

# Application of the Characteristic Function to Evaluate and Compare Analytical Variability in an External Quality Assessment Scheme for Serum Ethanol

Wim Coucke,<sup>1\*</sup> Corine Charlier,<sup>2</sup> Willy Lambert,<sup>3</sup> Frank Martens,<sup>4</sup> Hugo Neels,<sup>5,6</sup> Jan Tytgat,<sup>7</sup> Philippe Van de Walle,<sup>1</sup> André Vanescote,<sup>8</sup> Pierre Wallemacq,<sup>9</sup> Sarah Wille,<sup>10</sup> and Alain G. Verstraete<sup>11,12</sup>

**BACKGROUND:** As a cornerstone of quality management in the laboratory, External Quality Assessment (EQA) schemes are used to assess laboratory and analytical method performance. The characteristic function is used to describe the relation between the target concentration and the EQA standard deviation, which is an essential part of the evaluation process. The characteristic function is also used to compare the variability of different analytical methods.

**METHODS:** We fitted the characteristic function to data from the Belgian External Quality Assessment program for serum ethanol. Data included results from headspace gas chromatography and the enzymatic methods of Abbott, Roche, Siemens, and Ortho-Clinical Diagnostics. We estimated the characteristic function with weighted nonlinear regression. By introducing dummy variables, we rewrote the original formula of the characteristic function to assess statistical inference for comparing the variability of the different analytical methods.

**RESULTS:** The characteristic function fitted the data precisely. Comparison between methods showed that there was little difference between the estimated variability for low concentrations, and that the increase in SD with increasing target concentration was slower for Abbott and Roche than for the other methods.

**CONCLUSIONS:** The characteristic function can successfully be introduced in clinical schemes, although its applicability to fit the data should always be assessed. Because of its easy parameterization, it can be used to assess differences in performance between analytical methods and to assess laboratory performance. The characteristic

function also offers an alternative framework for coefficients of variation to describe variability of analytical methods.

© 2015 American Association for Clinical Chemistry

A correct estimation of the variability in External Quality Assessment (EQA)<sup>13</sup> schemes, an important parameter used for follow-up of laboratories and analytical methodologies, has received wide interest in the past (1–3). It is not only used to evaluate different laboratories or different assays, e.g., by means of *z*-scores, but it may also reveal information about the state of the art of the analytical methods.

To evaluate laboratories by means of *z*-scores, EQA organizers have ample choice of methods to obtain a standard deviation. They can calculate the SD directly from the reported results; they may also obtain SDs from external sources, such as legislative documents, published literature, or historical data (4). In particular, when not enough data are available, deriving SDs from external sources becomes interesting.

To evaluate analytical methods, EQA organizers limit themselves to reporting the variability of several methods for a particular target concentration and matrix. Various studies have underlined the advantage of combining results of different EQA rounds or samples (5–8). Moreover, methods have been published to combine data from different centers or samples of the same target concentration (9). However, to the best of our knowledge, combining data from samples of different target concentrations to obtain a reliable estimate of the SD has not yet been done in the domain of EQA for clinical laboratories.

<sup>1</sup> Quality of Medical Laboratories, Scientific Institute of Public Health, Brussels, Belgium; <sup>2</sup> Department of Toxicology, University Hospital of Liège, Liège, Belgium; <sup>3</sup> Laboratory of Toxicology, Ghent University, Ghent, Belgium; <sup>4</sup> Department of Clinical Chemistry, AZ Groeninge Hospital, Kortrijk, Belgium; <sup>5</sup> Laboratory of Toxicology, ZNA Stuivenberg, Antwerp, Belgium; <sup>6</sup> Toxicological Center, University of Antwerp, Antwerp, Belgium; <sup>7</sup> Laboratory of Toxicology and Pharmacology, Katholieke Universiteit Leuven, Leuven, Belgium; <sup>8</sup> Gamma Medic SPRL, Isnes, Belgium; <sup>9</sup> Louvain Centre for Toxicology and Applied Pharmacology, Université Catholique de Louvain, Brussels, Belgium; <sup>10</sup> Toxicology Department, Federal Public Service Justice, National Institute of Criminalistics and Criminology, Brussels, Belgium; <sup>11</sup> Department of Laboratory Medicine, Ghent University,

Ghent, Belgium; <sup>12</sup> Department of Clinical Chemistry, Microbiology, and Immunology, Ghent University, Ghent, Belgium.

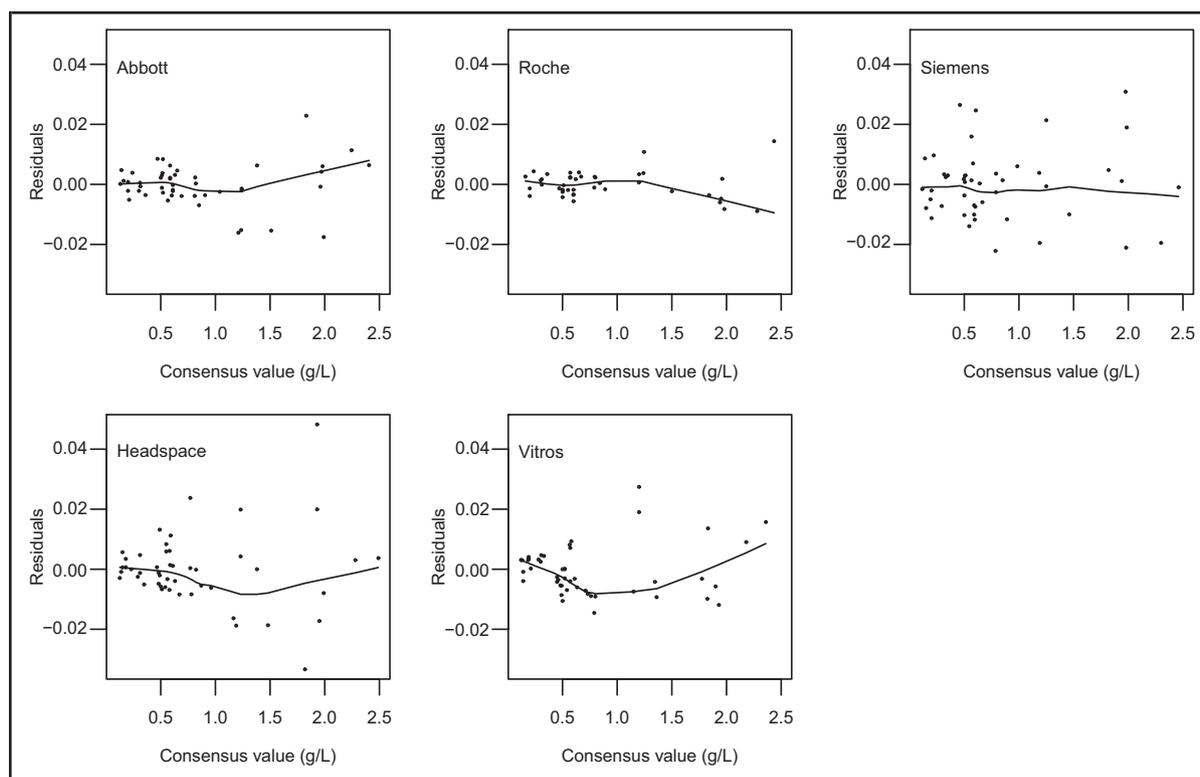
\* Address correspondence to this author at: Scientific Institute of Public Health, J. Wytsmanstraat 14, 1050 Brussels, Belgium. Fax +32-2-642-56-45; e-mail wim.coucke@wiv-isp.be.

Received February 23, 2015; accepted April 29, 2015.

Previously published online at DOI: 10.1373/clinchem.2015.240176

© 2015 American Association for Clinical Chemistry

<sup>13</sup> Nonstandard abbreviations: EQA, External Quality Assessment; LOWESS, locally weighted scatterplot smoothing.



**Fig. 1.** Residual dot plot for the curves of the characteristic function fit for the Qn estimator for each method.

Lines are the LOWESS regression obtained for each method. Flat LOWESS regression lines indicate perfect fit of the characteristic function; curved LOWESS regression lines indicate lack of ability of the characteristic function to model the data. A graph containing all the LOWESS regression lines together on 1 plot can be found in online Supplemental Fig. 1.

Recently, Thompson proposed a formula to describe the relation between the concentration of an analyte and the SD found in EQA programs (1). In basic form, the formula is given by Eq. (1):

$$SD = \sqrt{\alpha^2 + \beta^2 c^2} \quad (1)$$

where  $SD$  is the standard deviation of the EQA results,  $c$  is the concentration of the analyte; and  $\alpha$  and  $\beta$  are coefficients that have to be estimated via a nonlinear regression model.

By dividing both sides of Eq. (1) by the concentration, it can be rewritten (10) as follows:

$$CV = \sqrt{\frac{\alpha^2}{c^2} + \beta^2} \quad (2)$$

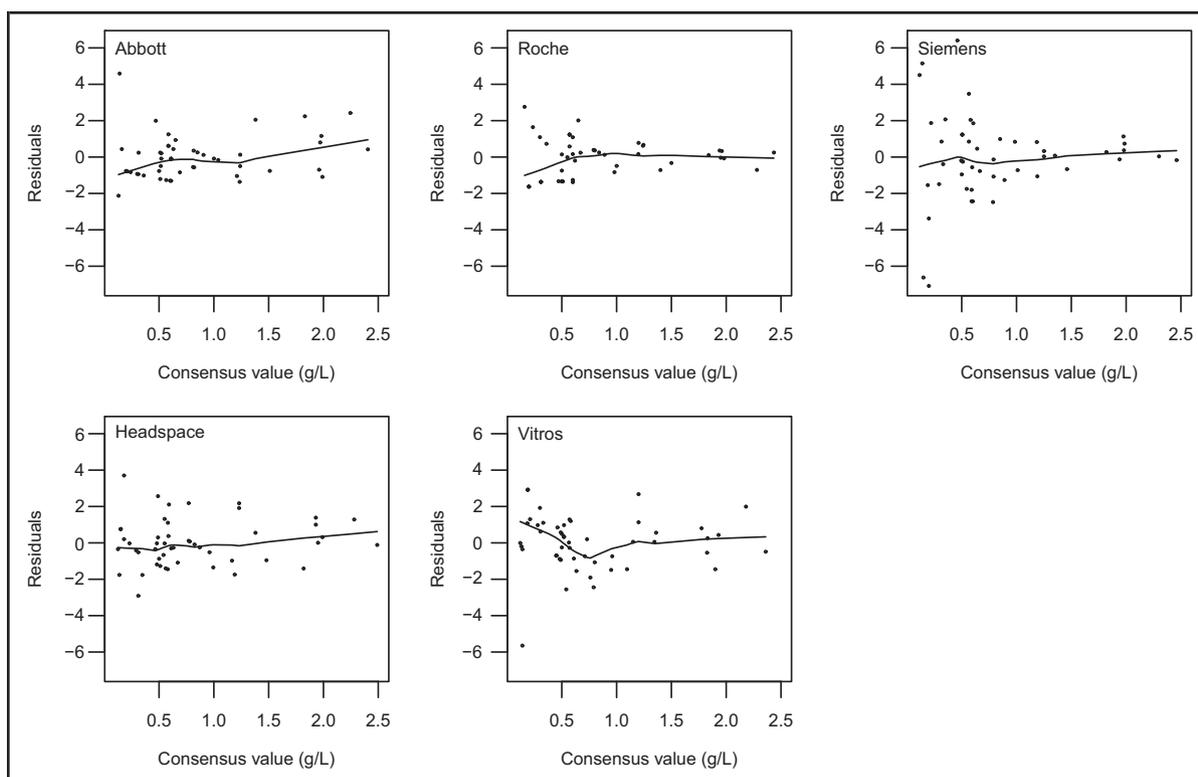
where  $CV$  is the coefficient of variation. The formula is also called the characteristic function (1), since it characterizes the performance of a particular system for a particular sample and measurand. The formula has been used in method validation (11, 12), total error estimation (13), and EQA schemes for analytical methods applied in the food industry and the environmental sector

(10, 14, 15). So far, only 1 application in the clinical laboratory has been described (16).

In EQA settings, the formula can be considered an alternative to the Horwitz function (3) with a wider applicability (1).

The parameters  $\alpha$  and  $\beta$  have a different meaning in describing the relation between concentration and the estimate of variability. The parameter  $\alpha$  principally determines the estimate of variability at low concentrations and could even be used to describe the limit of quantification (11). The parameter  $\beta$  affects the estimate of variability at higher concentrations and approaches the CV when  $\alpha$  is low or the concentration is high. The ratio  $\alpha/\beta$  determines the shape of the curve. The curve tends to a straight line when the ratio is small and a curved convex line when it is large.

Our aim here was to assess the feasibility of applying the characteristic function to EQA data and compare the different methods used in an EQA in the clinical setting. Serum ethanol, a parameter measured with totally different methods and relatively high precision, was used for this purpose.



**Fig. 2.** Residual dot plot for the curves of the characteristic function fit for the CV for each method.

Lines are the LOWESS regression obtained for each method. Flat LOWESS regression lines indicate perfect fit of the characteristic function; curved LOWESS regression lines indicate lack of ability of the characteristic function to model the data. A graph containing all the LOWESS regression lines together on 1 plot can be found in online Supplemental Fig. 2.

## Materials and Methods

### DATA

We used data from the Belgian EQA program for serum ethanol testing. In this program, the EQA organizers send 5–6 ethanol-spiked fresh serum samples twice a year. The participants are asked to treat the samples as routine samples and return the analysis results within 2 weeks.

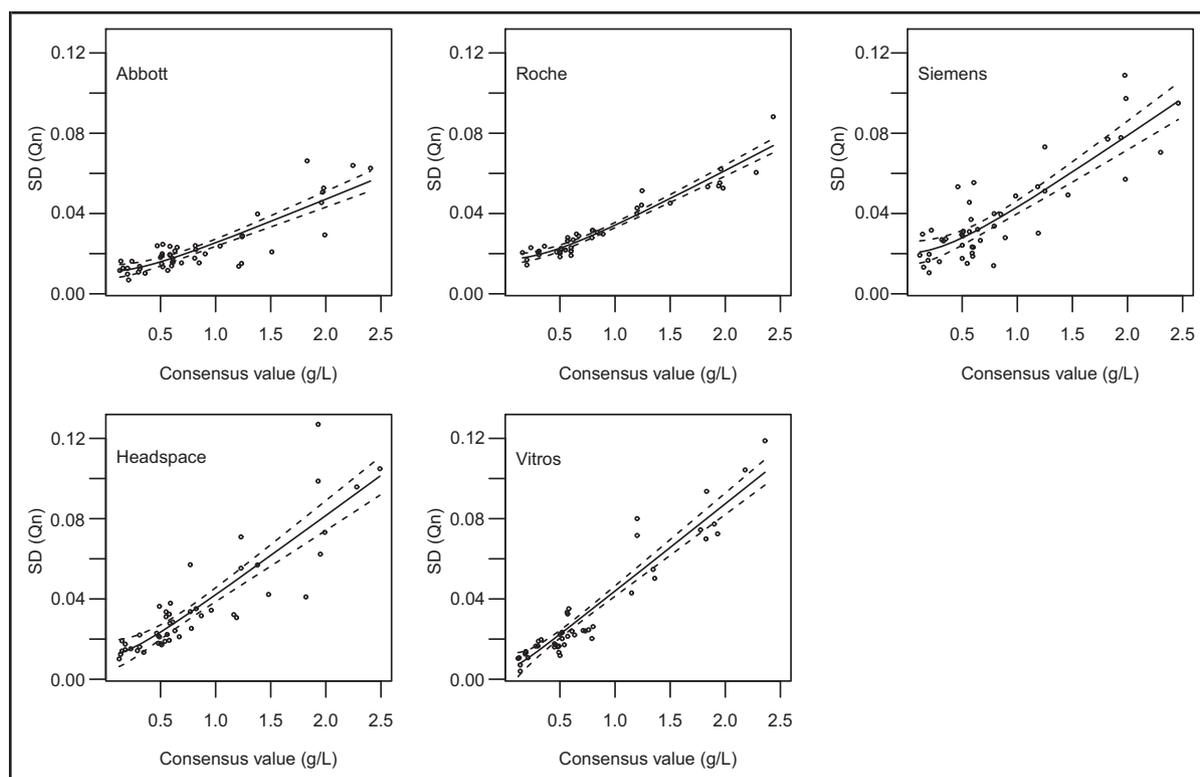
The median was used as a consensus value, a robust SD was calculated by the Qn estimator (17, 18), and the corresponding CV was obtained by dividing the estimated Qn by the consensus value.

Ethanol data included in the study were reported from 2010 to 2014. Overall, 50 samples were included in the study. Samples that contained ethanol concentrations below or equal to the limit of quantification (0.1 g/L) were discarded. Only SDs and CVs were evaluated for methods that were used for at least 15 different samples. SDs and CVs on the basis of <10 reported data points were not considered. Five methods were included in the study: headspace gas chromatography and the various

enzymatic methods of Abbott-Aeorset (Abbott Diagnostics), Dade (Dade Behring, Siemens), Roche (Roche Diagnostics), and Vitros (Ortho-Clinical Diagnostics). All kits from the same manufacturer were pooled each time into 1 group.

### CALCULATING THE CHARACTERISTIC FUNCTION

We solved Eq. (1) by means of nonlinear least squares regression with the Newton–Raphson algorithm and with  $n/e_i^2$  as weights (19), where  $n$  is the number of data that SDs were based on, and  $e_i$  is the residual value, which is the difference between the actual and predicted value by the modeled function of data point  $i$ . The nonlinear regression estimates were iterated, in which estimates of the residual values in the previous step were used to calculate weights in the next step. The iteration continued until the sum of the squared residual values was minimized at a precision of  $10^{-6}$ . An example of an iteration is provided in Supplemental Data, which accompanies the online version of this article at <http://www.clinchem.org/content/vol61/issue7>. For each estimated regression line, we calculated a 95% CI for the estimated



**Fig. 3.** Curves of the characteristic function for the Qn estimator for 5 different analytical methods for serum ethanol determination.

The graph labeled "All" shows the individual curves of each method together on 1 plot. A graph containing all the curves together on 1 plot can be found in online Supplemental Fig. 3.

line on the basis of the standard error of each predicted value.

For each curve, residual values were plotted against the target concentrations and visually inspected to check the goodness of fit of the curve by means of a residual dot plot. A nonparametric regression line on the basis of the locally weighted scatterplot smoothing (LOWESS) (20) line was drawn to indicate the direction of the points cloud. A line that followed a curvature up or down indicated a lack of ability of the characteristic function to model the data. A line that was flat with some potential peaks up or down pointed to the ability of the characteristic function to model the data.

#### COMPARISON BETWEEN METHODS

We compared methods by means of extra variables that can be used to distinguish between methods (so-called dummy variables). They allow 1 model to be fitted to data obtained for 2 different methods and allow an individual estimation of the parameter estimates for each method individually. Depending on the method, the extra variables have values of 0 or 1. Eq. (1) was transformed as follows:

$$SD = \sqrt{\alpha_1^2 + \alpha_2 X + (\beta_1^2 + \beta_2 X)c^2}. \quad (3)$$

The variable  $X$  was set at 0 for values from method 1 and at 1 for values from method 2. An example is given in the online Supplemental Data.

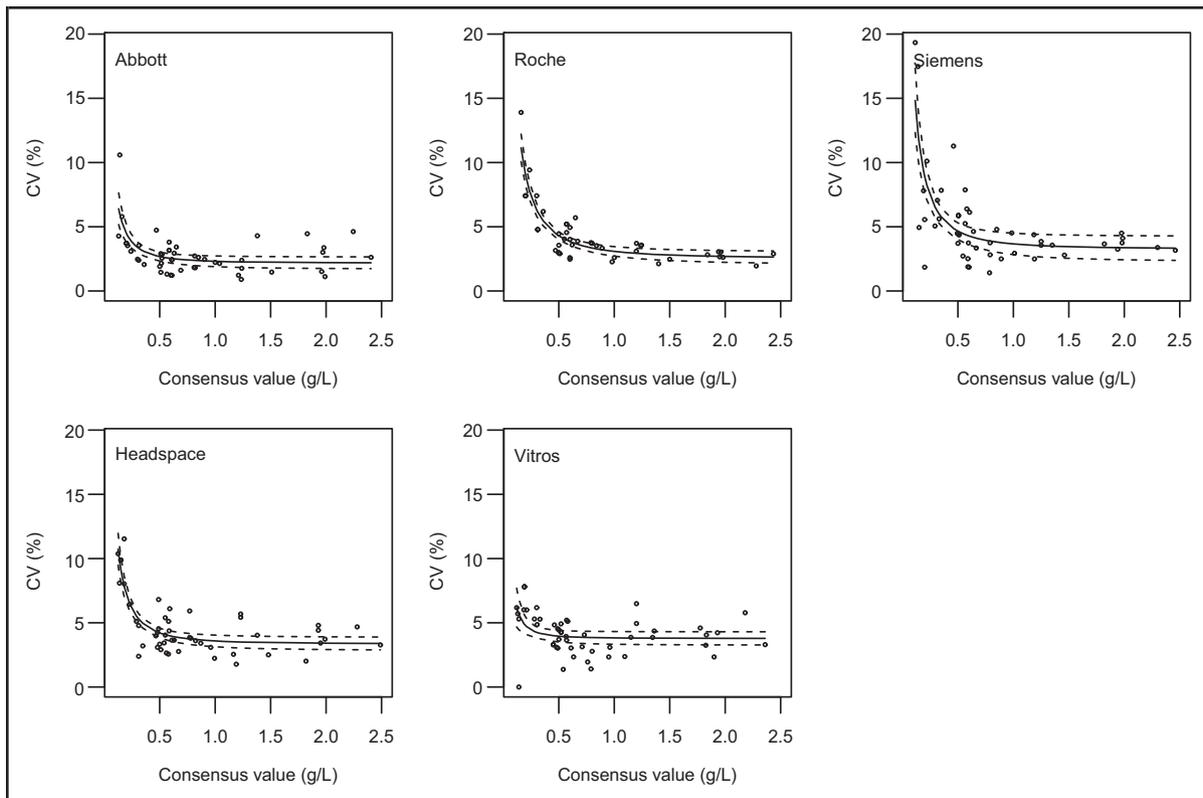
Equation (3) then becomes  $SD = \sqrt{\alpha_1^2 + (\beta_1^2)c^2}$  for method 1 and  $SD = \sqrt{\alpha_3^2 + (\beta_3^2)c^2}$  for method 2, where  $\alpha_3^2 = \alpha_1^2 + \alpha_2$  and  $\beta_3^2 = \beta_1^2 + \beta_2$ .

The values of  $\alpha_2$  and  $\beta_2$  inform about the differences between the methods, and their  $P$  values denote whether differences are significant.

We calculated  $P$  values for  $\alpha_2$  and  $\beta_2$  for each comparison between methods and corrected them for simultaneous hypothesis testing according to Tukey–Kramer. Dummy variables were used as well to compare the CVs of the methods.

#### Results

The residual dot plot for the Qn estimator for the 5 analytical methods for ethanol determination is shown in Fig. 1 and for the CV in Fig. 2. The LOWESS-smoothed lines showed a maximum deviation of 0.01 g/L, which



**Fig. 4.** Curves of the characteristic function for the CV for 5 different analytical methods for serum ethanol determination.

Graph labeled "All" shows the individual curves of each method together on 1 plot. A graph containing all the curves together on 1 plot can be found in online Supplemental Fig. 4.

was the rounding error of the reported results. This means that a possible lack of ability of the characteristic function to model the data was of a magnitude that was equal to or smaller than the rounding error of the reported results.

The curves of the characteristic function for modeling the Qn are depicted in Fig. 3 and for the CV in Fig. 4. The details of the values of  $\alpha$  and  $\beta$ , together with a

detail of the comparison between methods, are given in Table 1 for Qn and Table 2 for CV.

For Qn, the methods could be divided in 2 groups for the parameter  $\beta$ . The first group consisted of methods with a relatively slowly increasing SD with increasing ethanol concentration such as Abbott and Roche. The second group (consisting of headspace chromatography, Siemens, and Vitros) had a SD that increased relatively rapidly with in-

**Table 1.** Results from Eq. (1) ( $\alpha$ ,  $\beta$ ) for modeling the relation between the concentration and the SDs for each method individually, together with a detail of the comparison between methods.<sup>a</sup>

Laboratory method	Average reported results, n	Samples taken into account, n	$\alpha$	CI	$\beta$	CI
Abbott	49	13	0.011 <sup>b,c</sup>	0.0108-0.0113	0.0236 <sup>b</sup>	0.0232-0.024
Roche	45	96	0.0177 <sup>c</sup>	0.0169-0.0185	0.0288 <sup>b</sup>	0.0282-0.0295
Siemens	47	11	0.0206 <sup>c</sup>	0.0204-0.0208	0.0385 <sup>c</sup>	0.0383-0.0387
Headspace chromatography	48	16	0.0127 <sup>b,c</sup>	0.0126-0.0129	0.0406 <sup>c</sup>	0.0406-0.0407
Vitros	47	25	0.0064 <sup>b</sup>	0.0046-0.0078	0.0423 <sup>c</sup>	0.0411-0.0435

<sup>a</sup> Groups with the same letter next to the parameter estimate do not differ significantly for  $\alpha$  or  $\beta$ .

**Table 2.** Results from Eq. (2) ( $\alpha$ ,  $\beta$ ) for modeling the relation between the CV and the concentration for each method individually, together with a detail of the comparison between methods.<sup>a</sup>

Laboratory method	Average reported results, n	Samples taken into account, n	$\alpha$	CI	$\beta$	CI
Abbott	49	13	0.7831 <sup>b</sup>	0.571-0.949	2.173 <sup>b,c</sup>	1.631-2.603
Roche	45	96	1.7352 <sup>d</sup>	1.539-1.912	2.5472 <sup>b,c</sup>	1.986-3.005
Siemens	47	11	1.6637 <sup>c,d</sup>	1.328-1.942	3.265 <sup>b,c</sup>	2.062-4.131
Headspace chromatography	48	16	1.2736 <sup>c</sup>	1.089-1.434	3.3511 <sup>c,d</sup>	2.795-3.828
Vitros	47	25	0.5876 <sup>b</sup>	0.247-0.794	3.7776 <sup>d</sup>	3.22-4.263

<sup>a</sup> Groups with the same letter next to the parameter estimate do not differ significantly for  $\alpha$  or  $\beta$ .

creasing ethanol concentration. For the parameter  $\alpha$ , no clear distinction was observed between the groups. The parameter was highest for the Roche method, although significantly different only from Vitros and Abbott.

When the information of  $\alpha$  and  $\beta$  was combined and visually represented (Fig. 3), it was clear that the Abbott method had overall the lowest variability. In addition, the relation between SD and ethanol concentration was quite particular for the Vitros method. Although it was significantly different only from Roche, it had the second-lowest intercept of all methods, and its slope was the highest of all, although significantly higher than only the slopes of Roche and Abbott. The combination of a low  $\alpha$  and a high  $\beta$  resulted in its curve going from the second-lowest variability at low concentration to the highest variability at high concentration.

The results for CV are shown in Table 2 and Figs. 2 and 4. The parameter  $\alpha$ , which determines the behavior of the line for the lower concentrations, was lowest for Abbott and Vitros and highest for Roche and Siemens. The parameter  $\beta$ , which determines rather the position of the approximately horizontal part of the characteristic curve for the higher concentrations, was highest for Vitros. It was significantly higher than the  $\beta$  of Abbott, Roche, and Siemens. Headspace chromatography had an intermediate value.

## Discussion

This study demonstrates that the characteristic function, introduced earlier in the EQAs for soil and food chemistry, can be successfully applied to clinical EQAs. It should be noted, however, that some precautions must be taken before applying the characteristic function. First of all, its applicability always depends on its ability to describe the relation between the EQA standard deviation and the target concentration. The performance of the characteristic function must always be visually inspected with a residual plot. A nonparametric regression line may be used to indicate the direction

of the point cloud that consists of the residual values: the flatter the nonparametric regression line is, the better the characteristic function fits the data. One may consider accepting the goodness of fit if the difference between the highest and lowest point of the nonparametric regression line is less than 2 times the rounding error of the reported data, since in this case the deviation caused by a lack of fit is smaller than or equal to the error induced by rounding. Second, it has been suggested that only samples from the same origin or matrix can be used to establish a characteristic function (14).

The characteristic function's parameters are relatively easy to parameterize, and they deliver ample information about the reproducibility of analytical methods for any concentration in the investigated range, even for concentrations that have not been investigated before. When historical data are available, the estimated SD may be used in several aspects of the EQA process. For homogeneity testing, the interval variability should be compared with the SD that will be used to evaluate laboratories. If laboratories are evaluated on the basis of the SD of the reported results, the EQA organizer faces the problem of evaluating sample homogeneity with an unknown SD. The estimated SD of the characteristic function may act as a helpful estimate of the SD to be expected from the future reported EQA results. Subsequently, the estimated SD may be used to assess laboratories' performance on the basis of  $z$ -scores and may be preferred when the SD of the reported results is not reliable, for example, when there are not enough data available (21) or when results were reported near the limit of quantification.

Koch and Magnusson (10) already mentioned how to compare methods with the characteristic function. It may also be interesting to assess the statistical significance of the differences that are found, for example, to find the method that has the significantly lowest variability. Here, a small extension to the characteristic function helped to confirm significant differences between the examined methods for the low concentrations. For example, for increasing target

concentrations up to 2.5 g/L, the Roche and Abbott methods exhibited the smallest increase of variability.

Finally, a note should be made about the use of the CV to report data. In fact, Eq. (2) shows that the CV is equal to  $\beta$  for a model where  $\alpha$  is forced to be zero. It should be noted that, at least for serum ethanol testing, the  $\alpha$  parameter is not equal to zero: none of the confidence intervals in Table 2 contains zero. In these cases, and as can be seen in Fig. 4, constant CVs are a very rough approach to the real behavior of the variability and should not be considered independent of the concentration except when the concentration is large enough. Of course, it would be better to interpret  $\alpha$  and  $\beta$

rather than a single parameter that is not constant over the whole concentration range.

**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:** *No authors declared any potential conflicts of interest.*

**Role of Sponsor:** No sponsor was declared.

## References

1. Thompson M. The characteristic function, a method-specific alternative to the Horwitz function. *J AOAC Int* 2012;95:1803-6.
2. International Organization for Standardization. ISO 17043. Conformity assessment: general requirements for proficiency testing. Geneva: ISO; 2010.
3. Horwitz W, Britton P, Chirtel SJ. A simple method for evaluating data from an interlaboratory study. *J AOAC Int* 1998;81:1257-65.
4. International Organization for Standardization. ISO 13528:2005. Statistical methods for use in proficiency testing by interlaboratory comparisons. Geneva: ISO; 2005.
5. Whitehead T, Woodford F. External quality assessment of clinical laboratories in the United Kingdom. *J Clin Pathol* 1981;34:947-57.
6. Ehrmeyer S, Laessig R. Alternative statistical approach to evaluating interlaboratory performance. *Clin Chem* 1985;31:106-8.
7. Meijer P, de Maat MPM, Kluff C, Haverkate F, van Houwelingen HC. Long-term analytical performance of hemostasis field methods as assessed by evaluation of the results of an external quality assessment program for antithrombin. *Clin Chem* 2002;48:1011-5.
8. Coucke W, Van Blerk M, Libeer JC, Van Campenhout C, Albert A. A new statistical method for evaluating long-term analytical performance of laboratories applied to an external quality assessment scheme for flow cytometry. *Clin Chem Lab Med* 2010;48:645-50.
9. Kallner A, Khorovskaya L, Pettersson T. A method to estimate the uncertainty of measurements in a conglomerate of instruments/laboratories. *Scand J Clin Lab Invest* 2005;65:551-8.
10. Koch M, Magnusson B. Use of characteristic functions derived from proficiency testing data to evaluate measurement uncertainties. *Accreditation Qual Assur* 2012;17:399-403.
11. Duewer DL, Brown Thomas J, Kline MC, MacCrehan WA, Schaffer R, Sharpless KE, et al. NIST/NCI Micronutrients Measurement Quality Assurance Program: measurement repeatabilities and reproducibilities for fat-soluble vitamin-related compounds in human sera. *Anal Chem* 1997;69:1406-13.
12. Thompson M, Wood R. Using uncertainty functions to predict and specify the performance of analytical methods. *Accreditation Qual Assur* 2006;10:471-8.
13. Rozet E, Marini RD, Ziemons E, Hubert P, Dewé W, Rudaz S, et al. Total error and uncertainty: friends or foes? *TrAC Trends Anal Chem* 2011;30:797-806.
14. Thompson M, Mathieson K, Damant AP, Wood R. A general model for interlaboratory precision accounts for statistics from proficiency testing in food analysis. *Accreditation Qual Assur* 2008;13:223-30.
15. Jiménez-Chacón J, Alvarez-Prieto M. Modelling uncertainty in a concentration range. *Accreditation Qual Assur* 2009;14:15-27.
16. Côté I, Robouch P, Robouch B, Bisson D, Gamache P, LeBlanc A, et al. Determination of the standard deviation for proficiency assessment from past participant's performances. *Accreditation Qual Assur* 2012;17:389-93.
17. Rousseeuw PJ, Croux C. Alternatives to the median absolute deviation. *J Am Stat Assoc* 1993;88:1273-83.
18. Wilrich P-T. Robust estimates of the theoretical standard deviation to be used in interlaboratory precision experiments. *Accreditation Qual Assur* 2007;12:231-40.
19. Spang HA. A review of minimization techniques for nonlinear functions. *SIAM Rev* 1962;4:343-65.
20. Cleveland WS, Devlin SJ. Locally weighted regression: an approach to regression analysis by local fitting. *J Am Stat Assoc* 1988;83:596-610.
21. Coucke W, China B, Delattre I, Lenga Y, Van Blerk M, Van Campenhout C, et al. Comparison of different approaches to evaluate External Quality Assessment data. *Clin Chim Acta* 2012;413:582-6.