresponding decreases needed are 50% and 90% (1, 2). It has recently been recommended that a partial response in serum FLC concentration be defined like that of the serum M-spike. These guidelines have been proposed to standardize treatment-response criteria. Our data confirm some of these guidelines but suggest that the FLC guidelines should be more stringent and similar to the recommendations for urine M-spike.

During monitoring of stable disease, the serum electrophoretic M-spike has a CV of 8.1%. This variation reflects the interassay analytical CV (2.1%) and the long-term biological CV (7.8%). Although the patient selection criteria for this study required that the M-spike vary <5 g/L, the total and biological CVs are surprisingly low, considering concerns about variations in hydration and hematocrit over these long time frames. The serum M-spike CV is smaller than the IgG quantification CV (8.1% vs 13.0%), presumably owing to the added variability of polyclonal immunoglobulins contained in the IgG quantification. Either measurement, however, should suffice for monitoring patients with large serum M-spikes and suppressed polyclonal immunoglobulins. The 12.3% biological variation of the monoclonal IgG quantification (Table 4) compares to 4.5% for the within-person biological variation of normal IgG (8). This increase in withinperson biological variation for monoclonal IgG in our study is presumably due to the 2 additional factors of longer study intervals and the disease variation intrinsic to studying a cohort of diseased patients. Although IgG can be quantified in normal sera, serum and urine M-spikes as well as abnormal rFLC and increased iFLC are found only in the presence of disease, and thus, these quantities could be characterized only in diseased patients.

The urine M-spike has a CV of 35.8%. This variation is approximately 4 times larger than the serum M-spike CV and is due almost entirely to a large biological variation (35.3%) in

the 24-h urine M-spike. It should again be emphasized that we used the serum M-spike as part of the criteria for defining stable disease in this patient cohort, and this selection criterion may have biased the relationship of the assay CVs. When we attempted to select patients using urine M-spike and serum FLC criteria, we obtained smaller numbers of patients. The relationships between the 3 sets of CVs, however, remained essentially the same.

The serum iFLC has a CV of 28.4%. The iFLC has been the FLC format recommended for monitoring oligosecretory multiple myeloma (4), whereas the difference between the iFLC and uFLC has been recommended for patients with primary amyloid (9) to normalize the effects of declining renal clearance. Reduced renal clearance results in increased serum FLC concentrations, as indicated by reference interval studies and studies in renal failure patients (10, 11). The rFLC normalizes most of the changes seen with reduced renal function and is recommended for diagnostic studies. The rFLC had originally been thought to normalize biological variation, but it is the most variable of the FLC data formats and is not recommended for disease monitoring.

In comparison to the other serum measurements of M-spike and IgG, the serum iFLC variation is larger than expected, and like that for urine, has almost entirely a biological basis. Because of its low molecular weight and short serum half-life, the FLC concentration responds very quickly to changes in plasma cell synthesis and renal clearance. The short half-life may be the reason for the large CV of FLC measurements compared with intact immunoglobulin measurements. Small variations in synthesis or clearance of intact immunoglobulin get diluted in the large pool of long-lived serum immunoglobulins, whereas these variations for light chains are not smoothed by dilution in a large, long-lived serum pool of FLC. Although the large CV of urine M-spike assays usually has been attributed to the preanalytic variability of 24-h urine collections, this same FLC physiology is probably responsible for most of the large CV of the urine M-spike. The increased variability of the urine M-spike compared with the serum iFLC concentration may reflect the preanalytic variability inherent in a 24-h urine collection.

The percent reductions between samples listed in Table 3 are those needed for the different probabilities listed in the table, and these different reductions reflect the tradeoff between the need for sensitivity of the test for early detection of response and a low false-positive rate. As the percent reduction gets larger and approaches the 95% probability threshold, the rate of false positives in stable patients falls, but the rate of false negatives in unstable patients may increase. The reductions in serum and urine M-spikes needed for the 95% probability threshold in Table 3 are close to the current guidelines requiring a 25% or 50% decrease for a minimal response in serum or urine M-spike, respectively.

The FLC results in Table 3 indicate that the guidelines for serum FLC need to be more stringent and more closely resemble the UPEP guidelines rather than the SPEP guidelines. Several studies have correlated clinical response to changes in serum FLC concentration. Studies in primary amyloid (12) and multiple myeloma (13) indicated that a 50% reduction in FLC concentration correlated with response, and responses up to 90% improved prognostic predictions (14-18). Although some of these studies indicated that a 50% reduction in FLC was significant for predicting response, the biological variation data indicate that a 50% change has only an 80%–90% probability of being a significant change. These data suggest that the criteria for minimal and partial response using FLC should be comparable to those of the UPEP M-spike. Finally, these data show that serum FLC, when used as a monitoring tool, is measurable in more patients than the urine M-spike.

Acknowledgments

Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, or preparation or approval of manuscript.

References

1. Durie BG, Harousseau JL, Miguel JS, Blade J, Barlogie B, Anderson K, et al. International uniform response criteria for multiple myeloma. *Leukemia*. 2006; 20:1467–73. [PubMed: 16855634]
2. Smith A, Wisloff F, Samson D. Guidelines on the diagnosis and management of multiple myeloma 2005. *Br J Haematol*. 2006; 132:410–51. [PubMed: 16412016]
3. Murray DL, Ryu E, Snyder MR, Katzmann JA. Quantitation of serum monoclonal proteins: relationship between agarose gel electrophoresis and immunonephelometry. *Clin Chem*. 2009; 55:1523–9. [PubMed: 19520759]
4. Dispenzieri A, Kyle R, Merlini G, Miguel JS, Ludwig H, Hajek R, et al. International Myeloma Working Group guidelines for serum-free light chain analysis in multiple myeloma and related disorders. *Leukemia*. 2009; 23:215–24. [PubMed: 19020545]
5. Fraser, CG. AACC Press; Washington (DC): 2001. Biological variation: from principles to practice.
6. Smellie WSA. What is a significant difference between sequential laboratory results? *J Clin Pathol*. 2008; 61:419–25. [PubMed: 17938161]
7. Beal SL, Sheiner LB. Heteroscedastic nonlinear regression. *Technometrics*. 1988; 30:327–37.
8. Ricós, C.; García-Lario, J-V.; Alvares, V.; Cava, F.; Domenech, M.; Hernández, A., et al. [Accessed September 2011] Biological variation database, and quality specifications for imprecision, bias and total error (desirable and minimum): the 2010 update. <http://www.westgard.com/biodatabase-2010-update.htm>
9. Gertz MA, Comenzo R, Falk RH, Fermand JP, Hazenberg BP, Hawkins PN, 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. *Am J Hematol*. 2005; 79:319–28. [PubMed: 16044444]
10. Katzmann JA, Clark RJ, Abraham RS, Bryant S, Lymp JF, Bradwell AR, et al. Serum reference intervals and diagnostic ranges for free kappa and free lambda immunoglobulin light chains: relative sensitivity for detection of monoclonal light chains. *Clin Chem*. 2002; 48:1437–44. [PubMed: 12194920]
11. Hutchison CA, Harding S, Hewins P, Mead GP, Townsend J, Bradwell AR, et al. Quantitative assessment of serum and urinary polyclonal free light chains in patients with chronic kidney disease. *Clin J Am Soc Nephrol*. 2008; 3:1684–90. [PubMed: 18945993]
12. Lachmann HJ, Gallimore R, Gillmore JD, Carr-Smith HD, Bradwell AR, Pepys MB, Hawkins PM. Outcome in systemic AL amyloidosis in relation to changes in concentration of circulating free immunoglobulin light chains following chemotherapy. *Br J Haematol*. 2003; 122:78–84. [PubMed: 12823348]
13. Dispenzieri A, Zhang L, Katzmann JA, Snyder M, Blood E, DeGoey R, et al. Appraisal of immunoglobulin free light chain as a marker of response. *Blood*. 2008; 111:4908–15. [PubMed: 18364469]
14. Dispenzieri A, Lacy MQ, Katzmann JA, Rajkumar SV, Abraham RS, Hayman SR, et al. Absolute values of immunoglobulin free light chains are prognostic in patients with primary systemic amyloidosis undergoing peripheral blood stem cell transplantation. *Blood*. 2006; 107:3378–83. [PubMed: 16397135]
15. Sancharawala V, Seldin DC, Magnani B, Skinner M, Wright DG. Serum free light chain responses after high-dose intravenous melphalan and autologous stem cell transplantation for AL (primary) amyloidosis. *Bone Marrow Transplant*. 2005; 36:597–600. [PubMed: 16044137]
16. Cohen AD, Zhou P, Chou J, Teruya-Feldstein J, Reich L, Hassoun H, et al. Risk-adapted autologous stem cell transplantation with adjuvant dexamethasone +/- thalidomide for systemic light-chain amyloidosis: results of a phase II trial. *Br J Hematol*. 2007; 139:224–33.

17. Pinney JH, Lachmann HJ, Bansal L, Wechalekar AD, Gilbertson JA, Rowczenio D, et al. Outcome in renal AL amyloidosis after chemotherapy. *J Clin Oncol.* 2011; 29:674–81. [PubMed: 21220614]
18. Kumar SK, Dispenzieri A, Lacy MQ, Hayman SR, Buadi FK, Zeldenrust SR, et al. Changes in serum free light chain rather than intact monoclonal immunoglobulin levels predict outcome following therapy in primary amyloidosis. *Am J Hematol.* 2011; 86:251–5. [PubMed: 21328431]

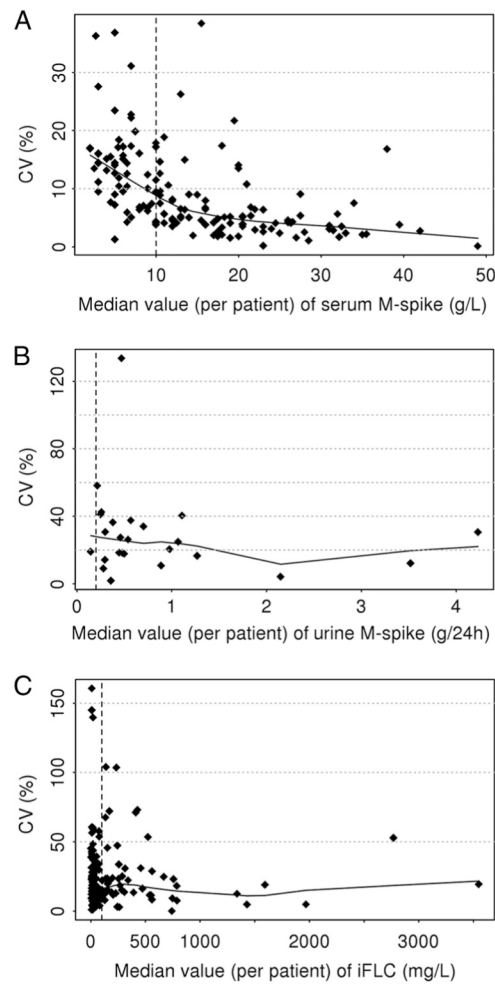


Fig. 1. Mean value of the serum M-spike (A), urine M-spike (B), and serum iFLC (C) for each patient's serial sample set graphed against the CV for that set of samples. Dashed lines indicate current recommendations for minimal values for monitoring.

Table 1
Number of patients and observations for each test

Assay	Unique patients, n	Mean observations per patient, n	Median observations per patient, n	Maximum observations per patient, n	Mean value
Serum M-spike, g/L	158	5.1	4	16	14.3
Measurable serum M-spike, g/L ^a	90	5.2	4	16	20
Involved immunoglobulin, g/L	148	5.1	4	16	19.5
Urine M-spike, g/24 h	25	5.2	4	11	1.0
iFLC, mg/L	158	5.2	4	16	209.4
Measurable iFLC, mg/L ^b	52	4.8	4	11	276.2
uFLC, mg/L	157	5.2	4	16	11.4
rFLC	157	5.2	4	16	26.2 ^c

^aValues >10 g/L.

^bValues >100 mg/L.

^cRange 0.001–1860.

Table 2
Total CVs for patient sample sets^a

Diagnosis	Measurable serum M-spike	Serum IgG	Measurable urine M-spike	Measurable serum iFLC	Serum iFLC	Serum uFLC	Serum rFLC
All diagnoses	8.1 (90)	13.0(148)	35.8 (25)	28.4 (52)	34.9 (158)	45.2 (157)	47.5 (157)
MGUS	6.3 (15)	16.9 (32)	58.2 (1)	15.9 (7)	32.8 (35)	34.7 (35)	41.2 (35)
SMM	6.1 (44)	9.1 (52)	19.2 (9)	17.0 (17)	24.6 (52)	42.8 (52)	46.2 (52)
MM	10.8 (31)	14.2 (64)	44.3 (15)	34.7 (28)	41.4 (71)	50.0 (70)	50.5 (70)

^aData are % (n).

Table 3
Percentage decrease needed in each assay for statistical significance at various probability thresholds

Test	n	Total CV	<i>P</i> = 0.5	<i>P</i> = 0.8	<i>P</i> = 0.9	<i>P</i> = 0.95
Serum M-spike	158	11.6	10.5	19.0	23.6	27.5
Measurable serum M-spike ^a	90	8.1	7.4	13.7	17.2	20.1
Involved immunoglobulin (IgG)	148	13.0	11.7	21.0	26.1	30.3
Urine M-spike	25	35.8	28.9	47.7	56.5	62.9
iFLC	158	34.9	28.3	46.9	55.6	62.0
Measurable iFLC ^b	52	28.4	23.7	40.2	48.3	54.5
uFLC	157	45.2	35.0	55.9	65.1	71.4
rFLC	157	36.4	36.4	57.7	66.9	73.2

^aValues >10 g/L.

^bValues >100 mg/L.

Table 4
Total, analytical, and biological CVs^a

	Measurable serum M-spike ^b	Urine M-spike (200 mg/24 h)	Measurable serum iFLC ^c	Serum IgG
Total CV	8.1	35.8	28.4	13
Analytical CV	2.1	4.5	5.8	4.2
Biological CV	7.8	35.5	27.8	12.3

^aTotal CVs are from Table 2, analytical CVs are from laboratory validation studies, and biological CVs are derived from the relationship of the biological CV being equal to the square root of the difference of the square of the total CV minus the square of the analytical CV.

^bValues >10 g/L.

^cValues >100 mg/L.