Corrie M. de Kat Angelino and Joannes F.M. Jacobs*

External quality assessment of M-protein diagnostics: a realistic impression of the accuracy and precision of M-protein quantification

https://doi.org/10.1515/cclm-2020-1810
Received December 10, 2020; accepted January 18, 2021;
published online January 28, 2021

Abstract

Objectives: Studies that investigate the accuracy and precision of M-protein quantification are scarce. These studies are prone to give a biased view, since they are exclusively performed by institutions with international top-expertise on M-protein diagnostics. To obtain a realistic impression of the accuracy and precision of M-protein quantification, we studied results of 73 laboratories participating in the Dutch External Quality Assessment (EQA) program for M-protein diagnostics.

Methods: To measure accuracy, healthy serum was spiked with respectively 1 and 5 g/L human IgG-kappa monoclonal antibody daratumumab. To measure precision, five sera were selected to be repeatedly send to all blinded EQA-participants.

Results: The reported concentrations for the EQA-sample spiked with 5 g/L daratumumab ranged from 2.6 to 8.0 g/L (mean 4.9 g/L, between-laboratory CV = 23%). 98% of the participants detected and correctly characterized the 1 g/L daratumumab band. Both the accuracy (mean 1.7 g/L) and precision (between-laboratory CV = 46%) of this 1 g/L M-protein was poor. In the five EQA-samples that were repeatedly send to the same 73 participating laboratories, between-laboratory precision (mean CV = 25%) was significantly different than the within-laboratory precision (mean CV = 12%). Relatively poor precision was observed in sera with small M-proteins.

Conclusions: The EQA-data reveal a large variation in reported M-protein concentrations between different laboratories. In contrast, a satisfactory within-laboratory precision was observed when the same sample was repeatedly analyzed. The M-protein concentration is correlated with both accuracy and precision. These data indicate that M-protein quantification to monitor patients is appropriate, when subsequent testing is performed within the same laboratory.

Keywords: accuracy; electrophoresis; external quality assessment; monoclonal gammopathy; M-protein diagnostics; precision; serum protein electrophoresis.

Introduction

Monoclonal gammopathies (MGs) are characterized by a clonal expansion of plasma cells and the excretion of a monoclonal immunoglobulin (M-protein). MGs encompass a broad clinical spectrum that ranges from MG of undermined significance (MGUS) to life-threatening diseases such as multiple myeloma and amyloid light chain (AL) amyloidosis [1, 2].

M-protein detection and quantification plays an important role in the diagnosis and management of patients with monoclonal gammopathies [3]. M-protein diagnostics is most commonly performed using electrophoretic methods, supplemented with immunoassays for quantification and clonality testing [4, 5]. Serum protein electrophoresis (SPEP) is often the first test to screen for MG. The presence of an M-protein is confirmed and further characterized using the more sensitive method of immunofixation electrophoresis (IFE). Using the SPEP densitogram, the M-protein concentration is calculated based on the percentage of the gated M-protein (M-spike) and the total serum protein concentration.

The procedure of M-protein quantification involves a manual step of selecting the limits of the M-spike [3]. Even though large parts of the M-protein analyses are fully automated, this judgment remains subjective. Variation in M-protein quantification is thus affected both by methodological- and inter-operator differences. This hampers accurate and precise M-protein quantification, which is critical for monitoring disease activity and to assess response to therapy [6].

***Corresponding author: Joannes F.M. Jacobs**, PhD, MD, Department of Laboratory Medicine, Laboratory Medical Immunology (Route 469), Radboud University Medical Center, Geert Grooteplein 10, 6525 GA Nijmegen, The Netherlands, Phone: +31 (0)24 3614625, Fax: +31 (0)24 3619415, E-mail: H.Jacobs@Radboudumc.nl

Corrie M. de Kat Angelino, Department of Laboratory Medicine, Laboratory Medical Immunology, Radboud University Medical Center, Nijmegen, The Netherlands

Several groups have studied variation of M-protein quantification either in clinically stable patients [7–9], or in sera spiked with human therapeutic monoclonal antibody [10–12]. It was shown that the concentration of the M-protein, the migration pattern and the polyclonal antibody background, all affect the accuracy and precision of M-protein quantification [7–12]. Depending on the exact nature of the selected samples, the mean analytical variation of M-protein quantification within a single institute ranged from CVs of 5.0–8.9%. However, these studies do not provide information on the level of variation between different centers. Moreover, the above mentioned studies are prone to give a biased view, since they are exclusively performed by institutions with international top-expertise on M-protein diagnostics.

To obtain a realistic impression of the accuracy and precision of M-protein quantification we studied results of 73 laboratories that participate in the Dutch External Quality Assessment (EQA) program for M-protein diagnostics. Participants were not aware that for this study specific EQA-sera were prepared that were repeatedly send to the same participants over a period of five years.

Materials and methods

This study was performed in collaboration with the Dutch EQA Organization for Quality Assessment of Laboratory Diagnostics (SKML). To assess accuracy, sera containing a known concentration of γ-migrating M-protein were prepared by spiking 1 and 5 g/L of the therapeutic monoclonal antibody daratumumab (Darzalex, Janssen Pharmaceutica) into serum of an individual without a monoclonal gammopathy and with a normal γ-fraction. All samples were carefully homogenized prior to aliquoting. To investigate precision we aliquoted EQA samples of five different patients and shipped these at room temperature 3–5 times to the same 73 laboratories over a period of five years. Large-volume sera pools with one M-protein could be prepared from patients with a monoclonal gammopathy who had a clinical indication for a plasmapheresis procedure. Each participant received 1 mL per sample to perform all the M-protein diagnostics on. Depending on the available sample volume, they were split in 3–5 aliquots, large enough to be shared with all 73 laboratories. Prior to shipment, these aliquots were stored at –20 °C.

Participants were not aware that the serum EQA-samples were either spiked with daratumumab or that a specific EQA-sample was repeatedly send around to the same participant over a period of five years. A total of 73 laboratories participated in the Dutch M-protein EQA program. These included both university medical centers (n=8) and general hospitals. In this study not all 73 participants returned results on all EQA-samples, for example because they missed the deadline. Laboratories reported which instrumentation was used and their results in the online reporting and scoring system named 'multi-sample evaluation (MUSE)' [13]. M-protein concentrations were rounded to 0.1 g/L. Precision calculations did not take into account if a participating laboratory switched methods or instrumentation during the five-year period of this EQA-study. Laboratories received feedback reports that follow the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) consensus.

All samples used in this study were obtained with informed consent, are coded and anonymized prior to be used in the Dutch EQA program. This study was performed in accordance with the Helsinki Declaration and was approved by the authors' Institutional Review Board (#2016-2356).

Results

Accuracy

The electrophoretic data of 5 g/L daratumumab spiked into serum with a normal γ-fraction are shown in Figure 1A. All 67 participating laboratories that returned their results detected the IgG-kappa monoclonal band. The mean reported concentration of this EQA sample was 4.9 g/L. The black bars in the histogram of Figure 1B show that the reported M-protein concentrations ranged from 2.6-8 g/L (CV = 23%). These results were compared to the reported M-protein concentrations of 16 international top-reference centers of a normal serum that was similarly spiked with 5 g/L daratumumab (white bars in Figure 1B), as recently published by Turner et al. [11]. The mean reported concentration by these 16 top reference centers was 4.9 g/L, with concentrations that ranged from 3-6.3 g/L (CV = 21%).

In the EQA-sample with 1 g/L daratumumab (Figure 1C), all but one laboratory (98%) were able to detect and correctly characterize the IgG-kappa M-protein. 24% of the participants judged that they could not accurately spike this M-protein and reported a concentration <3 g/L. The remaining 76% of the participants spiked the monoclonal band and reported an average M-protein concentration of 1.7 g/L (range 0.1–4.4 g/L, CV = 46%).

Between-laboratory precision

To investigate precision we aliquoted EQA samples of five different patients of whom we had ample material (because of a clinical indication for a plasmapheresis procedure) to send around at least three times to the same 73 laboratories. The characteristics of these five EQA samples are provided in Table 1. All five M-proteins migrated in the γ-fraction of the SPE (Figure 2) and were quantified by the participating laboratories based on the M-spike.

The between-laboratory CV values in these five EQA samples ranged from 15.5 to 36.9% (mean CV = 25%), see Table 1. The highest variation between laboratories was observed in samples with the smallest M-protein concentration.

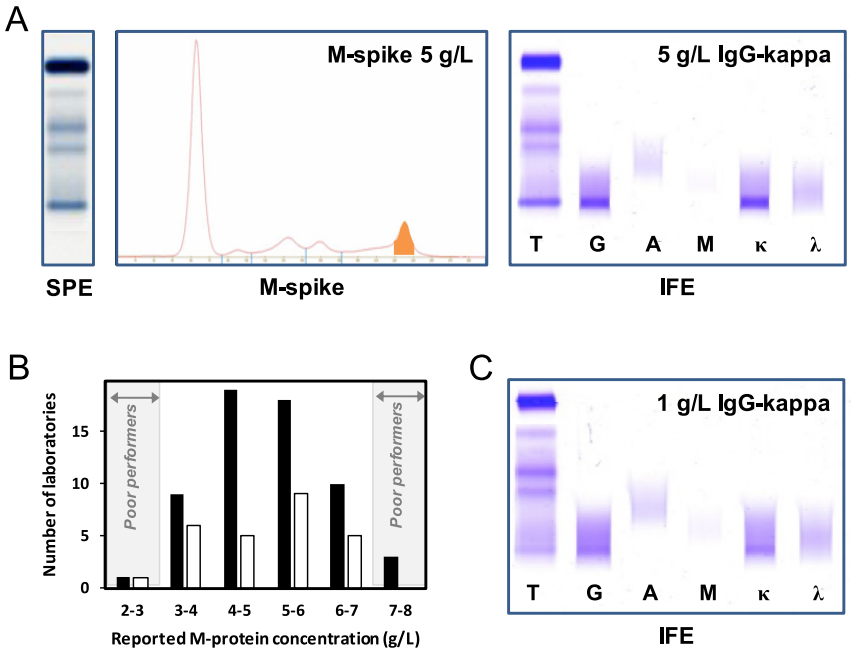

Figure 1: Detection and accuracy of M-protein quantification.
(A) Serum protein electrophoresis (SPE), densitogram and immunofixation electrophoresis (IFE) clearly identifies the daratumumab IgG-kappa band (5 g/L). The M-spike is shown in orange. (B) Histogram showing the reported M-protein concentration of the dara-spiked serum (5 g/L)of the Dutch laboratories that participate in the EQA program (black bars). As a reference, the white bars show the reported M-protein concentrations of 16 international top-reference centers of a normal serum that was similarly spiked with 5 g/L daratumumab, as recently published by Turner et al. [11]. The gray-shaded area highlights poor performer laboratories that quantitate the 5 g/L dara-spike as either <3 g/L or >7 g/L. (C) IFE of normal EQA-serum that was spiked with 1 g/L daratumumab. The IgG-kappa band is faintly visible at the cathodal end of the γ-fraction.

Table 1: Precision of M-protein quantification in five EQA-sera.

ID[a]	Number of participants[b]	M-protein	Mean conc., g/L	Between-lab CV, %[c]	Within-lab CV, %[c]
#1 (4×)	73	IgG-kappa	7.1	22.5	9.4
#2 (3×)	67	IgG-kappa	3.3	23.5	16.6
#3 (3×)	67	IgG-lambda	3.3	36.9	14.1
#4 (3×)	71	IgM-kappa	39.4	15.5	7.1
#5 (5×)	67	IgM-kappa	6.7	25.7	14.2

[a]Indicated in brackets is the number of times the EQA sample is distributed among participants. [b]A minimum of two analyses per EQA sample are required to be incorporated in the analyses. [c]Indicated are mean CV values.

Within-laboratory precision

In all five EQA samples the within-laboratory variation (mean CV = 12%) was significantly lower compared to the variation observed between laboratories (Table 1). Again, the variation was inversely correlated with the M-protein concentration. It is important to note that the within-laboratory CV indicated in Table 1, is the mean variation within laboratories. On average CV values were well below the acceptable upper limit of CV = 20% (arbitrary cut-off). Figure 3 illustrates that within the total group of laboratories it is possible to differentiate laboratories that perform excellent with CV values <5% for all five samples. In contrast, to that, we observed poor performers that have an average within-laboratory precision CV>30%.

Discussion

Whereas base-line M-protein measurements play a role in the diagnosis of monoclonal gammopathies [14], serial M-protein measurements serve to monitor the patient's disease progression or response to treatment [6]. Acceptable accuracy and between-laboratory variation is important to ensure that the diagnosis made in a certain patient is not dependent on where the M-protein diagnostics is performed. Acceptable within-laboratory precision is important to ensure that reported M-protein changes indeed reflect a true change in the patient's pathological state and is not based on analytical variation.

Despite its clinical importance, there is limited information available regarding the accuracy and analytical variation of M-protein quantification. We hypothesize that current literature creates an overly optimistic image since these studies are exclusively performed by institutions with an international top-reputation on M-protein diagnostics [10–12]. To obtain a realistic impression of the accuracy and precision of M-protein quantification we studied results of 73 laboratories that participate in the Dutch EQA program for M-protein diagnostics.

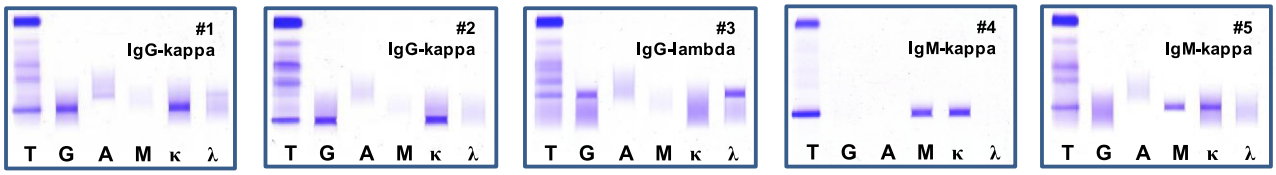

Figure 2: Immunofixation electrophoresis of the EQA samples.
Each panel shows a representative IFE of the five EQA-sera. Each M-protein is characterized by a single band that migrates in the γ-region. The exact migration pattern, the concentration of the M-protein and the polyclonal background varies for each EQA-sample.

Regarding the accuracy of M-protein quantification we conclude that on average the Dutch laboratories quantitate the 5 g/L M-protein EQA sample well. However, the range of reported concentrations was wide (CV = 23%), among them several laboratories that strongly underestimated (<3 g/L) or overestimated (>7 g/L) the true M-protein concentration. The accuracy further decreased in case the laboratories had to report on an EQA sample with 1 g/L M-protein (CV = 46%). These results are very much in line with data presented by Turner et al., who showed similar accuracy data among 16 top-reference centers [11]. Because of the poor accuracy observed for the quantification of small M-proteins, we agree with those laboratories that do not report concentrations of detected M-proteins <2 g/L.

To reach optimal within-laboratory precision over longer periods of time it is crucial that in each laboratory working procedures are defined for consistent manual gating of the M-spike. For reliable serial M-protein monitoring, that protocol should not change over the years and should be independent on the operator. On average the Dutch laboratories in this study have an acceptable within-laboratory variation (mean CV = 12%). This within-laboratory precision in the Dutch laboratories (CV = 12%) is clearly not as good as the precision reported by centers specialized in M-protein diagnostics. Depending on the exact nature of the selected samples, the mean analytical variation of M-protein quantification within a single institute ranges from CVs of 5.0–8.9% [10–12]. A possible limitation of this study is that some laboratories might have switched methods during the five-year period of this EQA-study. A laboratory might for example have changed their peak integration method which impacts the M-spike [15]. Since we could not correct for such methodological changes in this study, this may have negatively impacted precision in these laboratories.

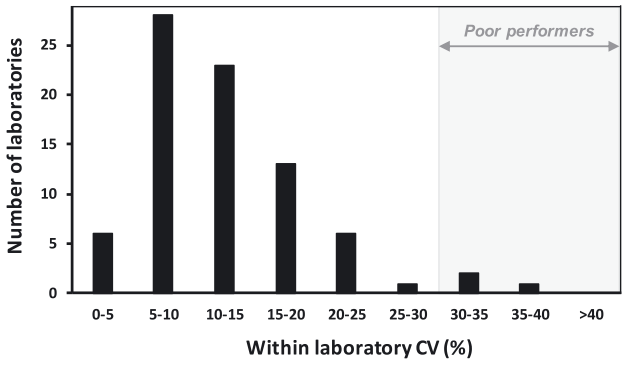

Figure 3: Within-laboratory precision.
Histogram showing the average within laboratory precision (CV%) of M-protein quantification performed in five EQA-samples. The gray-shaded area highlights poor performer laboratories with an average CV > 30%.

The International Myeloma Working Group (IMWG) Response Criteria are defined by relative changes in the M-protein concentration [6]. To correctly differentiate a patient with very good partial response (≥90% reduction in measurable intact M-protein) from partial response (≥50%), minimal response (≥25%), and progressive disease (≥25% increase), it is crucial that the variability in M-protein measurements is small. Especially since patients may receive therapy in different institutions over the course of their disease, and the patient's serum samples may be sent to different laboratories with different assay methods and gating practices. The average between-laboratory CV of 25% presented in this study, emphasizes that it is important that response assessment should be done with serial M-protein results obtained in the same laboratory. It also creates awareness that a small group of laboratories can be defined as poor performers, their within-laboratory imprecision (CV > 30%) is insufficient to perform a reliable response assessments that follows the IMWG guidelines.

To help interpret follow-up of M-protein quantification outside clinical studies, three independent groups calculated M-protein reference change values (RCV). RCV can act as a tool to detect significant changes in a serial laboratory result, in order to identify a true change in the pathological state of a patient with an MG. The RCV is based on a combination of analytical variation and biological variation. The M-protein RCV reported by these top-reference groups range from 25 to 38% [7–9]. In a minority of Dutch laboratories harmonized working procedures are clearly not in place, and within-laboratory CV > 30% were observed. In practice this means that these poor performers might

quantitate the IgG-kappa M-protein in the first EQA-sample as 6 g/L and the same EQA-sample in another year as 11 g/L. In these centers, clinicians might easily misinterpret this analytical variation as a sign of disease progression.

In our opinion the M-protein EQA program can educate participants to recognize rare M-protein patterns [16], and assist in the analytical validation of novel M-protein diagnostics [17]. In addition, it is recommended that an EQA-sample is independently interpreted by several operators within the same laboratory. Evaluation of these cumulative results can further facilitate the harmonization process of the manual peak integration within the team.

Theoretically M-protein-instability in stored sera could be a confounding factor in this study. To test for sample stability during the time that sera were stored at –20 °C, we calculated the mean reported M-protein concentration of both the first and last sample of each of the five EQA-sera in this study. No significant differences were observed between first-shipped samples and their counterparts that were stored in the freezer for several years (data not shown). We also observed no significant differences in reported results in samples that were sent in summer vs. other seasons (data not shown). From this we conclude that the M-protein is stable under these storage/shipment-conditions and variation in M-protein quantification is solely attributed to analytical variation in the participating laboratories.

Overall this study illustrates that on a national level strong differences are observed regarding the accuracy and precision with which laboratories quantitate M-proteins. The accuracy of the Dutch laboratories resembles the accuracy reached by 16 international top-reference centers. In contrast to that, the average within-laboratory precision of M-protein quantification in the Dutch laboratories (CV = 12%) is not as good as the precision reported by centers specialized in M-protein diagnostics. Overall, the performance of M-protein measurements of laboratories that participate in national EQA programs are comparable to those performed by top-expertise laboratories. With exception of a small group of poor-performing laboratories, our data indicate that M-protein quantification to monitor patients over time is appropriate when subsequent testing is performed within the same laboratory.

Acknowledgments: We would like to thank all patients who willingly donated blood required for the EQA program. The Dutch EQA program is made possible by the Dutch Foundation for Quality Assessment in Medical Laboratories (SKML).

Research funding: The SKML financed the acquisition of daratumumab to spike healthy serum with. JFMJ received a research grant from the Dutch Cancer Society (No. 10817).

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Competing interests: Authors state no conflict of interest.

Ethical approval: This study was performed in accordance with the Helsinki Declaration and was approved by the authors' Institutional Review Board (#2016-2356).

References

1. Glavey SV, Leung N. Monoclonal gammopathy: the good, the bad and the ugly. Blood Rev 2016;30:223–31.
2. Rollig C, Knop S, Bornhauser M. Multiple myeloma. Lancet 2015; 385:2197–208.
3. Willrich MAV, Murray DL, Kyle RA. Laboratory testing for monoclonal gammopathies: focus on monoclonal gammopathy of undetermined significance and smoldering multiple myeloma. Clin Biochem 2018;51:38–47.
4. Dimopoulos M, Kyle R, Fermand JP, Rajkumar SV, San Miguel J, Chanan-Khan A, et al. Consensus recommendations for standard investigative workup: report of the international myeloma workshop consensus panel 3. Blood 2011;117:4701–5.
5. Singh G. Serum and urine protein electrophoresis and serum-free light chain assays in the diagnosis and monitoring of monoclonal gammopathies. J Appl Lab Med 2020;5:1358–71.
6. Kumar S, Paiva B, Anderson KC, Durie B, Landgren O, Moreau P, et al. International Myeloma Working Group consensus criteria for response and minimal residual disease assessment in multiple myeloma. Lancet Oncol 2016;17:e328–46.
7. Evliyaoglu O, van Helden J, Jaruschewski S, Imohl M, Weiskirchen R. Reference change values of M-protein, free light chain and immunoglobulins in monoclonal gammopathy. Clin Biochem 2019; 74:42–6.
8. Katzmann JA, Snyder MR, Rajkumar SV, Kyle RA, Therneau TM, Benson JT, et al. 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. Clin Chem 2011;57:1687–92.
9. Salamatmanesh M, McCudden CR, McCurdy A, Booth RA. Monoclonal protein reference change value as determined by gel-based serum protein electrophoresis. Clin Biochem 2018;51: 61–5.
10. Jacobs JFM, Turner KA, Graziani MS, Frinack JL, Ettore MW, Tate JR, et al. An international multi-center serum protein electrophoresis accuracy and M-protein isotyping study. Part II: limit of detection and follow-up of patients with small M-proteins. Clin Chem Lab Med 2020;58:547–59.
11. Turner KA, Frinack JL, Ettore MW, Tate JR, Graziani MS, Jacobs JFM, et al. An international multi-center serum protein electrophoresis accuracy and M-protein isotyping study. Part I: factors impacting limit of quantitation of serum protein electrophoresis. Clin Chem Lab Med 2020;58:533–46.

12. Plebani M. New insights on the analytical performances for detecting and quantifying monoclonal proteins. Clin Chem Lab Med 2020;58:457–9.
13. Thelen MHM, Jansen RTP, Weykamp CW, Steigstra H, Meijer R, Cobbaert CM. Expressing analytical performance from multi-sample evaluation in laboratory EQA. Clin Chem Lab Med 2017;55:1509–16.
14. Rajkumar SV, Dimopoulos MA, Palumbo A, Blade J, Merlini G, Mateos MV, et al. International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma. Lancet Oncol 2014; 15:e538–48.
15. Schild C, Wermuth B, Trapp-Chiappini D, Egger F, Nuoffer JM. Reliability of M protein quantification: comparison of two peak integration methods on capillarys 2. Clin Chem Lab Med 2008;46: 876–7.
16. Bosman MCJ, Schreurs RHP, Nieuwenhuizen L, Bakkeren DL, Jacobs JFM. Broad bands observed in serum electrophoresis should not Be taken lightly. Clin Chem 2019;65:618–21.
17. Jacobs JFM, Haagen IA, Lodder A, van der Kroft C, de Kat Angelino CM, Croockewit S, et al. Analytical validation of the Hevylite assays for M-protein quantification. Clin Chem Lab Med 2018;56:1169–75.