

Laura E. Diepeveen, Coby M.M. Laarakkers, Gustavo Martos, Marta E. Pawlak, Fatih F. Uğuz, Kim E.S.A. Verberne, Rachel P.L. van Swelm, Siem Klaver, Anton F.J. de Haan, Kelly R. Pitts, Sukhvinder S. Bansal, Ioana M. Abbas, Marianne Fillet, Thibaud Lefebvre, Anneke J. Geurts-Moespot, Domenico Girelli, Annalisa Castagna, Matthias Herkert, Outi Itkonen, Gordana Olbina, Naohisa Tomosugi, Mark E. Westerman, Vincent Delatour, Cas W. Weykamp and Dorine W. Swinkels*

Provisional standardization of hepcidin assays: creating a traceability chain with a primary reference material, candidate reference method and a commutable secondary reference material

<https://doi.org/10.1515/cclm-2018-0783>

Received July 23, 2018; accepted October 4, 2018

Abstract

Background: Hepcidin concentrations measured by various methods differ considerably, complicating interpretation. Here, a previously identified plasma-based candidate secondary reference material (csRM) was modified into a serum-based two-leveled sRM. We validated its functionality to increase the equivalence between methods for international standardization.

Methods: We applied technical procedures developed by the International Consortium for Harmonization of Clinical Laboratory Results. The sRM, consisting of lyophilized serum with cryolyoprotectant, appeared commutable among nine different measurement procedures using 16 native human serum samples in a first round robin (RR1). Harmonization potential of the sRM was simulated in RR1 and evaluated in practice in RR2 among 11 measurement

procedures using three native human plasma samples. Comprehensive purity analysis of a candidate primary RM (cpRM) was performed by state of the art procedures. The sRM was value assigned with an isotope dilution mass spectrometry-based candidate reference method calibrated using the certified pRM.

Results: The inter-assay CV without harmonization was 42.1% and 52.8% in RR1 and RR2, respectively. In RR1, simulation of harmonization with sRM resulted in an inter-assay CV of 11.0%, whereas in RR2 calibration with the material resulted in an inter-assay CV of 19.1%. Both the sRM and pRM passed international homogeneity criteria and showed long-term stability. We assigned values to the low (0.95 ± 0.11 nmol/L) and middle concentration (3.75 ± 0.17 nmol/L) calibrators of the sRM.

Conclusions: Standardization of hepcidin is possible with our sRM, which value is assigned by a pRM. We propose the implementation of this material as an international calibrator for hepcidin.

***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
Laura E. Diepeveen, Coby M.M. Laarakkers, Rachel P.L. van Swelm and Siem Klaver: Department of Laboratory Medicine, Radboud University Medical Center, Nijmegen, The Netherlands; and Hepcidinanalysis.com, Nijmegen, The Netherlands
Gustavo Martos: Laboratoire National de Métrologie et d'Essais (LNE), Paris, France; and Departement de Chimie, Bureau International des Poids et Mesures (BIPM), Sèvres, France
Marta E. Pawlak, Fatih F. Uğuz, Kim E.S.A. Verberne and Anneke J. Geurts-Moespot: Department of Laboratory Medicine, Radboud University Medical Center, Nijmegen, The Netherlands
Anton F.J. de Haan: Department for Health Evidence, Radboud University Medical Center, Nijmegen, The Netherlands
Kelly R. Pitts: Corgenix Medical Corporation, Broomfield, CO, USA
Sukhvinder S. Bansal: Institute of Pharmaceutical Sciences, King's College London, London, UK

Ioana M. Abbas: Division 1.5 Protein Analysis, Federal Institute for Materials Research and Testing (BAM), Berlin, Germany
Marianne Fillet: Laboratory for the Analysis of Medicines, CIRM, University of Liège, Liège, Belgium
Thibaud Lefebvre: French Center of Porphyria, INSERM U1149, Sorbonne Paris Cité University, Labex GR-Ex, Louis Mourier Hospital APHP, Colombes, France
Domenico Girelli and Annalisa Castagna: Department of Medicine, University of Verona, Verona, Italy
Matthias Herkert: DRG Instruments, Marburg, Germany
Outi Itkonen: Laboratory Division HUSLAB, Helsinki University Central Hospital, Helsinki, Finland
Gordana Olbina and Mark E. Westerman: Intrinsic Life Sciences, La Jolla, CA, USA
Naohisa Tomosugi: Division of Advanced Medicine, Kanazawa Medical University, Ishikawa, Japan
Vincent Delatour: Laboratoire National de Métrologie et d'Essais (LNE), Paris, France
Cas W. Weykamp: Department of Clinical Chemistry, Queen Beatrix Hospital, Winterswijk, The Netherlands

Keywords: harmonization; hepcidin; iron metabolism; primary reference material; secondary reference material; standardization.

Introduction

Hepcidin is a liver-produced peptide hormone, which regulates systemic iron levels by counteracting the cellular iron exporter ferroportin [1, 2]. Disorders of iron metabolism often arise in combination with alterations in hepcidin levels. Therefore, hepcidin is considered a powerful diagnostic tool as well as a target of therapy. Consequently, many hepcidin measurement procedures (MPs) have been developed to quantify hepcidin concentrations in biological fluids, largely based on either mass spectrometry (MS) or immunochemistry (IC) [2].

Our previous round robin (inter-laboratory) studies (RRs) revealed hepcidin levels in the same clinical sample differed by a factor up to 9.1 between MPs, leading to non-equivalent results of results, complicating interpretation [3–5]. This hampers implementation of hepcidin MPs into clinical practice and for research purposes. Better agreement between hepcidin results obtained by various MPs will enable the establishment of global uniform reference intervals and decision limits, i.e. for guidance in oral iron therapy and diagnosis of iron refractory iron deficiency anemia [2, 6, 7]. Therefore, ideally, all hepcidin MPs need to be standardized. This allows a patient sample being traceable to the internationally recognized *Système International* (SI) via a calibration hierarchy of calibrators and MPs, and the establishment of a ‘true value’ by defining the measurand in SI units [8]. For standardization a primary reference material (pRM) is needed, which is a compound of well-characterized purity used to prepare a calibrator for a reference MP (RMP) [8]. This calibrated RMP is used to assign a value to a secondary reference material (sRM), as a sRM is matrix-based material which can be used for the calibration of routine MPs worldwide [9, 10]. However, until now no reference materials exist for hepcidin.

In case solely a sRM is developed, only harmonization can be achieved, which is a lower category of traceability leading to equalized results worldwide. In our previous hepcidin harmonization study [5], we applied technical procedures to achieve efficient harmonization according to those recently developed by the International Consortium for Harmonization of Clinical Laboratory Results (ICHCLR) [10–12]. Harmonization appeared technically achievable and lyophilized human plasma with

cryolyoprotectant (CLP), which improves stability, was identified as a commutable candidate sRM (csRM). Moreover, simulation studies showed that calibration with this csRM reduced the inter-method coefficient of variation (CV) from 28.6% to 7.7% [5].

Here, we report results for making standardization of worldwide hepcidin MPs possible by (i) modifying the previous identified csRM, producing a large batch of a two-leveled (i.e. covering low and middle concentration range of hepcidin) sRM consisting of lyophilized human serum with CLP for international use, confirming its commutability and estimating the sRM harmonization potential in a first RR study (RR1) by mathematical simulation, (ii) testing the sRM harmonization potential in practice in a small second RR study (RR2) and (iii) establishing a traceability chain allowing standardized results through value assignment of our sRM with a candidate RMP that relies on calibration with a pRM of well-characterized purity.

Materials and methods

Overview study design

The study was comprised of a stepwise approach towards standardization.

First, we produced a sRM, which is a modification of our previous identified csRM [5]. Here, we used serum instead of plasma, produced a two-leveled calibrator, covering both the low and middle range of hepcidin concentrations, and tested all included serum samples for HIV, hepatitis B and C. We assessed commutability and the harmonization potential of this modified material in RR1. Next, the sRM harmonization potential was tested in practice in RR2. We assessed the analytical performance of participating MPs in both studies, as harmonization is solely possible for MPs with good analytical performance.

To reach standardization, we established a traceability chain to the SI units through value assignment of the sRM with a candidate RMP calibrated with a pRM. Commercially obtained synthetic hepcidin was used as candidate pRM (cpRM). Amino-acid analysis and comprehensive impurity analysis of the cpRM, performed by liquid chromatography high-resolution tandem MS, were performed for value assignment [13].

Both the sRM and pRM were tested for homogeneity and stability.

Measurement procedures

MPs were selected to represent a variety of methodological approaches as described previously [5]. For RR1 nine MPs agreed to participate (4 MS- and 5 IC-based), and 11 MPs (8 MS- and 3 IC-based) for RR2. Relevant MP approaches and their participation in RR1 or/and RR2 are illustrated in Table 1.

Table 1: Methodological approaches of participating hepcidin MPs.

Participant code RR1	Participant code RR2	MP	Extraction	Standard	Manufacturer standard	Reference
MS-1	MS-1	MALDI-TOF MS	WCX	Heavy-isotope labeled synthetic hepcidin-25	Peptide Inst. (Japan)	[14]
N/A	MS-2	UPLC-MS/MS	Reversed phase	Heavy-isotope labeled synthetic hepcidin-25	Manufactured in-house	[15]
MS-3	MS-3	LC-MS/MS	Reversed phase	Heavy-isotope labeled synthetic hepcidin-25	Peptide Inst. (Japan)	[16]
MS-4	MS-4	LC-MS/MS	SPE-Oasis HLB	Heavy-isotope labeled synthetic hepcidin-25	Peptide Int. (USA)	[17]
MS-5	MS-5	LC-MS/MS	HLB SPE	Heavy-isotope labeled synthetic hepcidin-25	Peptide Inst. (Japan)	[18]
N/A	MS-6	LC-MS/MS	SPE-Oasis MAX	Heavy-isotope labeled synthetic hepcidin-25	Peptide Inst. (Japan)	[19]
N/A	MS-7	LC-MS/MS	Precipitation with ACN/TFA	Heavy-isotope labeled synthetic hepcidin-25	Peptide Inst. (Japan)	[20]
N/A	MS-8	UHPLC-MS/MS	SPE-Oasis uElution HLB	Heavy-isotope labeled synthetic hepcidin-25	Peptide Inst. (Japan)	^a
IC-1	N/A	cELISA	None	Synthetic hepcidin-25	Peptide Int. (USA)	[21]
IC-3	IC-3	cELISA (kit-commercially)	None	Synthetic hepcidin-25	Peptide Int. (USA)	^b
IC-4	IC-4	cELISA (test-automated)	None	Synthetic hepcidin-25	Peptide Int. (USA)	^c
IC-5	IC-5	cELISA	None	Synthetic hepcidin-25	Bachem, Germany	[22] ^d
IC-6	N/A	Direct-sandwich CL ELISA	None	Synthetic hepcidin-25	Bachem/Peptide Inst. (Japan)	^e

RR, round robin; 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; c, competitive; ELISA, enzyme-linked immunosorbent assay; CL, chemiluminescence; WCX, weak-cation exchange; HLB, hydrophilic lipophilic balanced reversed phase; N/A, not applicable, because these assays solely participated in one RR. Two procedures (MS-2 and IC-2) were not ready to measure the samples at the time of sample send-out for RR1. ^aNo reference available; a laboratory-developed MP for hepcidin-25 by Laboratory for the Analysis of Medicines, University of Liège, Liège, Belgium. ^bNo reference available; Intrinsic Hepcidin IDxTM Kit (product #ICE-007) – commercially available competitive ELISA from Intrinsic LifeSciences, La Jolla, USA. ^cNo reference available; Intrinsic Hepcidin IDxTM Test – automated competitive ELISA from Intrinsic LifeSciences, La Jolla, USA. ^dHepcidin 25 bioactive HS Kit (product #EIA-5782) commercially available assay from DRG Diagnostics, Marburg, Germany. ^eNo reference available; a second-generation MP based on the assay described in Butterfield et al. [23] developed by ELISA of Corgenix Inc., Broomfield, USA.

Collection and preparation of native samples and the sRM

For RR1, we collected 16 native human serum samples, covering the low (0.6–1.2 nmol/L), middle (2.6–7.9 nmol/L) and high (8.0–12.8 nmol/L) range of hepcidin concentrations in humans (Supplementary Table 1), measured by our validated in-house MP (MS-1, Table 1) [14]. Samples, all negative for HIV, hepatitis B and hepatitis C, were obtained from phlebotomized iron-depleted patients with hereditary hemochromatosis and healthy volunteers (see Supplementary Material), and were stored at –80 °C until sample set preparation.

Within 6 weeks after collection, the 16 serum samples were used to produce the sRM, as well as to prepare sample sets for each participating method, as previously described [5], based on technical procedures developed by the ICHCLR [12]. Each sample set consisted of (i) 16 individual samples with low (n=5), middle (n=6) and high (n=5) hepcidin concentrations to assess commutability of the sRM, (ii) mixtures of the individual samples to prepare a linearity panel to assess analytical performance of each MP and (iii) two-leveled sRM made of native samples with low and middle

hepcidin concentrations (Supplementary Table 1). Due to ethical issues (i.e. collecting high volume samples from acutely and chronically inflamed patients), we did not produce a calibrator for the high hepcidin concentration range.

In order to produce the sRM, low level samples were pooled and homogenized with addition of CLP at RT for approximately 1 h forming the low concentration calibrator. Likewise for middle level samples forming the middle concentration calibrator. All vials were lyophilized in a freeze-dryer (Zirbus Technology, Tiel, The Netherlands) for 63 h (Supplementary Table 2).

For RR2, we prepared a sample set consisting of six native samples (i.e. duplicate low-, middle- and high hepcidin concentration samples) and the two-leveled sRM. Samples, used for both assessment of harmonization potential, and analytical performance of participating MPs, were obtained from leftover heparin plasma of intensive care patients and from a phlebotomy of an iron-depleted hereditary hemochromatosis patient (see Supplementary Material). As in RR1, hepcidin concentrations were measured by MS-1 (Table 1) and samples were stored at –80 °C until sample set preparation.

Sample sets for RR1 and RR2, including the two-leveled sRM, were shipped on dry ice and participating laboratories were requested to prepare the samples according to provided instructions and to measure, in triplicates, each sample within one run.

Ethics

The study was approved by the Ethics Committee and the Board of Directors of the Radboudumc, Nijmegen, The Netherlands, and has been conducted in accordance with the Declaration of Helsinki. Both patients and volunteers signed informed consent prior to blood collection and all samples were blinded and randomized.

Use of leftover patient plasma in this study conformed to the code for proper secondary use of human tissue in the Netherlands.

Statistics

Exclusion of outlier results and analytical performance of MPs were assessed as previously described (see Supplementary Material) [5]. All statistical analyses were performed with SAS 9.4.

Commutability of the sRM: To validate commutability in RR1 of the modified sRM, we performed regression analysis with the results of 16 native samples and sRM for each method (y-axis) and the mean of all methods (x-axis), and assessed whether mean results of both calibrator samples were within the 95% prediction interval of the regression line of the mean of the individual samples [24].

Equivalence between MPs without harmonization: Linear regression was performed between the results of each of the MPs (y-axis) and the mean results of all MPs (x-axis), summarized in terms of the Pearson correlation coefficient r , the intercept a and slope b of the ordinary least squares (OLS) regression line, to explore the degree of equivalence without harmonization. Equivalence without harmonization was quantified as the mean inter-assay CV (%), calculated from the logarithmically transformed means of the results of the native samples (RR1: $n=16$; RR2: $n=3$).

Harmonization with the sRM: In RR1, the effect of harmonization on the equivalence between methods using the sRM was mathematically simulated by value reassignment based on regression of the results of the sRM samples per MP against the respective mean result of all MPs. The inter-assay dispersion in these simulated results was then expressed as the inter-assay CV (%) after harmonization with the sRM (see Supplementary Material).

In RR2, participants received two sets of identical samples, which allowed both prospective and retrospective calibration methods (see Supplementary Material). The between MP dispersion in the calibrated results was then expressed as the inter-assay CV (%).

To assess whether the achieved inter-assay CV is adequate, equivalence of MPs was placed in the context of the biological variation of hepcidin. Therefore we calculated limits for total allowable error (TEa) with $TEa(\%) = 1.65(f_1 \cdot CV_I) + f_2(CV_I^2 + CV_G^2)^{1/2}$, in which CV_I is the between-day intra-individual CV, CV_G is the inter-individual variation and f_1 and f_2 are factors for optimum (0.25 and 0.125), desirable (0.5 and 0.250), and minimum (0.75 and 0.375) TEa [25]. Values for CV_I (48.8%) and CV_G (154.1%) were derived from Murphy et al. [26].

Characterization of the sRM

Homogeneity was tested according to ISO13528 by means of duplicate measurements of 12 randomly selected low- and middle-concentrations calibrator samples by MS-1 [27]. Prior to use, the sRM was reconstituted with 0.30 mL deionized water, followed by leaving the vial at room temperature for 15 min, and subsequent careful mixing for 20 min (roller bench, 3.5 rpm). In the absence of between-lab variation data from external proficiency testing, and as within-lab between-run variation is always smaller than between-lab variation, we used data on between-run reproducibility from control charts of MS-1 for calculations on homogeneity, i.e. SD of 0.157 nM for the low concentration calibrator (lowest control level 2.34 nM, CV 6.70%) and SD 0.557 nM for the middle concentration calibrator (highest control level 10.1 nM, CV 5.50%).

Stability was studied by comparing hepcidin concentrations in aliquots stored at -20 , 4 and 20 °C. Measurements took place by MS-1 at baseline, after 1 month, 6 months, 1, 1.5 and 2 years of storage, and will be continued annually for the next 5 years. Significant changes in time concerning the value of the sRM, that exceeds the precision of MS-1, were considered as an instability. Statistical analysis was done using analysis of variance (ANOVA) and Bonferroni's multiple comparison test.

Characterization of the pRM

Synthetic human hepcidin-25 was purchased from Anaspec (Seraing, Belgium). Characterization of the cpRM was performed at the French Metrology Institute LNE (Laboratoire national de métrologie et d'essais). Hepcidin mass quantity was determined by amino acid analysis using isotope-dilution liquid-chromatography MS (ID-LC/MS) using a triple quadrupole MS (Waters Xevo TQ-MS, see Supplementary Material).

Since peptide impurities can positively bias amino acid analysis results, impurity identification and quantification was performed on a high resolution quadrupole-orbitrap MS (Thermo Q-Exactive Focus, Waltham, MA, USA) according to the procedure described previously [13]. Impurities were quantified on three hepcidin vials from Anaspec, stored at -20 °C and measured at three different time points (along an 8-month period). The average total impurity content determined by the q-orbitrap MS was subtracted from the amount determined by amino acid analysis using the TQ-MS.

Homogeneity testing was performed on randomly selected vials belonging to a single batch. Seven measurement replicates were performed on each vial by MS-1. Uncertainty associated to homogeneity was calculated by ANOVA.

Stability testing was performed by comparing impurity profiles, measured by ID-LC/MS, over an 8-month period. Impurities stability data were analyzed by regression analysis (see Supplementary Material).

Value assignment of sRM

Value assignment of the sRM was performed in two parts using a validated Weak-Cation-eXchange MALDI-Time of Flight – MS (WCX MALDI-TOF MS) assay as a candidate RMP (MS-1) [14]. In the first part, we reassigned the value of the internal standard of MS-1

(heavy hepcidin⁴⁰, Peptide Institute, [Japan]) using the certified pRM. Secondly, we used this adjusted internal standard to assign a value to the sRM. Details are described in the Supplementary Material. Statistical analysis was done using ANOVA.

Results

Production and validation of the sRM

In total, 1198 vials of both the low and middle concentrations calibrators were produced. Together they form sRM, consisting of lyophilized serum with CLP. The material was validated in RR1 by nine MPs with good analytical performance, assessed in terms of mutual correlations (Supplementary Table 3), linearity and reproducibility (Supplementary Table 4), which allows harmonization technically. Mean intra-assay variation appeared higher at low hepcidin concentrations (7.4%) compared to middle (3.4%) and high (2.9%) concentrations. In accordance with previous studies [3–5], the occurrence of large discrepancies in regression slopes (Supplementary Table 4) and in absolute hepcidin levels (Supplementary Table 5) despite high MPs correlations suggests calibration, not heterogeneity of the measurand, as a major cause of the current lack of equivalence between assays.

Without harmonization, the inter-assay CV between MPs in RR1 was calculated to be 42.1% (Table 2), of which 92.4% could be attributed to the lack of calibration (Supplementary Table 6). Mathematical simulation of harmonization of 16 native samples with the sRM revealed a significant improvement of equivalence between MPs, as the inter-assay CV decreased to 11.0% (Table 2).

To assess whether the achieved inter-assay CV is adequate for the biological variation of hepcidin, we used the criterion of TEa, combining bias and imprecision. Based on previous reported intra-individual CV and

inter-individual CV of hepcidin, TEa of 40.3% (optimum), 80.7% (desirable) and 121.0% (minimum) were calculated [26]. The higher the biological variation of a measurand, the higher the allowable bias after harmonization. Without harmonization, results of only half of the MPs were within the optimum TEa limits for hepcidin (Figure 1A). After calibration with the sRM, all simulated results were within the limits for the optimum TEa, which illustrates the decreased bias and the significant improvement in equivalence between the different MPs (Figure 1B).

In addition, commutability of the sRM was confirmed as for all MPs the mean of the triplicate measurements of the calibrator samples were within the 95% prediction interval of the regression line of the results of the 16 native samples for each MP and the mean of all MPs (Supplementary Figure 1). This corroborates the previously validated commutability of the now slightly modified sRM [5].

With respect to homogeneity testing, for the middle- and low-concentrations calibrators we found SDs of 0.098 nM and 0.041 nM, respectively, indicating that both passed the criteria as defined by ISO13528 [27]. The sRM was found to remain stable for at least 2 years at –20, 4 and 20 °C.

Harmonization potential of the sRM in practice

Next, we assessed the harmonization potential of the sRM in practice in RR2 among 11 MPs with good analytical performance in terms of reproducibility (Supplementary Table 7), as shown previously (RR1, [5]). Again, prior to harmonization large discrepancies in absolute hepcidin levels were found (Supplementary Table 8).

Without harmonization, the inter-assay CV between assays in RR2 was calculated to be 52.8%, which decreased to 19.1% after calibration with the sRM (Table 2). As in RR1, we placed these results in the context of biological variation of hepcidin using the TEa, and again all the results were within the limits or on the cut-off point of the optimum TEa after harmonization with the sRM illustrating decreased bias and significant improvement in equivalence between MPs (Figure 1C and D).

Value assignment of the sRM using a pRM

Characterization of the pRM

The average peptide mass fraction of hepcidin in the pRM was determined by combining results of triplicate

Table 2: The effect of the sRM on inter-assay CV after mathematical simulation of calibration (RR1) or in practice (RR2).

Equivalence (inter-assay CV, %)	RR1	RR2
Before calibration with sRM	42.1	52.8
After calibration with sRM	11.0	19.1

MP, measurement procedure; CV, coefficient of variation; sRM, secondary reference material; RR, round robin. Equivalence after calibration with the sRM calculated as the inter-assay CV (%) after value reassignment of the 16 native samples using OLS regression equations of the produced sRM set (RR1) or calculated as the inter-assay CV (%) of calibrated measurements (RR2).

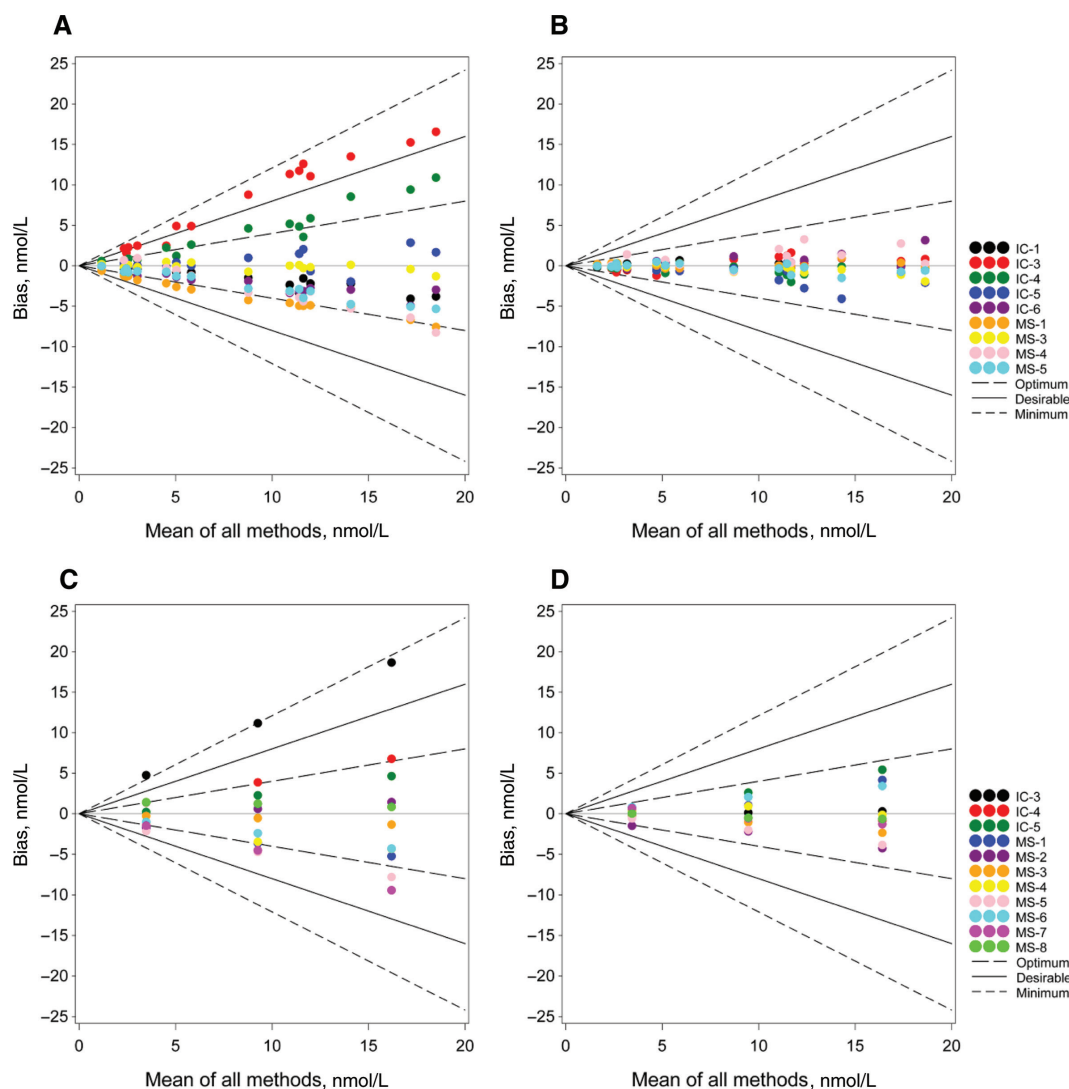


Figure 1: The effect of calibration with the sRM in RR1 and RR2, including the criterion of TEa. MP results before (A) and after (B) simulation of calibration with the sRM in RR1, and before (C) and after (D) calibration with the sRM in RR2, including the criterion of TEa. The data points above each other represent the measurement of the samples ($n=16$ in RR1, $n=3$ in RR2) by all different MPs. Because of the absence of a true value, the x-axis represents the mean results of the samples for all MPs. The y-axis shows the bias, i.e. the difference, between the results of the native samples of each MP and the mean of all MPs. The lines represent limits for optimum (large dashes), desirable (solid line) and minimum (small dashes) TEa.

hydrolyses performed on 3 different days (Supplementary Table 9). Results were consistent among hydrolysis replicates (RSD 4.1%) and between quantified amino acids (RSD 1.8%). This value was corrected by subtracting peptidic impurities quantified by high resolution MS, as previously described [13]. Relative abundances of major identified impurities, all structurally related to hepcidin-25 (N- and C-terminal truncated forms and modified forms, e.g. oxidation), are shown in Supplementary Table 10.

Impurities were quantified at 3 different time points, covering an 8-month stability period. No significant trend (99% confidence level) was detected in the impurity mass

fraction values. At the end of the testing period, the material was analyzed with ID-LC/MS and no new impurities were detected. The material is therefore stable at -20°C . The pRM was considered sufficiently homogeneous, as the calculated uncertainty associated with heterogeneity of the vials was 1.65%. Combined with the uncertainty from amino acid analysis, impurity quantification and material heterogeneity, the final expanded uncertainty was obtained. Hepcidin mass fractions were converted into mass quantity of hepcidin per vial by taking the gravimetrically measured reconstitution volume (0.5 mL) into account and correcting for peptidic impurities, leaving

the certified hepcidin mass quantity value assigned to the pRM to be $72.9 \pm 3.9 \mu\text{g}$ ($k=2$). The certified uncertainty is the expanded uncertainty with a coverage factor 2 corresponding to a level of confidence of about 95% estimated in accordance with ISO/IEC Guide 98-3 [28].

Value assignment of the sRM

Using value assigned pRM, and MS-1 as the provisional RMP, we assigned values to the low- and middle-concentrations calibrators of the sRM which are $0.95 \pm 0.11 \text{ nmol/L}$ and $3.75 \pm 0.17 \text{ nmol/L}$, respectively ($k=1$). These values differ approximately a factor 2 with the all method mean determined in RR1 ($2.38 \pm 1.22 \text{ nmol/L}$ and $7.03 \pm 3.15 \text{ nmol/L}$).

Discussion and conclusions

Previous RR studies showed the need for standardization of hepcidin MPs, as hepcidin concentrations measured by various methods differ considerably, hampering implementation of these MPs into clinical practice and for research purposes [3–5]. Therefore, we created all items needed to establish traceability to SI, i.e. a commutable two-leveled sRM made of lyophilized serum with CLP, with a value assigned by a candidate RMP calibrated with a pRM.

We showed, both by mathematical simulation (RR1) and in practice (RR2), that the use of the sRM to calibrate hepcidin MPs with good analytical performance can achieve a reduction of the inter-method CV that is within the optimum TEa, i.e. the inter-assay CV based on hepcidin's biological variation [26, 29]. We currently assessed TEa on limited data on a relatively high hepcidin between-day biological variability (48.8%) [26]. However, published [29–32] and our unpublished observations (L. Diepeveen) corroborate these data. Nevertheless expansion of data on between-day variability might be needed to achieve a more accurate estimation of the TEa that is better suited to assess the achievements of global standardization.

We found that inter-assay CVs before and after calibration differed slightly between RR1 (before: 42.1%, after: 11.0%) and RR2 (before: 52.8%, after: 19.1%). These differences may be explained by (i) varying numbers of measured samples (RR1: $n=16$, RR2: $n=3$), (ii) the use of pro- or retrospective calibration and (iii) variation in methodology of participating MPs (RR1: 4 MS, RR2: 8 MS) and associated differences in analytical performance.

We show that the sRM improves calibration bias, which was found to be the major contributor to measurement inaccuracy. It is conceivable to further reduce the inter-assay CV by improvement of other factors contributing to inaccurate measurements like analytical characteristics of MPs and heterogeneity of the measurand. The latter could be caused by differences in detection of the smaller hepcidin isoforms [5, 14]. Bioactive hepcidin-25 can be measured specifically by MS methods in contrast to most IC methods, which measure total hepcidin concentrations that include potential isoforms [2, 21]. Isoforms are increasingly present in patients with elevated hepcidin-25 levels (e.g. sepsis or chronic kidney disease). As defining the measurand is a requisite for standardization [6, 8], it is important that the value assignment of the sRM was performed by a MS method, assuring specific quantification of the defined measurand hepcidin-25, traceable to SI units (nmol/L).

We have produced two calibrators with respective concentrations of $0.95 \pm 0.11 \text{ nmol/L}$ and $3.75 \pm 0.17 \text{ nmol/L}$ ($k=1$). Production of a third, high concentration calibrator will further improve future standardization. However, this requires samples with high hepcidin levels originating from inflamed patients leading to ethical difficulties. Moreover, the provisional cRMP used in our study should be replaced by a validated RMP [8].

Usage of the sRM by clinical and research laboratories, and commercial providers of ELISA kits, will allow the standardization of hepcidin MPs worldwide. We anticipate that standardization of calibration will have a significant impact on the success of hepcidin as a biomarker as it allows the definition and application of generally accepted uniform reference intervals and decision limits, facilitating medical research and its translation to the clinic [7]. Standardization might become especially advantageous for using hepcidin in the diagnosis of iron refractory iron deficiency anemia to predict the effectiveness of oral iron supplementation to replenish deficient body iron stores [2].

The sRM will be made available to the medical, scientific and pharmaceutical community by Hepcidinanalysis.com at a fee for the service. We also intend to submit the material for review and possible publication in the JCTLM Database. We must emphasize that worldwide standardization can only be achieved if the sRM is implemented as the international calibrator for hepcidin.

Acknowledgments: We would like to recognize our friend and colleague Dr. Mark Westerman (deceased), Founder and CEO of Intrinsic LifeSciences, LLC, La Jolla, California, for his major contribution to hepcidin research and

for the development of hepcidin assays and products for use in research and clinical laboratories. We thank Drs Alexander Rennings and Mirian Janssen for their help in HFE-hemochromatosis patient inclusion, Ankie Sneek for arranging blood collection, Jochem Pastoor for his help in the design of the figures, and all patients and volunteers for their participation in the study.

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

Research funding: We would like to acknowledge Fondazione Cariverona. The funding organization played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

Employment or leadership: LD, CL, SK, RvS and DS are employees of Radboudumc, that via its Hepcidinanalysis.com initiative offers high quality hepcidin measurements to the medical, scientific and pharmaceutical community at a fee for the service basis. KRP is an employee of Corgenix Medical Corporation specializing in IVD development, manufacturing, and distribution. MW (deceased) was President and CEO of Intrinsic LifeSciences LLC. GO is an employee of Intrinsic LifeSciences LLC and has ownership interest in the company. Intrinsic LifeSciences is engaged in the commercial development of hepcidin and related diagnostic products and provides CLIA compliant hepcidin testing to research, clinical and pharmaceutical clients worldwide.

Honorarium: None declared.

Competing interests: The funding organization(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

References

- 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.
- Girelli D, Nemeth E, Swinkels DW. Hepcidin in the diagnosis of iron disorders. *Blood* 2016;127:2809–13.
- 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.
- 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.
- 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.
- Vesper HW, Myers GL, Miller WG. Current practices and challenges in the standardization and harmonization of clinical laboratory tests. *Am J Clin Nutr* 2016;104:907S–12S.
- Hoofnagle AN. Harmonization of blood-based indicators of iron status: making the hard work matter. *Am J Clin Nutr* 2017;106:1615S–9S.
- ISO17511:2003. In vitro diagnostic medical devices – measurement of quantities in biological samples – metrological traceability of values assigned to calibrators and control materials. <https://www.iso.org/standard/30716.html>. Accessed: 10 Sept 2018.
- Miller WG. Harmonization: its time has come. *Clin Chem* 2017;63:1184–6.
- Myers GL, Miller WG. The international consortium for harmonization of clinical laboratory results (ichclr) – a pathway for harmonization. *Ejifcc* 2016;27:30–6.
- Myers GL, Miller WG. The roadmap for harmonization: status of the international consortium for harmonization of clinical laboratory results. *Clin Chem Lab Med* 2018;56:1667–72.
- Weykamp C, Eckfeld J, Vesper H, Thienpont L, Burns C, Caliendo A, et al. AACC harmonization initiative; toolbox of technical procedures to achieve harmonization for a measurand 2012. www.harmonization.net. Accessed: 10 Sept 2018.
- Bros P, Josephs RD, Stoppacher N, Cazals G, Lehmann S, Hirtz C, et al. Impurity determination for hepcidin by liquid chromatography-high resolution and ion mobility mass spectrometry for the value assignment of candidate primary calibrators. *Anal Bioanal Chem* 2017;409:2559–67.
- Laarakkers CM, Wiegierinck 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.
- 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.
- Murao N, Ishigai M, Yasuno H, Shimomura 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.
- Udali S, Castagna A, Corbella M, Ruzzenente A, Moruzzi S, Mazzi F, et al. Hepcidin and DNA promoter methylation in hepatocellular carcinoma. *Eur J Clin Invest* 2018;48.
- 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.
- 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.
- Abbas IM, Hoffmann H, Montes-Bayon M, Weller MG. Improved lc-ms/ms method for the quantification of hepcidin-25 in clinical samples. *Anal Bioanal Chem* 2018;410:3835–46.
- Kroot JJ, Laarakkers CM, Geurts-Moespot AJ, Grebenchtchikov N, Pickkers P, van Ede AE, et al. Immunochemical and

- mass-spectrometry-based serum hepcidin assays for iron metabolism disorders. *Clin Chem* 2010;56:1570–9.
22. Zipperer E, Post JG, Herkert M, Kundgen A, Fox F, Haas R, et al. Serum hepcidin measured with an improved elisa correlates with parameters of iron metabolism in patients with myelodysplastic syndrome. *Ann Hematol* 2013;92:1617–23.
23. Butterfield AM, Luan P, Witcher DR, Manetta J, Murphy AT, Wroblewski VJ, et al. A dual-monoclonal sandwich elisa specific for hepcidin-25. *Clin Chem* 2010;56:1725–32.
24. Budd JR, Weykamp C, Rej R, MacKenzie F, Ceriotti F, Greenberg N, et al. IFCC working group recommendations for assessing commutability part 3: using the calibration effectiveness of a reference material. *Clin Chem* 2018;64:465–74.
25. Fraser CG, Hyltoft Petersen P, Libeer JC, Ricos C. Proposals for setting generally applicable quality goals solely based on biology. *Ann Clin Biochem* 1997;34:8–12.
26. 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.
27. ISO13528:2005. Statistical methods for use in proficiency testing by interlaboratory comparisons. <https://www.iso.org/standard/35664.html>. Accessed: 10 Sept 2018.
28. ISO/IEC Guide 98-3:2008. Uncertainty of measurement – part 3: guide to the expression of uncertainty in measurement (gum:1995). <https://www.iso.org/standard/50461.html>. Accessed: 10 Sept 2018.
29. 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.
30. 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.
31. Kemna E, Tjalsma H, Laarakkers C, Nemeth E, Willems H, Swinkels D. Novel urine hepcidin assay by mass spectrometry. *Blood* 2005;106:3268–70.
32. Ganz T, Olbina G, Girelli D, Nemeth E, Westerman M. Immunoassay for human serum hepcidin. *Blood* 2008;112:4292–7.

Supplementary Material: The online version of this article offers supplementary material (<https://doi.org/10.1515/cclm-2018-0783>).

Supplemental material

Supplemental methods

Samples RR1

Since RR1 is designed to produce a large batch of two-leveled sRM and as a validation of our previous study [1], we recruited 16 individuals instead of 32 that are required for a harmonization study according to ICHCLR. [2] These 16 individual serum samples were selected to cover the low (0.6-1.2 nmol/L), middle (2.6-7.9 nmol/L) and high (8.0-12.8 nmol/L) range of hepcidin concentrations in humans, as measured by MS-1 (Table 1). [3] To this end, we collected 300 mL whole blood from individuals with low or middle hepcidin concentrations by venipuncture in phlebotomy bags (Macopharma VSL7001PD, Tourcoing, France) and 50 mL whole blood from individuals with a high hepcidin concentration in five 10 mL tubes (Becton Dickinson Vacutainer® SST™ II Advance serum Tubes, Franklin Lakes, United States). We chose to collect 300 mL whole blood from individuals with low or middle hepcidin concentrations to ensure obtaining a minimum of 90 mL serum after processing, in order to produce the reference material besides study purposes. We chose to collect 50 mL whole blood from individuals with a high hepcidin concentration to ensure obtaining a minimum of 10 mL serum for the samples in the high concentration range needed for study purposes. Blood collections from phlebotomized iron-depleted HFE-hemochromatosis patients and healthy volunteers (age: 18-75 years) took place at the Radboudumc, Nijmegen. Collected blood was transported on ice (~5 minutes) and centrifuged at 3450g (4°C) for 8 minutes. After centrifugation, plasma was pipetted into glass vials to coagulate for 4 hours, centrifuged (1570g, 19 minutes, RT) and stored at -80°C until sample set preparation.

Within 6 weeks after collection, the 16 native human serum samples were used to produce the sRM, consisting of a calibrator for the low range and for the middle range, and to prepare sample sets for each participating method, as previously described. [1] Each sample set consisted of i) 16 individual samples with low (n=5), middle (n=6) and high (n=5) hepcidin concentrations to assess commutability of the sRM, ii) mixtures of the individual samples to prepare a linearity panel to assess analytical performance of each participating method and iii) the two-leveled sRM made of the individual samples with low and middle hepcidin concentrations (Supplemental Table 1). The individual collected serum samples were transported on dry ice to the Queen Beatrix Hospital (Winterswijk, the Netherlands), where sample sets were prepared on the same day and stored at -80°C until shipment. Reference material vials were placed in -70°C and lyophilized after three days (Supplemental Table 2). Lyophilized material was stored at 2-8°C and shipped with the sample sets on dry ice to the participating laboratories within a week after preparation, where the sample sets arrived in frozen condition. All labs reported results within 3 weeks after receipt of the samples.

Samples RR2

To test the sRM harmonization potential in practice heparin plasma samples were needed with low, middle and high hepcidin concentrations. To this end, we collected one extra vial of blood from a phlebotomy of an iron-depleted hereditary hemochromatosis patient to produce a low hepcidin concentration sample for each MP (1.8 nM). A pool with a high hepcidin concentration was obtained from leftover heparin plasma samples of intensive care patients, once used for diagnostics, to produce a high hepcidin concentration sample for each MP (10.9 nM). By mixing both plasma with a low hepcidin concentration and the high hepcidin concentration plasma pool in a 50/50 ratio, we prepared a pool with a middle level

hepcidin concentration from which we produced a middle hepcidin concentration sample for each MP (5.7 nM). All together we prepared a sample set for each participating MP in RR2 consisting of 6 native samples (i.e. a low-, middle- and high hepcidin concentration sample in duplicate) and the sRM. As in RR1, hepcidin concentrations were measured by MS-1 (Table 1) and samples were stored at -80°C until sample set preparation.

Statistics

Exclusion of outlier measurements

Outlier measurements, assessed as previously described [1], were left out of analyses, restricted to 2 per method to minimize data manipulation (see Supplemental Table 5 and 8).

Analytical performance of measurement procedures

In both RR1 and RR2, reproducibility of the MPs was determined as the intra-assay coefficient of variation (CV), based on the logarithmically transformed (LN) results of triplicate measurements of the individual samples (RR1; 16, RR2; 3). Since standard deviations showed to be approximately proportional to the hepcidin level, logarithmic transformation of hepcidin concentrations was applied to stabilize standard deviations along the concentration range. The intra-assay CV of each MP was then calculated as $CV = 100\% \times \sqrt{\exp(\sigma^2) - 1}$ where σ is the standard deviation of the LN-transformed results and \exp is e^x . In addition, the intra-assay CV of each MP was calculated separately for the LN-transformed results of the low, middle and high hepcidin concentrations. Subsequently, we calculated the average of these values to compare the intra-assay variation at the three different hepcidin concentration levels globally.

In RR1, mutual correlations were calculated as the Spearman rank coefficient between each pair wise combination of MP, based on the 16 individual samples, and linearity was assessed from the linearity panel results plotted against the mixture ratios, expressed as the intercept a , slope b and Pearson correlation coefficient (r), as previously described. [1]

Harmonization with the secondary reference material

In RR1, the effect of harmonization on the equivalence between methods using the sRM was mathematically simulated by value reassignment based on the regression of the results of the sRM samples per method against the respective mean result of all methods. Assuming linearity, warranted by the analyses of linearity of each measurement procedure in RR1, linear regression equations were calculated for plots presenting the mean results of the calibrator samples for all methods (x-axis) and the respective results for these samples of each individual method (y-axis). Using these equations with the actual measured result of each of the individual samples before harmonization (Calibrated value= (measured value- intercept)/slope) simulates calibration with the produced secondary reference material, yielding the virtually harmonized results of these individual samples for all methods. The dispersion in these simulated results was then expressed as the inter-assay CV (%) after harmonization with the sRM, by using $CV = 100\% \times \sqrt{\exp(\sigma^2) - 1}$ where σ is the standard deviation of the LN-transformed means of the virtually harmonized results and \exp is e^x . At this point, the value of the two-leveled sRM was set as the mean measured by participating MPs of RR1.

In RR2, the effect of harmonization on the equivalence between methods using the sRM was tested in practice. To this end participants were allowed to use either prospective and

retrospective calibration methods. Participants that used the prospective calibration method in RR2, measured the first set of samples with their original calibration method. After recalibration with the sRM, they measured the second set of samples. Participants that used the retrospective method, measured the two-leveled sRM and individual samples in one run. The values obtained for the sRM were then used to create a regression analysis using the consensus value assigned to the sRM after RR1. All results measured in triplicate with their original assay were then recalculated using this regression analysis.

Primary reference material

Candidate pRM synthetic human hepcidin-25 peptide (amino acid sequence DTHFPICIFCCGCCHRSKCGMCCKT) was purchased from Anaspec (Seraing, Belgium). Hepcidin amount per vial was quantified by amino acid analysis using isotope-dilution, liquid-chromatography, mass spectrometry (ID-LC/MS). Briefly, the content of a vial was gravimetrically reconstituted in 500 μ L of an aqueous solution of 5% acetonitrile, 0.1% formic acid and subsamples were mixed with isotopically labelled Phe, Ile and Pro (Cambridge Isotope Laboratories, Inc.), as internal standards. After evaporation of the solvent, hepcidin was hydrolyzed in a vacuum sealed vessel in presence of HCl in the vapor phase. Hydrolysis proceeded for 40h at 130°C. Samples were reconstituted in water before analysis. Calibration was performed using solutions of certified reference materials for Phe (NMIJ CRM 6014a), Ile (NMIJ CRM 6013-a) and Pro (NMIJ CRM 6016a). These reference materials are pure amino acid substances of certified purity, whose mass fraction values are traceable to the SI. Peak area ratios of natural : labeled amino acids were plotted against their mass ratios to produce the calibration function. The concentration of internal standard

was similar in the certified reference materials and sample, and the molar ratio of the internal standard to sample was adjusted to be close to 1.

Stability testing was performed by comparing impurity profiles, measured by ID-LC/MS, over an 8-month period. The variation of each impurity mass fraction in time was subject to regression analysis. The t_{ratio} , calculated as the slope of the regression line divided by its uncertainty, was compared with the critical value of a two tailed t-distribution for a 0.01 significance level (t_{critical}). The alternative hypothesis that the slope was significantly different from zero was rejected when $t_{\text{ratio}} < t_{\text{critical}}$, indicating the absence of a significant trend and therefore stability.

Value assignment of the sRM

Value assignment of the sRM was performed in two parts using a validated Weak-Cation-exchange MALDI-Time of Flight – mass spectrometry (WCX MALDI-TOF MS) assay as candidate RMP (MS-1) [4].

In the first part, we reassigned the value of the internal standard of MS-1 (heavy hepcidin⁺⁴⁰, Peptide Inst. (JP)) using the certified pRM. To this end, the internal standard (Lot No. 221-006221) was reconstituted according to the manual of Peptide Institute. The pRM, synthetic hepcidin-25, was reconstituted according to instructions provided by LNE (France). In every run, we added 10 μL of the internal standard (100 μM) and 20 μL of the pRM (52.1 μM) gravimetrically to 10 mL 20% acetonitril as first dilution step. Secondly, in quadruplicate, we added 10 μL diluted material to 50 μL blanc plasma, from an iron-depleted juvenile hemochromatosis patient for dilution to nM measurement range. We used blanc plasma to

avoid oxidation of hepcidin in both the internal standard and the pRM. This was followed by the regular WCX MALDI-TOF MS protocol. In total 8 different runs were performed for value reassignment of the internal standard using pRM.

In the second part, we used this adjusted heavy hepcidin internal standard to assign a value to the sRM. Both the low and middle concentrations calibrator, consisting of individual serum samples, were placed at room temperature for 1 hour prior to reconstitution, which was done gravimetrically in 300 μ L deionized water. In every run, we diluted the internal standard (0.1 mM) gravimetrically by addition of 10 μ L to 10 mL 20% acetonitril. Secondly, we added 10 μ L diluted internal standard to 50 μ L sRM gravimetrically and in quadruplicate. This was followed by the regular WCX MALDI-TOF MS protocol. In total 4 different runs were performed by MS-1 as candidate RMP to assign a value to the sRM.

Statistical analysis was done using ANOVA.

Supplemental Table 1. Overview of sample set RR1.

Sample Type	Sample nr.	Description	Hepcidin concentration (nmol/L) ^e
Individual samples ^a	1	Subject 1	0.6
	2	Subject 2	1.1
	3	Subject 3	1.2
	4	Subject 4	1.1
	5	Subject 5	1.1
	6	Subject 6	2.6
	7	Subject 7	7.9
	8	Subject 8	4.4
	9	Subject 9	6.3
	10	Subject 10	2.7
	11	Subject 11	3.0
	12	Subject 12	12.8
	13	Subject 13	8.5
	14	Subject 14	11.3
	15	Subject 15	9.7
	16	Subject 16	8.0
Mixtures ^b	17	M1/L1 (sample nr 1-4)	
	18	M2 (sample nr 5-8)	
	19	M3 (sample nr 9-12)	
	20	M4/L5 (sample nr 13-16)	
	21	M5 (sample nr 1-16)	
Linearity panel ^c	22	L2 (75%/25% M1/M4)	
	23	L3 (50%/50% M1/M4)	
	24	L4 (25%/75% M1/M4)	
Reference material ^d	25	Lyophilized serum with CLP. Low: pooled sample nr 1-5	
	26	Lyophilized serum with CLP. Middle: pooled sample nr 6-11	

CLP, cryolyoprotectant; M, mixture; L, linearity panel sample

- a. 16 native human serum samples from either hemochromatosis patients or healthy volunteers: 5 low, 6 middle and 5 high hepcidin concentrations samples.

- b. Each mixture consists of 4 individual samples, except for M5 which consists of all samples. The individual samples each mixture consists of are described between the brackets.
- c. The samples of the linearity panel consist of mixture samples M1 and M4 in different ratios. L1 has a ratio of 100/0 of M1/M4 and is therefore equal to M1. The same applies to L5, in which the ratio 0/100 of M1/M4 is used, and therefore equals M4.
- d. The two-leveled reference material consists of 2 vials; a calibrator for the lower concentration range, made of the 5 samples with low hepcidin concentrations, and for the middle range, made of the 6 samples with middle hepcidin concentrations.
- e. Determined with MS-1. [4]

Supplemental Table 2. Lyophilization procedure of the secondary reference material.

Phase	Nr	Temperature in °C	Time (min)
Freeze temperature (FT)	01	-40	60
Main drying (MD)	01	-40	120
	02	-30	60
	03	-30	1080
	04	20	2100
	05	20	60
Second Drying (SD)	01	20	60
	02	20	240
Total duration in hours/days			63 h/2.6 d

Supplemental Table 3. Analytical characteristics of hepcidin MPs participating in RR1, in terms of spearman rank correlation coefficients between the MPs of the 16 individual samples.

	MS-1	MS-3	MS-4	MS-5	IC-1	IC-3	IC-4	IC-5	IC-6
MS-1	1.000								
MS-3	0.993	1.000							
MS-4	0.978	0.984	1.000						
MS-5	0.979	0.994	0.991	1.000					
IC-1	0.990	0.985	0.972	0.976	1.000				
IC-3	0.987	0.985	0.960	0.970	0.988	1.000			
IC-4	0.985	0.996	0.988	0.996	0.976	0.974	1.000		
IC-5	0.958	0.964	0.935	0.958	0.968	0.979	0.950	1.000	
IC-6	0.989	0.993	0.981	0.990	0.991	0.982	0.985	0.974	1.000

MP, measurement procedure; MS, mass spectrometry-based MP; IC, immunochemical-based MP.

Supplemental Table 4. Analytical characteristics of hepcidin MPs participating in RR1, in terms of reproducibility and linearity.

MP	Reproducibility		Regression relationship compared to all MPs mean		
	Intra-assay CV, % ^a	r ^b	r ^c	Slope (95% CI)	Intercept (95% CI)
MS-1	5.4	0.996	0.997	0.631 (0.605-0.657)	-0.530 (-0.787- -0.273)
MS-3	3.0	0.999	0.998	0.958 (0.921-0.995)	0.220 (-0.144-0.584)
MS-4	4.2	0.987	0.981	0.476 (0.413-0.538)	2.103 (1.421-2.784)
MS-5	2.7	0.995	0.996	0.688 (0.655-0.722)	0.238 (-0.091-0.568)
IC-1	6.3	0.990	0.995	0.781 (0.736-0.826)	0.226 (-0.220-0.671)
IC-3	7.8	0.989	0.997	1.970 (1.874-2.067)	-0.314 (-1.292-0.665)
IC-4	5.8	0.993	0.996	1.595 (1.510-1.680)	-1.018 (-1.852- -0.183)
IC-5	5.7	0.994	0.985	1.084 (0.975-1.192)	-0.243 (-1.309-0.824)
IC-6	2.8	0.996	0.993	0.793 (0.741-0.846)	-0.386 (-0.904-0.132)

MP, measurement procedure; MS, mass spectrometry-based MP; IC, immunochemical-based MP; CV, coefficient of variation.

- Reproducibility, expressed as the intra-assay CV, calculated from the SD of the logarithmically transformed results of the 16 individual samples measured in triplicate (Supplemental Table 5).
- Linearity is assessed by the linearity panel samples (Supplemental Table 1), expressed as the Pearson correlation coefficient r.
- Pearson correlation coefficient r of the regression relationship between results of each individual measurement procedure with the all measurement procedures mean.

Supplemental Table 5. Raw results of triplicate measurements of all samples for RR1 obtained by the participating MPs.

			MP																	
			IC-1		IC-3		IC-4		IC-5		IC-6		MS-1		MS-3		MS-4		MS-5	
			Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Sample	MM	SD																		
1	1.37	0.59	1.13	0.15	1.47	0.23	1.77	0.25	1.13	0.15	1.07	0.06	0.57	0.06	1.37	0.06	2.67	0.15	1.13	0.06
2	2.45	0.90	2.23	0.31	4.07	0.55	3.03	0.15	2.80	0.26	1.67	0.06	1.10	0.10	2.60	0.10	ND	ND	2.10	0.10
3	3.02	1.27	2.63	0.06	5.50	0.60	3.77	0.06	3.07	0.12	1.77	0.06	1.27	0.06	2.87	0.06	3.97	0.06	2.33	0.06
4	2.66	1.11	2.27	0.15	4.83	0.25	3.43	0.21	2.53	0.06	1.63	0.06	1.30	0.10	2.67	0.15	3.50	0.44	1.77	0.06
5	2.33	1.03	1.90	0.17	4.47	0.83	2.73	0.21	2.57	0.06	1.40	0.00	1.10	0.00	2.13	0.06	3.03	0.21	1.60	0.00
6	4.52	1.52	3.57	0.06	7.00	0.53	6.77	0.55	4.40	0.00	3.57	0.06	2.37	0.12	5.00	0.17	4.23	0.38	3.77	0.06
7	11.40	5.40	8.00	0.00	23.17	1.19	16.27	0.21	12.90	0.17	8.30	0.30	6.47	0.21	11.50	0.36	7.50	0.10	8.53	0.21
8	8.77	4.22	7.30	0.44	17.57	1.36	13.40	0.26	9.77	0.40	6.90	0.10	4.53	0.25	8.03	0.21	5.47	0.12	5.97	0.15
9	10.91	5.17	8.57	0.47	22.27	0.12	16.10	0.36	10.93	0.15	7.57	0.12	6.33	0.35	10.93	0.40	7.70	0.26	7.80	0.10
10	5.04	2.15	4.27	0.12	9.97	0.67	6.27	0.06	5.47	0.25	3.67	0.15	2.43	0.15	5.00	0.10	4.53	0.12	3.77	0.12
11	5.81	2.41	5.00	0.17	10.73	0.21	8.43	0.15	5.83	0.21	4.10	0.10	2.90	0.10	6.23	0.12	4.60	0.26	4.50	0.00
12	18.49	8.48	14.70	0.30	35.07	1.42	29.40	1.95	20.17	2.06	15.53	0.31	10.93	0.31	17.20	0.10	10.23	0.38	13.17	0.32
13	11.99	5.17	9.80	0.10	23.07	0.59	17.87	0.93	11.33	0.12	9.23	0.15	7.10	0.17	11.80	0.10	8.80	0.20	8.87	0.12
14	17.17	7.76	13.10	0.40	32.43	0.76	26.60	0.87	20.03	0.06	12.23	0.21	10.50	0.00	16.77	0.21	10.77	0.06	12.13	0.15
15	14.08	6.61	11.90	0.66	27.60	0.36	22.63	0.85	12.13	0.12	11.13	0.15	8.97	0.25	14.20	0.30	8.80	0.20	9.33	0.21
16	11.62	5.56	10.07	0.21	24.23	1.47	15.20	0.72	13.67	1.10	8.50	0.20	6.67	0.31	11.33	0.35	7.23	0.21	7.67	0.25
17	2.25	0.67	2.00	0.17	3.23	0.25	2.73	0.21	2.60	0.00	1.67	0.06	1.03	0.06	2.50	0.10	2.67	0.15	1.83	0.06
18	6.46	2.71	5.73	0.29	12.47	1.63	9.00	0.46	5.73	0.06	5.10	0.20	3.27	0.06	6.40	0.17	5.40	0.56	5.03	0.12
19	9.22	3.92	6.63	0.25	17.00	1.35	12.87	0.55	11.67	0.21	6.87	0.23	5.37	0.23	9.50	0.35	5.70	0.10	7.37	0.46
20	13.13	6.11	10.83	0.40	27.00	2.10	17.63	0.15	14.63	0.21	10.10	0.20	8.50	0.10	12.97	0.15	7.83	0.25	8.67	0.15
21	7.60	3.03	6.47	0.25	13.70	1.47	10.77	0.58	8.73	0.31	5.30	0.26	4.30	0.17	7.83	0.06	5.77	0.15	5.53	0.06
22	5.14	2.04	4.97	0.15	9.63	0.40	6.07	0.15	6.00	0.10	3.77	0.06	2.43	0.06	5.17	0.25	4.37	0.31	3.83	0.06

			MP																	
			IC-1		IC-3		IC-4		IC-5		IC-6		MS-1		MS-3		MS-4		MS-5	
			Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
23	7.76	3.31	6.23	0.06	14.50	0.98	11.67	0.25	8.00	0.10	6.00	0.35	4.33	0.35	7.97	0.25	5.33	0.31	5.80	0.00
24	10.33	4.17	8.37	0.59	19.30	1.47	14.47	0.31	10.83	0.49	8.23	0.61	6.67	0.06	10.67	0.12	7.10	0.10	7.37	0.25
25	2.38	1.22	1.83	0.06	5.03	2.16	3.13	0.25	2.17	0.21	1.47	0.06	0.90	0.00	2.20	0.17	3.00	0.10	1.70	0.10
26	7.03	3.15	5.33	0.21	13.20	0.61	10.80	0.66	8.07	0.21	4.83	0.21	3.80	0.10	7.07	0.23	5.03	0.21	5.10	0.17

MM, mean of all methods; SD; standard deviation; MP, measurement procedure; MS, mass spectrometry-based MP; IC, immunochemical-based MP; ND, not determined; RR, round robin.

Sample number 1-16 represent the individual native human serum samples, as sample number 17-24 represent the mixture and linearity samples and sample number 25-26 the calibrator samples (Table 1 provides information regarding the design of the samples). The shaded results were identified as outliers and were thus left out of the analyses. These samples were identified by normalizing the mean result of each sample for every measurement procedure against the mean result of all methods and expressed as the ratio mean/selected measurement procedures' mean $\frac{M}{SMM}$. Subsequently the relative deviation (%) for each of the samples from the mean normalized results for that specific measurement procedure (calculated as $\frac{M/SMM}{\text{mean } M/SMM} \times 100\%$) was determined. The number of outlier samples or measurement errors, defined as having a ratio below 60% or above 140%, identified in this manner and left out of the analyses was restricted to 2 per MP to prevent data manipulation. MP MS-2 and IC-2 are missing as they were not ready to measure the samples at the time of sample send-out.

Supplemental Table 6. Relative contributions of the assessed analytical characteristics of the 9 participating MPs in RR1 to the equivalence without harmonization.

Analytical characteristic of the method	Mean CV (%)	Contribution to the equivalence (%) ^f
Imprecision ^a	4.8	1.3
Non-linearity ^b	5.9	2.0
Differences in calibration ^c	40.5	92.4
Heterogeneity ^d	8.7	4.3
Equivalence without harmonization ^e	42.1	100.0

CV, coefficient of variation.

- Mean CV of the imprecision was calculated as the arithmetic mean of the intra-assay CVs.
- Mean CV of non-linearity was calculated as the arithmetic mean of the CVs for linearity, which represent the deviation of each individual linearity panel sample from the regression line with correlation coefficient r . By applying the sum of squares of these deviations, we could calculate the concerned CVs.
- Contribution of differences in calibration (CV_C) were calculated from the other CVs for equivalence without harmonization (CV_T), imprecision (CV_R), non-linearity (CV_L) and heterogeneity (CV_H) using the formula $CV_C = \sqrt{(CV_T)^2 - \{(CV_R)^2 + (CV_L)^2 + (CV_H)^2\}}$.
- The contribution of heterogeneity represents the CV that remains after harmonization, i.e. the achievable equivalence, by calculation of the inter-assay CV after value reassignment of the 16 individual samples using the mean of all methods.
- The degree of equivalence without harmonization is expressed as the inter-assay CV.

- f. The relative contributions can be calculated from the squared CVs of each characteristic as a percentage of the squared CV of the current equivalence.

Supplemental Table 7. Reproducibility of MPs participating in RR2.

MP	Intra-assay CV
MS-1	4.5 %
MS-2	5.4 %
MS-3	4.0 %
MS-4	5.3 %
MS-5	2.7 %
MS-6	4.7 %
MS-7	3.3 %
MS-8	1.4 %
IC-3	5.0 %
IC-4	10.9 %
IC-5	9.2 %

MP, measurement procedure; CV, coefficient of variation; RR, round robin.

The presented intra-assay CVs are calculated using the mean of the duplicate measurement of the 3 individual sample CVs.

Supplemental Table 8. Raw results of triplicate measurements of all samples for RR2 obtained by the participating MPs.

			MP																					
			IC-3		IC-4		IC-5		MS-1		MS-2		MS-3		MS-4		MS-5		MS-6		MS-7		MS-8	
			Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Sample	MM	SD																						
1	3.32	2.04	8.23	0.60	4.87	0.67	3.70	0.30	1.83	0.12	1.73	0.15	3.20	0.20	2.20	0.10	1.33	0.06	2.47	0.06	2.03	0.06	4.90	0.10
2	9.26	4.69	20.43	0.78	13.13	1.32	11.53	1.25	5.67	0.06	9.87	0.70	8.73	0.25	5.80	0.10	4.57	0.06	6.87	0.21	4.77	0.23	10.50	0.10
3	16.19	7.97	34.87	0.91	22.97	1.59	20.83	1.60	10.97	0.46	17.63	0.47	14.87	0.12	11.87	0.91	8.40	0.10	11.90	0.85	6.77	0.06	17.03	0.15

MM, mean of all methods; SD, standard deviation; MP, measurement procedure; MS, mass spectrometry-based MP; IC, immunochemical-based MP; RR, round robin.

The shaded results were identified as outliers and were thus left out of the analyses. These samples were identified by normalizing the mean result of each sample for every measurement procedure against the mean result of all methods and expressed as the ratio mean/selected measurement procedures' mean $\frac{M}{SMM}$. Subsequently the relative deviation (%) for each of the samples from the mean normalized results for that specific measurement procedure (calculated as $\frac{M/SMM}{\text{mean } M/SMM} \times 100\%$) was determined. The number of outlier samples or measurement errors, defined as having a ratio below 60% or above 140%, identified in this manner and left out of the analyses was restricted to 2 per MP to prevent data manipulation.

Supplemental Table 9. Amino acid analysis results and final mass fraction assignment of hepcidin pRM.

Hepcidin mass fraction in reconstituted vial ($\mu\text{g/g}$)				
		Quantification amino acid		
Day	Rep.	Phe	Ile	Pro
1	1	148	159	156
	2	155	159	159
	3	158	165	162
2	1	154	154	153
	2	157	159	156
	3	163	167	159
3	1	159	158	156
	2	154	161	172
	3	140	142	163
Single mass fraction ($\mu\text{g/g}$)		154	158	160
Combined mass fraction ($\mu\text{g/g}$) ^a		157.3 \pm 5.6		
Impurity-corrected mass fraction ($\mu\text{g/g}$) ^a		148.5 \pm 6.3		
Hepcidin mass quantity per vial (μg) ^b		72.9 \pm 3.9		

a. Mass fraction and associated expanded uncertainty (k=2).

b. Mass quantity of hepcidin per vial and calculated expanded uncertainty (k=2) considering amino acid analysis, impurity correction and material heterogeneity.

Supplemental Table 10. Identified peptidic impurities and their relative amount to hepcidin-25 by high resolution mass spectrometry analysis.

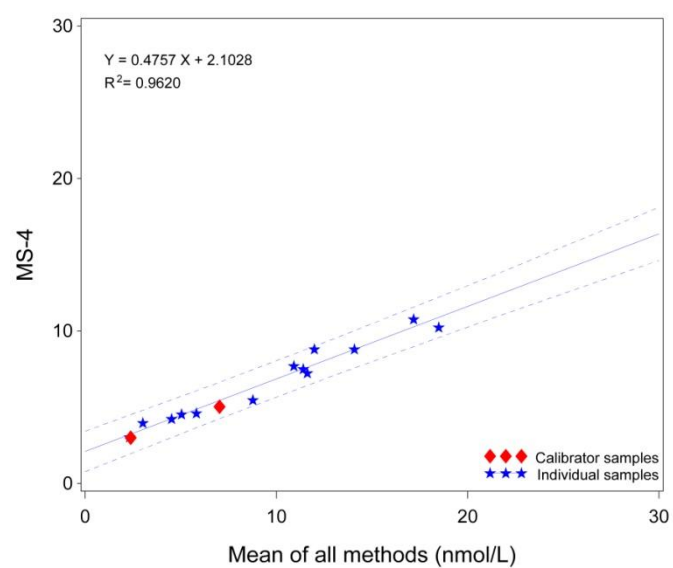
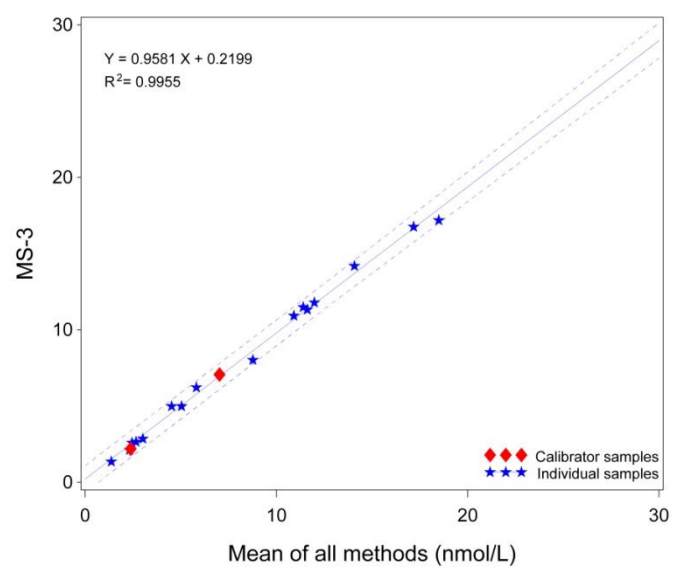
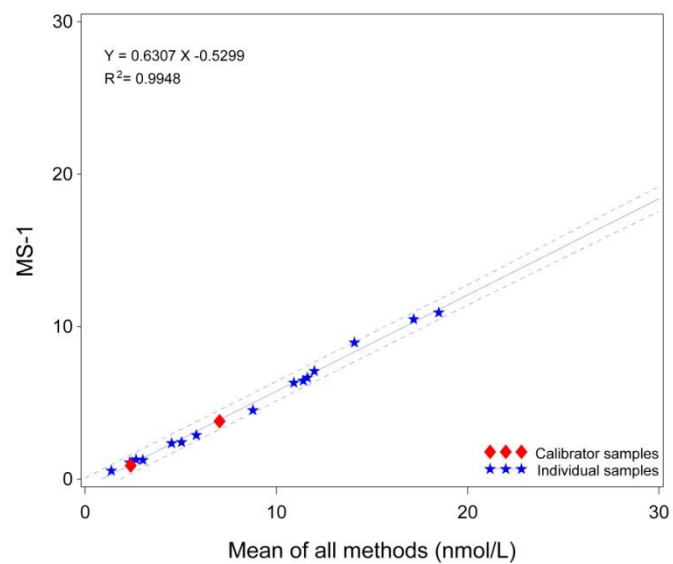
Peptide impurity	RT (min)	Monoisotopic mass (Da)	Amount relative to hepcidin 1-25 (%)
H1-25 (Thr-2 O-sulfonation)	19.21	2866.98	0.5%
H1-25(+150.1)	22.02	2937.13	0.6%
H1-24	17.88	2685.98	0.6%
H2-25	16.84	2672.00	0.9%
H1-25 (oxd Met-21)	15.15	2803.02	1.3%
H3-25	17.05	2570.95	0.3%
H1-25 (Gly20->Glu)	16.79	2859.05	1.0%
H5-25 (N-term acetylated)	16.99	2911.04	0.3%
Total impurity content (%)			5.6%
Standard uncertainty (%)			1.1%

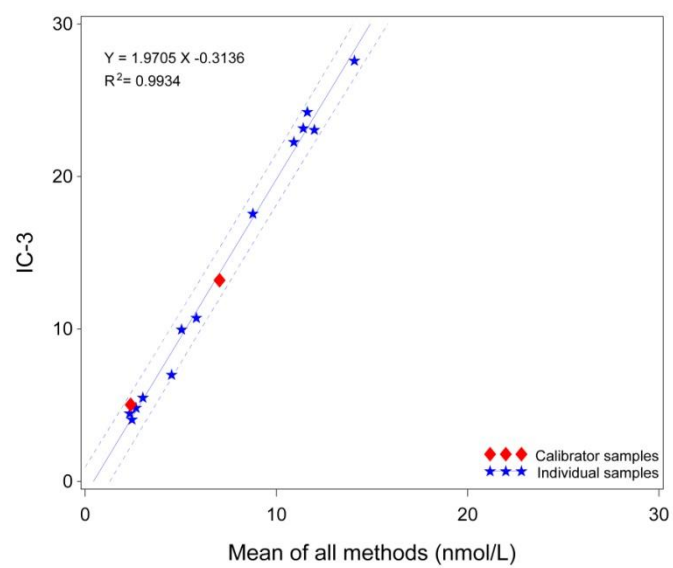
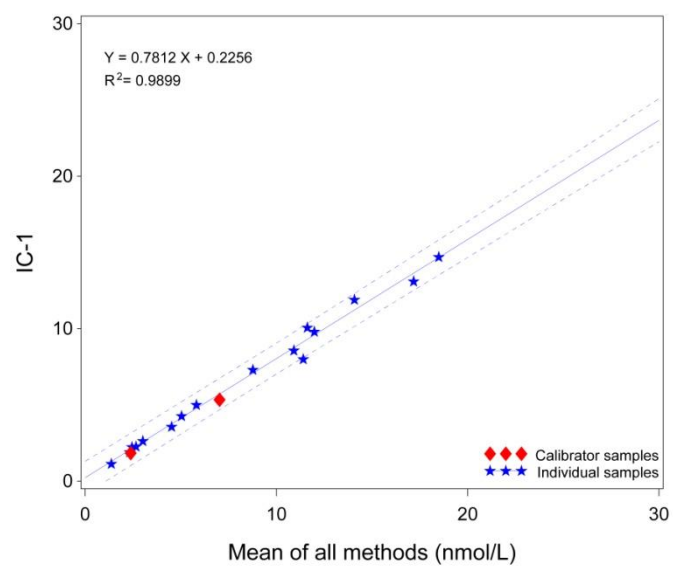
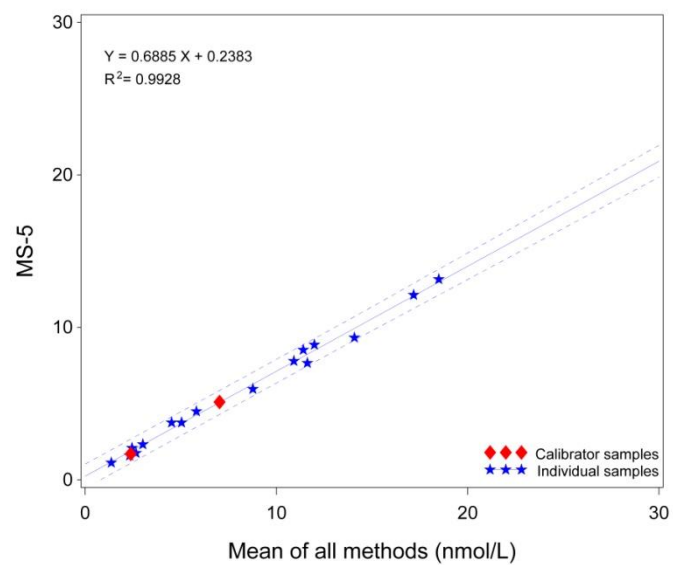
RT, retention time.

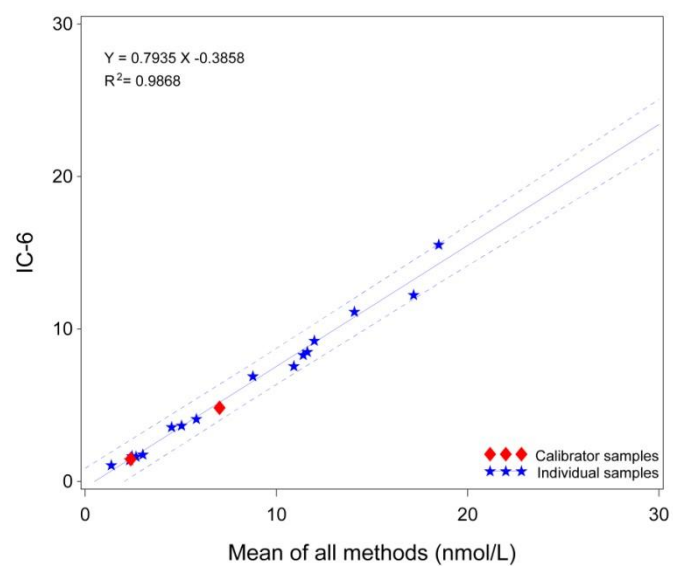
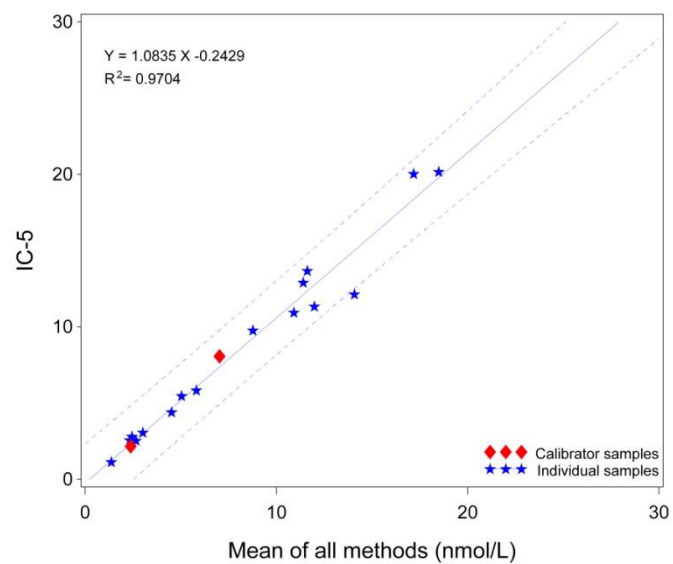
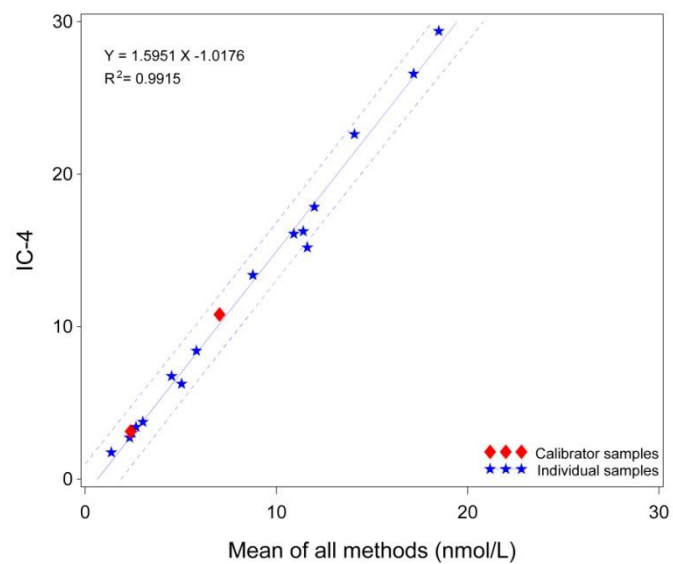
All major identified impurities are structurally related to hepcidin-25 (N- and C- terminal truncated forms and modified forms, e.g. oxidation). Therefore H1-25 represents the whole hepcidin peptide, with modifications explained between brackets, H1-24 represents hepcidin with one amino acid missing at the C-terminal, H3-25 is missing two amino acids at the N-terminal and H5-25 missess four amino acids at the N-terminal. [5]

Supplemental Figure 1. Commutability plots of each MP of RR1 against the mean of all MPs.

The mean of the triplicate measurements of the 16 individual samples (blue stars) and calibrator samples (red diamond) of each MP (y-axis) are plotted against the corresponding mean results of all methods (x-axis). The regression line (blue) is presented with its 95% prediction interval of the individual samples (blue dotted lines).







References

1. 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.
2. Weykamp C, Eckfeld J, Vesper H, Thienpont L, Burns C, Caliendo A, et al. Aacc harmonization initiative; toolbox of technical procedures to achieve harmonization for a measurand 2012. Available at: www.harmonization.net. Accessed: 10 Sept 2018.
3. 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*. 2011;117:e218-25.
4. 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.
5. Bros P, Josephs RD, Stoppacher N, Cazals G, Lehmann S, Hirtz C, et al. Impurity determination for hepcidin by liquid chromatography-high resolution and ion mobility mass spectrometry for the value assignment of candidate primary calibrators. *Anal Bioanal Chem*. 2017;409:2559-67.