Ellis T. Aune, Laura E. Diepeveen, Coby M. Laarakkers, Siem Klaver, Andrew E. Armitage, Sukhvinder Bansal, Michael Chen, Marianne Fillet, Huiling Han, Matthias Herkert, Outi Itkonen, Daan van de Kerkhof, Aleksandra Krygier, Thibaud Lefebvre, Peter Neyer, Markus Rieke, Naohisa Tomosugi, Cas W. Weykamp and Dorine W. Swinkels*

Optimizing hepcidin measurement with a proficiency test framework and standardization improvement

Abstract

Objectives: Hepcidin measurement advances insights in pathophysiology, diagnosis, and treatment of iron disorders, but requires analytically sound and standardized measurement procedures (MPs). Recent development of a two-level secondary reference material (sRM) for hepcidin assays allows worldwide standardization. However, no proficiency testing (PT) schemes to ensure external quality assurance (EQA) exist and the absence of a high calibrator in the sRM set precludes optimal standardization.

Methods: We developed a pilot PT together with the Dutch EQA organization Stichting Kwaliteitsbewaking Medische Laboratoriumdiagnostiek (SKML) that included 16 international hepcidin MPs. The design included 12 human serum samples that allowed us to evaluate accuracy, linearity, precision and standardization potential. We manufactured, value-assigned, and validated a high-level calibrator in a similar manner to the existing low- and middle-level sRM.

Results: The pilot PT confirmed logistical feasibility of an annual scheme. Most MPs demonstrated linearity ($R^2 > 0.99$) and precision (duplicate CV > 12.2%), although the need for EQA was shown by large variability in accuracy. The high-level calibrator proved effective, reducing the inter-assay CV from 42.0% (unstandardized) to 14.0%, compared to 17.6% with the two-leveled set. The calibrator passed international homogeneity criteria and was assigned a value of 9.07 ± 0.24 nmol/L.

Conclusions: We established a framework for future PT to enable laboratory accreditation, which is essential to

Ellis Aune and Laura E. Diepeveen contributed equally to this work.

*Corresponding author: Prof. Dr. Dorine W. Swinkels, Department of Laboratory Medicine, Translational Metabolic Laboratory (830), Radboud University Medical Center, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands; and Hepcidinanalysis.com, Nijmegen, The Netherlands, Phone: +31 (0)24-3618957, Fax: +31 (0)24-3668754, E-mail: Dorine.Swinkels@Radboudumc.nl
Ellis T. Aune, Laura E. Diepeveen, Coby M. Laarakkers and Siem Klaver, Department of Laboratory Medicine, Radboud University Medical Center, Nijmegen, The Netherlands; and Hepcidinanalysis.com, Nijmegen, The Netherlands
Andrew E. Armitage, MRC Human Immunology Unit, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK
Sukhvinder Bansal, Department of Pharmacy, School of Cancer and Pharmaceutical Science, King's College London, London, UK
Michael Chen, Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, Canada; and Division of Medical Sciences, University of Victoria, Victoria, Canada
Marianne Fillet, Laboratory for the Analysis of Medicines, CIRM, University of Liège, Liège, Belgium

Huiling Han, Intrinsic Life Sciences, La Jolla, USA
Matthias Herkert, DRG Instruments, Marburg, Germany
Outi Itkonen, Laboratory Division HUSLAB, Helsinki University Central Hospital, Helsinki, Finland
Daan van de Kerkhof, Algemeen Klinisch Laboratorium, Catharina Ziekenhuis, Eindhoven, The Netherlands
Aleksandra Krygier, Department of Endocrinology, Metabolism and Internal Medicine, Poznan University of Medical Sciences, Poznan, Poland
Thibaud Lefebvre, French Center of Porphyria, INSERM UMR1149, Labex GR-Ex, Louis Mourier Hospital, APHP.Nord-Université de Paris, Paris, France
Peter Neyer, Institute of Laboratory Medicine, Kantonsspital Aarau, Aarau, Switzerland
Markus Rieke, IPH GMBH, Alfeld (Leine), Germany
Naohisa Tomosugi, Division of Systems Bioscience for Drug Discovery, Medical Research Institute, Kanazawa Medical University, Kahoku, Japan
Cas W. Weykamp, Department of Clinical Chemistry, Queen Beatrix Hospital, Winterswijk, The Netherlands; and and SKML, Nijmegen, The Netherlands

ensure quality of hepcidin measurement and its use in patient care. Additionally, we showed optimized standardization is possible by extending the current sRM with a third high calibrator, although international implementation of the sRM is a prerequisite for its success.

Keywords: external quality assurance; hepcidin; iron metabolism; proficiency testing; standardization; secondary reference material.

Introduction

The liver-derived hormone hepcidin is the key regulator of iron homeostasis by inhibiting the only known cellular iron exporter ferroportin [1, 2]. Since dysregulation of hepcidin causes a variety of iron disorders, including anemia of inflammation, its measurement and its ratio to ferritin and transferrin saturation can be used to diagnose certain iron disorders and guide iron therapies, making it an important diagnostic biomarker [1, 3, 4]. Furthermore, hepcidin is a therapeutic target for both iron-overload disorders, such as β-thalassemia and hereditary hemochromatosis, and iron-restrictive anemias as observed with iron refractory iron deficiency anemia (IRIDA), inflammatory diseases, certain tumors and chronic kidney disease [5–7].

Both mass spectrometry (MS) and immunochemistry (IC) based measurement procedures (MPs) have been developed to quantify hepcidin concentrations. However, our previous studies revealed that hepcidin levels in the same clinical sample may vary up to a factor of 9 among different MPs [8–11]. This lack of worldwide standardization causes confusion in interpretation of hepcidin levels and hepcidin-related ratios, which hampers both research collaborations and multicenter medical consultations [12]. Effective use of hepcidin measurement in patient care and clinical research require both comparability and analytical reliability to establish uniform clinical decision limits and reference ranges [13]. This is essential to compare results across studies or monitor a patient's treatment at different facilities to prevent inconsistent or incorrect conclusions.

As a first step, we developed a two-leveled (low and middle) commutable secondary reference material (sRM) made of human serum that was value-assigned by a primary reference material (pRM) [11]. We showed that calibration using this sRM reduced the inter-method coefficient of variation (CV) from 42.1 to 11.0% when standardization was simulated and from 52.8 to 19.1% when standardization was performed in practice. The sRM, with concentrations of 0.95 ± 0.11 nmol/L and 3.75 ± 0.17 nmol/L (k=1),

increases comparability between MPs but calibrates solely the lower part of the pathophysiological hepcidin range. Therefore, in this current study, we produced and validated a third high-level calibrator to cover the higher hepcidin levels. Global implementation of the sRM allows standardization of all hepcidin MPs, meaning measurements can be traced back to the Système International (SI) and a "true" value can be established [14, 15]. As a next step, to evaluate the analytical performance of hepcidin assays and ensure reliability of hepcidin MPs, we aimed to create the first external quality assurance (EQA) program for hepcidin assays to pave the way for laboratory accreditation.

Here, we report the results of a pilot proficiency test (PT) organized and implemented in collaboration with Dutch external quality assurance (EQA) organizer Stichting Kwaliteitsbewaking Medische Laboratoriumdiagnostiek (SKML) [16]. The aims of this proficiency initiative were to set-up a framework for a worldwide EQA program for hepcidin assays, in which the analytical performance and current agreement among international hepcidin MPs was determined, and to evaluate the calibration potential of the three-leveled sRM.

Materials and methods

Study overview

The aims of our study were two-fold. First, we wanted to evaluate the current analytical performance and agreement of hepcidin MPs worldwide and determine if standardization has already been achieved regarding recent production of a sRM. To this end, we established the framework for an EQA scheme in order to provide participating laboratories with a summary of their analytical performance to allow opportunities for accreditation and ultimately improve the standard of diagnostics and patient care internationally.

Second, we produced a high-level calibrator in the same manner as those already developed [11] and aimed to validate its potential to improve standardization compared to the two-leveled sRM using retrospective calibration of the PT samples.

To this end, in collaboration with SKML (Supplementary Figure 1), we developed a PT that included a variety of international hepcidin MPs. We produced a set of 12 lyophilized human serum samples with target values determined by a candidate reference measurement procedure (cRMP, Supplementary Table 1), designed to address accuracy, linearity, precision and standardization potential. These samples included the existing two calibrators [17], the newly produced third candidate calibrator, a linearity panel with three blinded duplicates and three additional samples. These additional samples were selected to cover the upper end of the (patho) physiological range, which was not included in the linearity panel, to ensure good coverage of the whole clinically relevant range as such make the sample set robust for the purposes of a thorough pilot PT scheme.

Proficiency test program development

Laboratory recruitment and participation: Laboratories housing hepcidin MPs were invited to participate based on previous collaborations [10, 11], expressed interest in purchasing the sRM, or published on hepcidin as a diagnostic biomarker in 2018 and 2019 in peer reviewed journals.

The initial group included 15 laboratories running 19 MPs (10 MS and 9 IC) from 12 countries and 3 continents. All were asked to run the samples within two weeks of receipt and to perform their assays in the same manner as they would for their routine use. IC-2 experienced calibrator errors resulting in unreliable data and IC-4 encountered significant equipment errors that prevented them from running their assay and reporting results. MS-3 did not consent to de-anonymization, excluding their results. Therefore, the final cohort included 16 MPs (9 MS and 7 IC, Table 1).

Data collection: All labs were provided with both a digital and hard copy of a Standard Data Report Form (Supplementary Figure 2) that included questions about the measurement method, a table to report results in the units they were measured, and space for remarks. Laboratories were asked to return the completed form within two weeks of receiving the samples.

Samples

Collection and preparation: To produce the linearity panel of three duplicates, three additional samples within the physiological hepcidin range [26] and a high-level calibrator, we periodically collected and processed anonymized leftover serum from routine diagnostics and therapeutic phlebotomies in December 2019 and January 2020. Details are described in the Supplementary Methods.

Distribution: All lyophilized sample sets were shipped at room temperature (RT) on the same day from Streekziekenhuis Koningin Beatrix in Winterswijk, The Netherlands. All were instructed to store the samples at 4 °C upon arrival until the assay was performed and information about sample storage and handling was provided both digitally over email and in hard copy with shipment.

Ethics

This study was conducted in accordance with the Declaration of Helsinki. All leftover patient serum was anonymized upon collection and was handled in accordance with the code for proper secondary use of human tissue in The Netherlands.

Table 1: Methodological approaches of participating hepcidin MPs.

Participant code	MP	Extraction	Standard	Standard manufacturer	Ref
MS-1	MALDI-TOF MS	WCX	Heavy-isotope labeled synthetic hepcidin-25	Peptides International	[18]
MS-2	LC-MS/MS	Precipitation with ACN	Heavy-isotope labeled synthetic hepcidin-25	Peptides International	[19]
MS-4	LC-MS/MS	Reversed phase	Hepcidin/LEAP-1 from Peptides Int.	Peptides International	a
MS-5	LC-MS/MS	Mixed anion exchange	Heavy-isotope labeled synthetic hepcidin-25	Peptide Institute	[20]
MS-6	LC-MS/MS	HBL/Waters	Heavy-isotope labeled synthetic hepcidin-25	Peptide Institute	[21]
MS-7	LC-MS/MS	SPE	Heavy-isotope labeled synthetic hepcidin-25	Peptide Institute	b
MS-8	LC-MS/MS	SPE	Heavy-isotope labeled synthetic hepcidin-25	Synthesized in house	[22]
MS-9	UPLC-MS/MS	Reversed phase SPE	Heavy-isotope labeled synthetic hepcidin-25	Synthesized in house	[23]
MS-10	LC-MS/MS	Reversed phase SPE	Heavy-isotope labeled synthetic hepcidin-25	Peptide Institute Inc	[24]
IC-1	cELISA	None	Synthetic hepcidin-25	Bachem	c
IC-3	cELISA	None	Synthetic hepcidin-25	Bachem	d
IC-5	cELISA	None	Synthetic hepcidin-25	Bachem	d
IC-6	cELISA	None	Synthetic hepcidin-25	Peptides International/AmbioPharm	e
IC-7	cELISA	None	Synthetic hepcidin-25	Peptides International/AmbioPharm	f
IC-8	cELISA	None	Synthetic Hepcidin-25	Bachem	c
IC-9	cELISA	None	Synthetic Hepcidin-25	Bachem	[25]

MP, measurement procedure; MS, mass spectrometry-based MP; IC, immunochemical-based MP; MALDI, matrix-assisted laser desorption/ionization; TOF, time of flight; LC, liquid chromatography; UPLC, ultra-performance liquid chromatography; SPE, solid phase extraction; cELISA, competitive enzyme-linked immunosorbent assay; WCX, weak-cation exchange; HLB, hydrophilic lipophilic balanced reverse phase; ACN, acetonitrile. [a]No reference available, an MS method based on the assay described in Schmitz et al. [24], developed by the Institute of Laboratory Medicine, Kantonsspital Aarau, Aarau, Switzerland. [b]Manuscript under preparation; a laboratory-developed MP for hepcidin-25 by Laboratory for the Analysis of Medicines, University of Liège, Liège, Belgium. [c]Hepcidin 25 bioactive automated ELISA (Cat. #HYE-5769) from DRG Diagnostics, Marburg, Germany. [d]Hepcidin 25 bioactive HS Kit (Cat. #EIA-5782) commercially available assay from DRG Diagnostics, Marburg, Germany. [e]No reference available; Intrinsic Hepcidin IDxTM Kit (product #ICE-007) – commercially available competitive ELISA from Intrinsic LifeSciences, La Jolla, USA. [f]No reference available; Intrinsic Hepcidin IDxTM Test – automated competitive ELISA from Intrinsic LifeSciences, La Jolla, USA.

Data analysis

Proficiency test: Results reported in ng/mL were converted to nmol/L, using the molecular weight of hepcidin-25 (2789.4 g/mol) [27]. The values determined by MS-1 were used as target values for evaluating the proficiency of all laboratories, as MS-1 was previously described as cRMP that is calibrated using the reference material [11]. For the purposes of the pilot, potential outliers were not removed in order to avoid biasing the data.

Equivalence between MPs was assessed in terms of accuracy of each MP, a ratio of each laboratory-assigned value to the target value converted to percentage, and bias (nmol/L) of each result compared to the cRMP, calculated by subtracting the values obtained by each laboratory for each sample from the target value determined by MS-1. Additionally, the intra-assay CVs for each sample (n=9, excluding the three calibrators) were calculated among all laboratories (n=18) as well as within each method group (IC or MS). The resulting CVs were then averaged and quantified as the mean inter-assay CV (%).

Analytical performance was assessed in terms of linearity and precision. For evaluation of linearity, the duplicate linearity samples were averaged and linear regression was performed to find an R^2 value. Precision was evaluated by determining the CV for each of the three duplicate samples. To evaluate adequacy of precision for hepcidin measurements, optimal precision was calculated as $f1*CV_i$ [28], where CV_i is the intra-individual CV (48.8%) [29] and $f1$ is 0.25 for an optimal threshold.

Calibration: Commutability of the low and middle calibrators was assessed previously with regression analysis of 16 native serum samples for all 9 MPs (y-axis) against the mean of all MPs (x-axis) [11]. As the mean results of both calibrators fell within the 95% prediction interval of the regression line, commutability was confirmed. Since the third high calibrator was produced in the same manner as the previously developed calibrators, commutability was assumed here.

All laboratories received the samples blinded, therefore the effect of standardization by using the sRM was performed retrospectively by value reassignment based on linear regression of the results of the sRM samples per MP against the respective results of the cRMP MS-1. The inter-assay dispersion in these simulated results was then expressed as the inter-assay CV (%) after standardization with the sRM and compared with the inter-assay CV (%) before standardization. It is important to note that good analytical performance is a prerequisite to evaluating standardization potential.

Hepcidin exhibits relatively high biological variation, i.e. a between-day intra-individual variation of 48.8% and an inter-individual variation of 154.1% [29]. Therefore, to place the bias of all hepcidin measurement compared to MS-1 in a relevant diagnostic context, total allowable error (TEa) was calculated using

$$TE_A, \% = 1.65*f1*CV_i + f2\left(CV_i^2 + CV_g^2\right)^{\frac{1}{2}}$$

[27]. CV_i is the between-day intra-individual CV (48.8%), CV_g is the inter-individual variation (154.1%) [29], and $f1$ and $f2$ are factors for optimum (0.25 and 0.125), desirable (0.5 and 0.25) and minimum (0.75 and 0.327) TEa.

Characterization of the third high calibrator

Homogeneity was evaluated according to ISO13528 by means of duplicate measurements of 12 randomly selected calibrator samples by MS-1 [11, 30]. The sRM was reconstituted with 0.30 mL deionized water and left at RT for 15 min, followed by careful mixing for 20 min (roller bench, 3.5 rpm). We compared within-vial to between-vial variation to assess if the calibrator passes homogeneity criteria.

Stability was evaluated by storing aliquots of the sRM at 4 °C. Measurements were performed by MS-1 at baseline and after 1 and 6 months. These will be continued at 12 and 18 months, and then annually for five years. Concentration changes are considered significant, and indicative of instability, if they exceed the precision of MS-1. Statistical analysis was done using analysis of variance (ANOVA) and Bonferroni's multiple comparison test.

The true value of the high calibrator was assigned using the cRMP, a validated Weak-Cation-eXchange MALDI-Time of Flight-MS (MS-1) [11]. We used the pRM to reassign the internal standard of MS-1 (stable isotope, manufactured by Peptide International) and subsequently used this internal standard to assign a value to the candidate high-level calibrator, as described previously [11].

Results

Organizational aspects of proficiency testing

A primary goal of the pilot PT was to assess the feasibility of sample preparation and send-out. No significant problems were encountered in this process. Anonymous sample collection from diagnostic leftovers and therapeutic phlebotomies was efficient, and the process of developing PT samples of particular concentrations based on initial hepcidin measurements was successful. All samples were delivered to laboratories within three days of shipment from The Netherlands and all laboratories reported that samples arrived without any visible damage.

Measurement by the laboratories was generally uncomplicated, though six MPs (from five laboratories) reported after the two-week deadline but still within four weeks of receiving the samples. Laboratories reported late due to equipment malfunction, scheduling conflicts, or commercial ELISA shipping delays. No laboratories reported errors with sample reconstitution and handling. All laboratories correctly and completely filled out the standard data report form.

Laboratory proficiency

Data analysis of the uncalibrated results showed a high level of variation among the absolute hepcidin values of

the methods evaluated (Supplementary Table 2), confirming the need for standardization. Analytical performance of each MP is summarized in Table 2. For IC methods, the value for HPT2020-S9 (21.18 nmol/L, Supplementary Table 1) was reported as out of range for three MPs. For the purpose of data analysis, these values were excluded for those assays.

Accuracy and bias

On average, the accuracy was 145% and ranged from 76 to 540% (Table 2), again stressing the current lack of standardization. IC methods reported higher results on average (Supplementary Table 3). The bias of each measurement from the target values determined by cRMP MS-1 without standardization is shown in Figure 1A. By placing these results in the context of the TEa, we assessed if the inter-assays CVs are adequate for the biological variation of hepcidin, as described in Diepeveen et al. [11]. Based on reported inter- and intra-individual CVs for hepcidin, TEa of 40.3% (optimum), 80.7% (desirable), and 121.0% (minimum) were calculated and subsequently plotted. Many results fall outside of the optimum range and although most fall within the minimum ranges, one MP did not meet the minimum TEa criteria.

Linearity

In general, laboratories showed good analytical performance in terms of linearity, with a linear regression R^2 average of 0.9959 (range: 0.9704–1, Table 2). These results suggest that the linearity of the assays is acceptable, at least up to a concentration of 12.2 nmol/L (highest linearity sample). While for most laboratories R^2 values above 0.99 were found, MS-5 reported data with a lower R^2 value (0.9704).

Precision

Analytical performance assessed in terms of precision was, on average, less than the calculated optimal minimum CV of 12.2% for most MPs (Table 2). The exception is MS-5. Three additional assays reported at least one of the three duplicates with a CV>12.2% (MS-7, MS-8, IC-6).

Characteristics of the high-level calibrator

Calibration potential

The third high calibrator, made of lyophilized serum with CLP, was validated during the proficiency test solely with

Table 2: Summary of assay performance before calibration with the reference material.

MP	Accuracy [Range], %	R^2	Precision, %CV			
			Duplicate 1	Duplicate 2	Duplicate 3	Average
MS-1	100 [N/A]	0.9995	1.8	2.3	2.6	2.2
MS-2	112 [95–151]	0.9985	10.7	4.3	2.7	5.9
MS-4	118 [101–150]	1	3.0	1.0	0.0	1.3
MS-5	100 [55–172]	0.9704	29.7	6.7	35.8	24.1
MS-6	114 [97–139]	0.9985	1.0	1.7	2.9	1.8
MS-7	173 [147–255]	0.9959	20.0	4.1	1.5	8.5
MS-8	131 [114–154]	0.9941	2.6	16.4	6.1	8.4
MS-9	89 [76–108]	0.9991	1.7	1.3	6.4	3.1
MS-10	170 [151–209]	1	0.1	1.5	0.6	0.7
IC-1	125 [109–147]	1	0.3	1.4	4.3	2.0
IC-3	116 [94–125]	0.9992	2.1	8.4	3.7	4.7
IC-5	132 [118–141]	0.9941	5.7	1.2	6.2	4.4
IC-6	340 [274–540]	0.9948	15.0	8.8	0.9	8.2
IC-7	205 [174–252]	0.9991	0.6	8.9	2.5	4.0
IC-8	127 [118–154]	0.9993	3.3	3.9	3.0	3.4
IC-9	168 [142–199]	0.9923	8.2	2.3	2.2	4.3
Average	145 [76–540]	0.9959	6.6	4.6	5.1	5.4

MP, measurement procedure; MS, mass spectrometry-based MP; IC, immunochemical-based MP; N/A, not applicable. Accuracy (%) is expressed as a ratio of each laboratory-assigned value to the target value as determined by MS-1. R^2 was calculated using linear regression of the three linearity panel samples. Precision, expressed as percent coefficient of variation, was calculated using the blinded duplicates of the three linearity samples. R^2 and precision are indicators of analytical assay performance, whereas accuracy is indicative of the need for standardization. Values above the optimal precision threshold (12.2% CV) are depicted in blue.
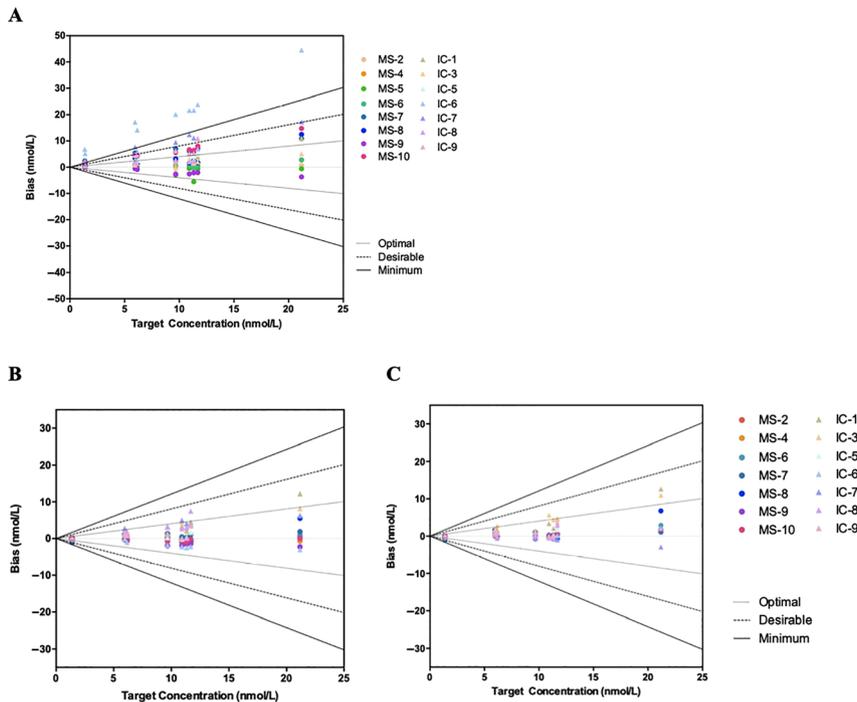
Figure 1: Bias from target values before calibration (A), after calibration with the two-leveled sRM (B), and after calibration with the three-leveled sRM (C).
Bias (nmol/L, y-axis) was calculated by subtracting the target value (nmol/L, x-axis), as determined by MS-1, from the reported value for each sample (n=9) from each measurement procedure. Calibration with the sRM was done using a linear regression with the calibration samples (either S2 and S7 or S2, S7, and S12) to recalculate the reported values. For this evaluation of calibration potential, MS-3 and MS-5 were excluded based on poor analytical performance. Optimal, desirable, and minimum TEa lines were defined as 40.3, 80.7, and 121.0% respectively based on reported inter- and intra-individual CVs for hepcidin [28, 29].

MPs that met our criteria of good analytical performance assessed in terms of linearity and precision. To this end, MS-5 was not included in this evaluation of the calibration potential.

Without standardization, the overall inter-assay CV was found to be 42.0% (Table 3). Looking at MS and IC methods separately, we found an inter-assay CV of 25.3% for MS MPs and an inter-assay CV of 45.9% for IC MPs. As expected, mathematical simulation of standardization with the two-leveled sRM showed a great reduction of the inter-assay CV (overall; 17.6%, MS; 11.0%, IC; 17.2%, Table 3). Mathematical simulation of standardization using the three-leveled sRM, including our newly produced third high calibrator, shows an even better improvement in the inter-assay CV (overall; 14.0%, MS; 8.8%, IC; 15.7%, Table 3), achieved in large part by improving equivalency at higher concentrations. Additionally, the average accuracy of all of the MPs was found to be improved from 145% unstandardized to 106.4% with the two-level calibrator and 105.8% with the three-level calibrator (Table 3). When visualizing bias, the spread is clearly reduced using the two-leveled calibrator (Figure 1B) compared to the non-calibrated data (Figure 1A). However, in particular the IC methods still tend to show higher variability both above and below the target values. With the use of the three-leveled calibrator (Figure 1C), nearly all results fall within the minimum bias allowance and most meet the desirable bias allowance for both MS and IC methods. It is important to note that even though MS-5 did not meet the analytical

performance criteria to be included in this standardization evaluation, when retrospectively calibrated, its results still fall within the desirable bias range (Supplementary Figure 3).

Homogeneity, stability and value assignment

The calibrator passed homogeneity criteria as described by ISO13528 [30], as the between-vial variation (SD: 0.236 nM) was smaller than the within-vial variation (SD: 0.322 nM).

The material was found to be stable for up to 6 months (stability testing ongoing), although stability up to 5 years is assumed since this is confirmed for lyophilized material with CLP in previous studies [10, 11]. Its value was assigned using the pRM and MS-1, as the candidate RMP, and is defined as 9.07 ± 0.24 nM (k=1).

Discussion

Multiple studies have shown that absolute hepcidin levels reported for the same clinical sample vary tremendously depending on the MP used, which complicates utility of the biomarker [8–11]. As a first step towards uniform hepcidin measurement, a two-leveled commutable sRM was produced, enabling worldwide standardization [11]. To optimize this, we now have established a framework for future quality assurance and extended the sRM by adding a third high calibrator.

Table 3: Impact of two-level and three-level calibration on inter-assay CV and accuracy.

	Overall			MS			IC		
	Pre	**2-L**	**3-L**	**Pre**	**2-L**	**3-L**	**Pre**	**2-L**	**3-L**
Mean inter-assay CV, %	42.0	17.6	14.0	25.3	11.0	8.8	45.9	17.2	15.7
Mean accuracy, %	151.5	106.4	105.8	129.6	97.0	100.3	173.4	115.7	111.2

MS, mass spectrometry-based MP; IC, immunochemical-based MP; CV, coefficient of variation. Inter-assay CV (%) and accuracy (%) before calibration (Pre), calibration with the low- and middle-level calibrators (2-L) and all three calibrators (3-L) were evaluated for all methods and MS/IC separately.

Here, we showed that PT is feasible and most MPs perform well on linearity and precision, which is a prerequisite for standardization and ensures reliable hepcidin measurement. However, the average accuracy of all MPs was found to be 145%, which stresses the clear need for EQA and reveals that even though an sRM is available, standardization has not yet been achieved. Furthermore, our previous research suggested that calibration bias was the major contributor to measurement inaccuracy [11], which we tried to further reduce with expanding the sRM with a high calibrator extending the calibration potential to the upper hepcidin range. Although its assumed commutability will ideally be verified in a larger PT study, we validated its potential to reduce the inter-assay CV with retrospective standardization of the laboratory data using concentrations the laboratories obtained for the calibrators included in the PT set. The three-level calibrator reduced the inter-assay CV even more than the two-leveled calibrator (overall 2-L: 17.6%; 3-L: 14.0%) compared to nonstandardized data (overall: 42.0%). Furthermore, MS-5 did not meet our criteria of acceptable precision, which afterward appeared due to internal standard inconsistencies that had gone undetected in standard practice, emphasizing the need for, and utility of, EQA. However, MS-5 results still fall within the desirable TEa when standardized, elucidating that even when optimal analytical performance is not achieved the sRM is still valuable in reducing calibration bias. When translated to patient care, these results cumulatively suggest that instituting EQA can ensure reliable, standardized hepcidin measurements. This will facilitate, for example, international communication among medical doctors regarding diagnosis of rare hepcidin-related iron disorders such as IRIDA and comparison of hepcidin-related research studies, making study outcomes more meaningful in clinical practice.

Besides decreasing calibration bias and improving the analytical performance of MPs, optimization of hepcidin standardization, and therefore utility of PT, can be further improved by reducing the heterogeneity of the measurand. A first step was made by studying the degree of hepcidin

protein binding in the circulation [31], which was inconclusive. Further research is needed to understand if this might influence hepcidin quantification, which in turn is crucial for correct interpretation of its measurement in patients. Additionally, differences in MS and IC performance can be due to measurand heterogeneity, since we observe higher variation and less accuracy in IC compared to MS methods, which is important to clarify. Although this difference has been documented for more biomarkers [32], IC MPs are certainly valuable in research and diagnostics, especially where MS systems are not accessible and less accuracy may be allowed practically due to a high biological variation and therefore TEa. For hepcidin MPs, these observed differences between IC and MS MPs may be due to cross-reactivity of hepcidin isoform detection by IC methods, which is problematic since hepcidin-25 is the only biologically active isoform and the one that should be evaluated. [8, 10] Currently, there is inconclusive data regarding the influence of isoform detection on hepcidin-25 quantification, which must be studied further to assess if it affects clinical decision making [33, 34]. Furthermore, several IC methods also reported the sample with the highest target concentration (S9) to be out of range instead of providing a value, which may influence IC data. This suggests that these assays have more difficulty to measure hepcidin levels in the upper reference range and elucidates the need for a standardized protocol for handling out-of-range measurements. All in all, future efforts will be directed towards achieving a consensus on best practice for clinical hepcidin measurement.

Last, larger studies into the between- and within-subject variation of hepcidin would allow optimal assessment of the achievements of global standardization and validity of PT, since these parameters are used to place the achieved inter-assay CV after standardization within a biological context. The higher the biological variation, the higher the allowable bias after standardization. Currently, the TEa was based on relatively limited intra- and inter-individual variation data [28], which, though similar to

other studies [35–38], is not guaranteed to provide the most accurate estimate.

Altogether, this pilot program was designed to assess the current performance of MPs and lays important groundwork for an annual PT scheme. Based on the minor logistical challenges we encountered, we will extend the notification, shipment and data reporting timelines in the future enabling more laboratories to participate. Also, a scoring system for standardized laboratory evaluation will be included and a formal report will be generated in accordance with other SKML schemes. This EQA program will ultimately pave the way for international laboratory accreditation, remediation of analytically poorly performing MPs through comprehensive performance feedback, and universal definition of reference ranges and clinical decision limits. All will directly contribute to enhanced quality of hepcidin results and hepcidin-related ratios in both research and diagnostics, and consequently also in quality of publications and increased utility of hepcidin measurement in patient care. Here, we demonstrate the potential for achieving worldwide standardization, ensured by PT, although international implementation the three-leveled sRM is a prerequisite for the success of such a program. The material is available at HepcidinAnalysis.com.

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

Competing interests: EA, LD, CL, SK and DS are employees of Radboudumc, which via its Hepcidinanalysis.com initiative offers high quality hepcidin measurements to the medical, scientific and pharmaceutical community at a fee for the service basis. All other authors state no conflict of interest.

Ethical approval: This study was conducted in accordance with the Declaration of Helsinki. All leftover patient serum was anonymized upon collection and was handled in accordance with the code for proper secondary use of human tissue in The Netherlands.

References

1. Kroot JJ, Tjalsma H, Fleming RE, Swinkels DW. Hepcidin in human iron disorders: diagnostic implications. Clin Chem 2011;57: 1650–69.
2. Nemeth E, Tuttle MS, Powelson J, Vaughn MB, Donovan A, Ward DM, et al. Hepcidin regulates cellular iron efflux by binding to ferroportin and inducing its internalization. Science 2004;306: 2090–3.
3. Girelli D, Nemeth E, Swinkels DW. Hepcidin in the diagnosis of iron disorders. Blood 2016;127:2809–13.
4. Nemeth E, Ganz T. The role of hepcidin in iron metabolism. Acta Haematol 2009;122:78–86.
5. van Swelm RPL, Wetzels JFM, Swinkels DW. The multifaceted role of iron in renal health and disease. Nat Rev Nephrol 2020;16:77–98.
6. Liu J, Sun B, Yin H, Liu S. Hepcidin: a promising therapeutic target for iron disorders: a systematic review. Medicine (Baltim) 2016; 95: e3150.
7. Ganz T, Nemeth E. The hepcidin-ferroportin system as a therapeutic target in anemias and iron overload disorders. Hematol Am Soc Hematol Educ Progr 2011;2011:538–42.
8. Kroot JJ, Kemna EH, Bansal SS, Busbridge M, Campostrini N, Girelli D, et al. Results of the first international round robin for the quantification of urinary and plasma hepcidin assays: need for standardization. Haematologica 2009;94:1748–52.
9. Kroot JJ, van Herwaarden AE, Tjalsma H, Jansen RT, Hendriks JC, Swinkels DW. Second round robin for plasma hepcidin methods: first steps toward harmonization. Am J Hematol 2012;87:977–83.
10. van der Vorm LN, Hendriks JC, Laarakkers CM, Klaver S, Armitage AE, Bamberg A, et al. Toward worldwide hepcidin assay harmonization: identification of a commutable secondary reference material. Clin Chem 2016;62:993–1001.
11. Diepeveen LE, Laarakkers CM, Martos G, Pawlak ME, Uğuz FF, Verberne KE, et al. Provisional standardization of hepcidin assays: creating a traceability chain with a primary reference material, candidate reference method and a commutable secondary reference material. Clin Chem Lab Med 2019;57:864–72.
12. Hoofnagle AN. Harmonization of blood-based indicators of iron status: making the hard work matter. Am J Clin Nutr 2017;106: 1615S–9S.
13. Vesper HW, Myers GL, Miller WG. Current practices and challenges in the standardization and harmonization of clinical laboratory tests. Am J Clin Nutr 2016;104(3 Suppl):907S–12S.
14. Miller WG, Jones GR, Horowitz GL, Weykamp C. Proficiency testing/external quality assessment: current challenges and future directions. Clin Chem 2011;57:1670–80.

15. Jones GR, Albarede S, Kesseler D, MacKenzie F, Mammen J, Pedersen M, et al. Analytical performance specifications for external quality assessment–definitions and descriptions. Clin Chem Lab Med 2017;55:949–55.

16. Stichting Kwaliteitsbewaking Medische Laboratoriumdiagnostiek. Available from: https://www.skml.nl/ [Accessed 25 May 2020].

17. Hepcidin Analysis. Hepcidin reference material: secondary reference material for standardization of hepcidin assays. Available from: http://www.hepcidinanalysis.com/ contentmenu/provided-service/reference-material/ [Accessed 25 May 2020].

18. Laarakkers CM, Wiegerinck ET, Klaver S, Kolodziejczyk M, Gille H, Hohlbaum AM, et al. Improved mass spectrometry assay for plasma hepcidin: detection and characterization of a novel hepcidin isoform. PloS One 2013;8: e75518.

19. Chen M, Liu J, Wright B. A sensitive and cost-effective high-performance liquid chromatography/tandem mass spectrometry (multiple reaction monitoring) method for the clinical measurement of serum hepcidin. Rapid Commun Mass Sp 2020; 34: e8644.

20. Lefebvre T, Dessendier N, Houamel D, Ialy-Radio N, Kannengiesser C, Manceau H, et al. Lc-ms/ms method for hepcidin-25 measurement in human and mouse serum: clinical and research implications in iron disorders. Clin Chem Lab Med 2015;53:1557–67.

21. Itkonen O, Parkkinen J, Stenman UH, Hamalainen E. Preanalytical factors and reference intervals for serum hepcidin lc-ms/ms method. Clin Chim Acta 2012;413:696–701.

22. Pechlaner R, Kiechl S, Mayr M, Santer P, Weger S, Haschka D, et al. Correlates of serum hepcidin levels and its association with cardiovascular disease in an elderly general population. Clin Chem Lab Med 2016;54:151–61.

23. Schmitz EM, Leijten NM, van Dongen JL, Broeren MA, Milroy LG, Brunsveld L, et al. Optimizing charge state distribution is a prerequisite for accurate protein biomarker quantification with LC-MS/MS, as illustrated by hepcidin measurement. Clin Chem Lab Med 2018;56:1490–7.

24. Murao N, Ishigai M, Yasuno H, Shimonaka Y, Aso Y. Simple and sensitive quantification of bioactive peptides in biological matrices using liquid chromatography/selected reaction monitoring mass spectrometry coupled with trichloroacetic acid clean-up. Rapid Commun Mass Spectrom 2007;21:4033–8.

25. Wray K, Allen A, Evans E, Fisher C, Premawardhena A, Perera L, et al. Hepcidin detects iron deficiency in Sri Lankan adolescents with a high burden of hemoglobinopathy: a diagnostic test accuracy study. Am J Hematol 2017;92:196–203.

26. Galesloot TE, Vermeulen SH, Geurts-Moespot AJ, Klaver SM, Kroot JJ, van Tienoven D, et al. Serum hepcidin: reference ranges and biochemical correlates in the general population. Blood Am J Hematol 2011;117:e218–25.

27. Park CH, Valore EV, Waring AJ, Ganz T. Hepcidin, a urinary antimicrobial peptide synthesized in the liver. J Bio 2001;276: 7806–10.

28. White GH, Farrance I. Uncertainty of measurement in quantitative medical testing: a laboratory implementation guide. Clin Biochem Rev 2004;25:S1.

29. Murphy AT, Witcher DR, Luan P, Wroblewski VJ. Quantitation of hepcidin from human and mouse serum using liquid chromatography tandem mass spectrometry. Blood 2007;110: 1048–54.

30. ISO13528. Statistical methods for use in proficiency testing by interlaboratory comparisons; 2005. Available from: https:// www.iso.org/standard/35664.html [Accessed 10 May 2020].

31. Diepeveen LE, Laarakkers CM, Peters HP, van Herwaarden AE, Groenewoud H, IntHout J, et al. Unraveling hepcidin plasma protein binding: evidence from peritoneal equilibration testing. J Pharm 2019;12:123.

32. Hoofnagle AN, Wener MH. The fundamental flaws of immunoassays and potential solutions using tandem mass spectrometry. J Immunol Methods 2009;347:3–11.

33. Hepcidin 25 (bioactive) HS ELISA. DRG diagnostics. Available from: https://www.drg-diagnostics.de/files/2015-11_hepcidin_ hybrid-xl_elisa.pdf [Accessed 26 May 2020].

34. HEPCIDIN-25 Chemiluminescent ELISA (Hepcidin-25 Peptide Detection). Broomfield, Colorado, USA: Corgenix; 2016. [package insert].

35. Kroot JJ, Hendriks JC, Laarakkers CM, Klaver SM, Kemna EH, Tjalsma H, et al. (Pre)analytical imprecision, between-subject variability, and daily variations in serum and urine hepcidin: implications for clinical studies. Anal Biochem 2009;389:124–9.

36. Schaap CC, Hendriks JC, Kortman GA, Klaver SM, Kroot JJ, Laarakkers CM, et al. Diurnal rhythm rather than dietary iron mediates daily hepcidin variations. Clin Chem 2013;59:527–35.

37. Kemna EH, Tjalsma H, Podust VN, Swinkels DW. Mass spectrometry-based hepcidin measurements in serum and urine: analytical aspects and clinical implications. Clin Chem 2007;53: 620–8.

38. Ganz T, Olbina G, Girelli D, Nemeth E, Westerman M. Immunoassay for human serum hepcidin. Blood 2008;112:4292–7.