Results of the first international round robin for the quantification of urinary and plasma hepcidin assays: need for standardization

Joyce J.C. Kroot,1 Erwin H.J.M. Kemna,1 Sukhvinder S. Bansal,7 Mark Busbridge,7 Natascia Campostrini,4 Domenico Girelli,4 Robert C. Hider,2 Vasiliki Koliaraki,6 Avgi Mamalaki,5 Gordana Olbina,6 Naohisa Tomosugi,7 Chris Tselepis,6 Douglas G. Ward,7 Tomas Ganz,6,9 Jan C.M. Hendriks,10 and Dorine W. Swinkels1

1Department of Clinical Chemistry, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 2Pharmaceutical Sciences Division, King’s College London, London, UK; 3Department of Clinical Chemistry, Imperial College HealthCare NHS Trust, Hammersmith Hospital Campus, London, UK; 4Department of Clinical and Experimental Medicine, Section of Internal Medicine, University of Verona, Verona, Italy; 5Laboratory of Molecular Biology & Immunobiotechnology, Department of Biochemistry, Hellenic Pasteur Institute, Athens, Greece; 6Intrinsic Life Sciences, La Jolla, CA, USA; 7Division of Advanced Medicine, Medical Research Institute/Division of Nephrology, Kanazawa Medical University, Ishikawa, Japan; 8CRUK Institute for Cancer Studies, University of Birmingham, Birmingham, UK; 9Department of Medicine, David Geffen School of Medicine at UCLA, LA, USA, and 10Department of Epidemiology and Biostatistics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands

ABSTRACT

The recently discovered iron regulatory peptide hormone hepcidin holds promise as a novel biomarker in iron metabolism disorders. To date, various mass spectrometry and immunochemical methods have been developed for its quantification in plasma and urine. Differences in methodology and analytical performance hinder the comparability of data. As a first step towards method harmonization, several hepcidin assays were compared. Worldwide eight laboratories participated in a urinary and plasma round robin in which hepcidin was analyzed. For both urine and plasma: (i) the absolute hepcidin concentrations differed widely between methods, (ii) the between-sample variation and the analytical variation of the methods are similar. Importantly, the analytical variation as percentage of the total variance is low for all methods, indicating their suitability to distinguish hepcidin levels of different samples. Spearman correlations between methods were generally high. The round robin results inform the scientific and medical community on the status and agreement of the current hepcidin methods. Ongoing initiatives should facilitate standardization by exchanging calibrators and representative samples.

Key words: hepcidin, iron, quality control.


©2009 Ferrata Storti Foundation. This is an open-access paper.

Introduction

Hepcidin plays a central role in iron metabolism, and could become a useful biomarker for the diagnosis and monitoring of iron disorders.10 Progress in human studies of hepcidin in normal physiology and various disease states has been hampered by the limited availability of hepcidin assays. Assays have been developed on mass spectrometry platforms including surface enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF MS), matrix assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) and liquid chromatography tandem-MS techniques (LC-MS/MS). Some methods use an internal standard, either hepcidin analogs or bioactive hepcidin-25 synthesized with stable isotopes.9,10 Recently, immunochemical (IC) assays for hepcidin-25 have also been developed, which comprise of competitive radio-immunoassays (RIA)11 and enzyme-linked immunosorbent assays (ELISA).11,12 Currently there is no reference method for hepcidin measurements. Therefore, to increase comparability of hepcidin data across clinical studies we evaluated the levels, between-sample variation and the analytical variation of hepcidin assays on a panel of urine and plasma samples in a so-called Round Robin.14

Design and Methods

Study design and participants

A prospective repeated measurement design with 12 replicates was used to assess concordance in urine and plasma hepcidin
analysis. Six and eight laboratories participated in the urinary and plasma analyses, respectively, in six different countries. The study was coordinated by the Department of Clinical Chemistry of the Radboud University Nijmegen Medical Centre. All laboratories performed 12 replicates, that consisted of triplicate assays of each sample on four consecutive days. The only information provided about the samples were urinary creatinine levels.

**Specimens**

Eight urine samples with a wide range of hepcidin concentrations were collected from healthy subjects (samples 1-5) and patients (samples 6-8) (October 2007) with informed consent according to the declaration of Helsinki. Synthetic hepcidin-25 (Peptide Institute Inc., Osaka, Japan) was used for external mass calibration. ProteinChip; Bio-Rad Laboratories). Synthetic 25-hepcidin directly applied to hydrophilic Nor mal Phase chips (NP20 Clinprot beads (Bruker Daltonik) and analyzed by MALDI-TOF MS. Plasma samples were assayed by SELDI-TOF-MS using Cu²⁺ loaded IMAC chips. Synthetic human hepcidin-25 (Peptide Institute Inc.) was used as an external standard in both assays.

Method IV is based on LC-MS/MS. [15N,13C2]Gly12,20-hepcidin (heavy hepcidin) was added to plasma or urine as internal standard. Magnetic nanoparticles (Bruker Daltonik) were used to extract hepcidin from the samples followed by LC-MS/MS analysis using selected reaction monitoring of the triple charged precursor fragmenting to the double charged product ion for both heavy hepcidin and hepcidin.

Method V is a competitive ELISA. 96-well plates were coated with in house-prepared anti-human hepcidin antibody and biotinylated hepcidin-25 as tracer. Custom synthesized hepcidin-25 was used as an external standard. Method VI is a competitive RIA using a 125I labeled synthetic hepcidin-25 (Bachem) with an in-house rabbit anti-hepcidin polyclonal antibody (against hepcidin-KLH conjugate) using a secondary antibody-PEG assisted separation. Synthetic hepcidin-25 (Bachem) was used as an external standard.

Method VII is based on LC-MS/MS. Isotopic human synthetic hepcidin-25 (Peptide Institute Inc.) was used as an internal standard.

Method VIII is a competitive ELISA; plates were coated with an in -house prepared recombinant hepcidin-25-His peptide and a polyclonal antibody against recombinant (in-house prepared) hepcidin-25-His was used to establish competition. Horseradish peroxidase labeled anti-rabbit antibody was used as secondary antibody. Hecpidin-25-His was used as an external standard.

**Hepcidin methods**

Characteristics of the methods used for the plasma and urine hepcidin measurements of the present study are schematically presented in Table 1. Method I is based on SELDI-TOF MS. Samples were directly applied to hydrophilic Normal Phase chips (NP20 ProteinChip; Bio-Rad Laboratories). Synthetic 25-hepcidin (Peptide Institute Inc.) was used for external mass calibration.

Method II is also based on SELDI-TOF MS. Hepcidin was first extracted from the sample using Macro-Prep® CM Support beads (Bio-Rad Laboratories). The extract was applied to NP20 chips (Bio-Rad Laboratories). Synthetic hepcidin-24 peptide (custom made, Peptide Institute Inc.) was used as an internal standard.

**Statistical methods**

The study was designed to compare hepcidin levels as well as the repeatability of the methods used for serum and urine, respectively. With respect to the repeatability, the magnitude of variation that exists between samples and between measurements of the same sample relative to

---

**Table 1. Characteristics of methods used for plasma and urine hepcidin measurements.**

<table>
<thead>
<tr>
<th>Method</th>
<th>Principle</th>
<th>Method</th>
<th>Hepcidin extraction</th>
<th>Standard</th>
<th>Urine</th>
<th>Plasma</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Mass spectrometry</td>
<td>SELDI-TOF MS</td>
<td>Normal phase</td>
<td>None</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>II</td>
<td>Mass spectrometry</td>
<td>SELDI-TOF MS</td>
<td>Weak cation exchange</td>
<td>Internal</td>
<td>Synthetic hepcidin-24</td>
<td>X</td>
</tr>
<tr>
<td>III</td>
<td>Mass spectrometry</td>
<td>MALDI-TOF MS</td>
<td>Reversed phase Immobilized metal affinity Chromatography</td>
<td>External</td>
<td>Synthetic hepcidin-25</td>
<td>X</td>
</tr>
<tr>
<td>IV</td>
<td>Mass spectrometry</td>
<td>LC-MS/MS</td>
<td>Weak cation exchange</td>
<td>Internal</td>
<td>[15N,13C2]Gly12,20-hepcidin</td>
<td>X</td>
</tr>
<tr>
<td>V</td>
<td>Immunochemical</td>
<td>Competitive ELISA</td>
<td>None</td>
<td>External</td>
<td>Synthetic hepcidin-25</td>
<td>X</td>
</tr>
<tr>
<td>VI</td>
<td>Immunochemical</td>
<td>Competitive RIA</td>
<td>None</td>
<td>External</td>
<td>Synthetic hepcidin-25</td>
<td>X</td>
</tr>
<tr>
<td>VII</td>
<td>Mass spectrometry</td>
<td>LC-MS/MS</td>
<td>None</td>
<td>Internal</td>
<td>Synthetic hepcidin-25</td>
<td>X</td>
</tr>
<tr>
<td>VIII</td>
<td>Immunochemical</td>
<td>Competitive ELISA</td>
<td>None</td>
<td>External</td>
<td>Recombinant hepcidin-25-His</td>
<td>X</td>
</tr>
</tbody>
</table>

* X: participation in hepcidin round robin for urine and/or plasma.
of including the latter two isoforms in the total variation are of interest. Accordingly, we partitioned the total variance of each hepcidin method into the following components: i) the between-sample variance and ii) the analytical variance. Our design allowed us to divide the latter into three subcomponents: the between-day variance, the between triplicate variance and the residual analytical variance, i.e. the part of the analytical variance that cannot be attributed to the other two.

A linear mixed model was used to estimate these variance components of each method separately. The dependent variable was hepcidin outcome, and the independent random variables were: sample (plasma: 7 levels, urine: 8 levels), day (4 levels) and repeated measurement (3 levels). We found that the estimated percentage variance due to the triplicate measurements was very small and considerably smaller than the variance due to the between-day variation. Consequently, this term was omitted from the final model and not presented separately. The SD (absolute error), the CV (relative error) and the percentage variance relative to the total variance of each random variable are presented and for each method separately.

Results and Discussion

Mean hepcidin levels for all (Table 2) samples differ considerably between all methods. Notably, hepcidin values of method V are relatively high, but trends and variability are similar to all other methods (Table 2). It is also of note that for this Round Robin study, we only used native urine and serum samples, except for one, i.e. urine sample 5 to which synthetic hepcidin was added to a final concentration of 13.1 nmol/mmol creatinine. We found that the mean hepcidin outcome of some methods (i.e. methods II, IV and VI) for this spiked urine sample are closer to this concentration than others. However, irrespective of whether the value assigned of the spiked hepcidin-25 by Peptide International is correct, these results can not simply be extrapolated to native urine or serum samples, since it is presently unclear whether the various methods evaluated behave differently for spiked and native samples. In other words, the so called commutability of the spiked samples with native clinical samples for the various methods is unknown and its assessment may be part of future standardization efforts for hepcidin measurements.15 25,27

In general, differences in hepcidin levels between methods might be due to: (i) the use of different calibration solutions with level assignments based on different techniques; (ii) to hepcidin aggregation of either the standard solution or the sample; or (iii) hepcidin binding to α2 macroglobuline or albumine26 or iv) the existence of three hepcidin isoforms hepcidin-25, 22 and 20. These four points may differentially affect IC and MS measurements, and urine and serum quantifications. More specifically, it was recently learned from the measurements of steroid and thyroid hormones. Another cause for differences between IC and MS methods is that IC methods lack the selectivity to distinguish hepcidin-25 from hepcidins-20 and -22. However, the implications of including the latter two isoforms in the

the total variation are of interest. Accordingly, we partitioned the total variance of each hepcidin method into the following components: i) the between-sample variance and ii) the analytical variance. Our design allowed us to divide the latter into three subcomponents: the between-day variance, the between triplicate variance and the residual analytical variance, i.e. the part of the analytical variance that cannot be attributed to the other two.

A linear mixed model was used to estimate these variance components of each method separately. The dependent variable was hepcidin outcome, and the independent random variables were: sample (plasma: 7 levels, urine: 8 levels), day (4 levels) and repeated measurement (3 levels). We found that the estimated percentage variance due to the triplicate measurements was very small and considerably smaller than the variance due to the between-day variation. Consequently, this term was omitted from the final model and not presented separately. The SD (absolute error), the CV (relative error) and the percentage variance relative to the total variance of each random variable are presented and for each method separately.

Results and Discussion

Mean hepcidin levels for all (Table 2) samples differ considerably between all methods. Notably, hepcidin values of method V are relatively high, but trends and variability are similar to all other methods (Table 2). It is also of note that for this Round Robin study, we only used native urine and serum samples, except for one, i.e. urine sample 5 to which synthetic hepcidin was added to a final concentration of 13.1 nmol/mmol creatinine. We found that the mean hepcidin outcome of some methods (i.e. methods II, IV and VI) for this spiked urine sample are closer to this concentration than others. However, irrespective of whether the value assigned of the spiked hepcidin-25 by Peptide International is correct, these results can not simply be extrapolated to native urine or serum samples, since it is presently unclear whether the various methods evaluated behave differently for spiked and native samples. In other words, the so called commutability of the spiked samples with native clinical samples for the various methods is unknown and its assessment may be part of future standardization efforts for hepcidin measurements.15

In general, differences in hepcidin levels between methods might be due to: (i) the use of different calibration solutions with level assignments based on different techniques; (ii) to hepcidin aggregation of either the standard solution or the sample; or (iii) hepcidin binding to α2 macroglobuline or albumine26 or iv) the existence of three hepcidin isoforms hepcidin-25, 22 and 20. These four points may differentially affect IC and MS measurements, and urine and serum quantifications. More specifically, it was recently found that around 90% of the circulating hepcidin is bound to α2 macroglobulin in the blood. These observations not only raise the question whether we should measure total, bound or unbound hepcidin, but also what the methods evaluated actually measure. The search for an answer to these questions is a new challenge for which much can be learned from the measurements of steroid and thyroid hormones. Another cause for differences between IC and MS methods is that IC methods lack the selectivity to distinguish hepcidin-25 from hepcidins-20 and -22. However, the implications of including the latter two isoforms in the

the total variation are of interest. Accordingly, we partitioned the total variance of each hepcidin method into the following components: i) the between-sample variance and ii) the analytical variance. Our design allowed us to divide the latter into three subcomponents: the between-day variance, the between triplicate variance and the residual analytical variance, i.e. the part of the analytical variance that cannot be attributed to the other two.

A linear mixed model was used to estimate these variance components of each method separately. The dependent variable was hepcidin outcome, and the independent random variables were: sample (plasma: 7 levels, urine: 8 levels), day (4 levels) and repeated measurement (3 levels). We found that the estimated percentage variance due to the triplicate measurements was very small and considerably smaller than the variance due to the between-day variation. Consequently, this term was omitted from the final model and not presented separately. The SD (absolute error), the CV (relative error) and the percentage variance relative to the total variance of each random variable are presented and for each method separately.

Results and Discussion

Mean hepcidin levels for all (Table 2) samples differ considerably between all methods. Notably, hepcidin values of method V are relatively high, but trends and variability are similar to all other methods (Table 2). It is also of note that for this Round Robin study, we only used native urine and serum samples, except for one, i.e. urine sample 5 to which synthetic hepcidin was added to a final concentration of 13.1 nmol/mmol creatinine. We found that the mean hepcidin outcome of some methods (i.e. methods II, IV and VI) for this spiked urine sample are closer to this concentration than others. However, irrespective of whether the value assigned of the spiked hepcidin-25 by Peptide International is correct, these results can not simply be extrapolated to native urine or serum samples, since it is presently unclear whether the various methods evaluated behave differently for spiked and native samples. In other words, the so called commutability of the spiked samples with native clinical samples for the various methods is unknown and its assessment may be part of future standardization efforts for hepcidin measurements.15

In general, differences in hepcidin levels between methods might be due to: (i) the use of different calibration solutions with level assignments based on different techniques; (ii) to hepcidin aggregation of either the standard solution or the sample; or (iii) hepcidin binding to α2 macroglobuline or albumine26 or iv) the existence of three hepcidin isoforms hepcidin-25, 22 and 20. These four points may differentially affect IC and MS measurements, and urine and serum quantifications. More specifically, it was recently found that around 90% of the circulating hepcidin is bound to α2 macroglobulin in the blood. These observations not only raise the question whether we should measure total, bound or unbound hepcidin, but also what the methods evaluated actually measure. The search for an answer to these questions is a new challenge for which much can be learned from the measurements of steroid and thyroid hormones. Another cause for differences between IC and MS methods is that IC methods lack the selectivity to distinguish hepcidin-25 from hepcidins-20 and -22. However, the implications of including the latter two isoforms in the

the total variation are of interest. Accordingly, we partitioned the total variance of each hepcidin method into the following components: i) the between-sample variance and ii) the analytical variance. Our design allowed us to divide the latter into three subcomponents: the between-day variance, the between triplicate variance and the residual analytical variance, i.e. the part of the analytical variance that cannot be attributed to the other two.

A linear mixed model was used to estimate these variance components of each method separately. The dependent variable was hepcidin outcome, and the independent random variables were: sample (plasma: 7 levels, urine: 8 levels), day (4 levels) and repeated measurement (3 levels). We found that the estimated percentage variance due to the triplicate measurements was very small and considerably smaller than the variance due to the between-day variation. Consequently, this term was omitted from the final model and not presented separately. The SD (absolute error), the CV (relative error) and the percentage variance relative to the total variance of each random variable are presented and for each method separately.

Results and Discussion

Mean hepcidin levels for all (Table 2) samples differ considerably between all methods. Notably, hepcidin values of method V are relatively high, but trends and variability are similar to all other methods (Table 2). It is also of note that for this Round Robin study, we only used native urine and serum samples, except for one, i.e. urine sample 5 to which synthetic hepcidin was added to a final concentration of 13.1 nmol/mmol creatinine. We found that the mean hepcidin outcome of some methods (i.e. methods II, IV and VI) for this spiked urine sample are closer to this concentration than others. However, irrespective of whether the value assigned of the spiked hepcidin-25 by Peptide International is correct, these results can not simply be extrapolated to native urine or serum samples, since it is presently unclear whether the various methods evaluated behave differently for spiked and native samples. In other words, the so called commutability of the spiked samples with native clinical samples for the various methods is unknown and its assessment may be part of future standardization efforts for hepcidin measurements.15

In general, differences in hepcidin levels between methods might be due to: (i) the use of different calibration solutions with level assignments based on different techniques; (ii) to hepcidin aggregation of either the standard solution or the sample; or (iii) hepcidin binding to α2 macroglobuline or albumine26 or iv) the existence of three hepcidin isoforms hepcidin-25, 22 and 20. These four points may differentially affect IC and MS measurements, and urine and serum quantifications. More specifically, it was recently found that around 90% of the circulating hepcidin is bound to α2 macroglobulin in the blood. These observations not only raise the question whether we should measure total, bound or unbound hepcidin, but also what the methods evaluated actually measure. The search for an answer to these questions is a new challenge for which much can be learned from the measurements of steroid and thyroid hormones. Another cause for differences between IC and MS methods is that IC methods lack the selectivity to distinguish hepcidin-25 from hepcidins-20 and -22. However, the implications of including the latter two isoforms in the
Mass spectrometry methods are highlighted in black.

nmol/L for urine and plasma, respectively, unless otherwise stated; #, mean and SD in Mint/mmol creatinine for urine and in Mint/L for plasma.

CV: coefficient of variation, i.e. relative error; % of total variance, % of variance-segment to the total variance; ‡, sample means and SDs in nmol/mmol creatinine and

Between-sample, segment due to variation between samples; analytical variation, segment due to repeated measurements; SD, standard deviation, i.e. absolute error;

Table 3. Sample means and variations by method for urine and plasma hepcidin levels.

<table>
<thead>
<tr>
<th>Body fluid</th>
<th>Method</th>
<th>Mean ALL sample</th>
<th>SD</th>
<th>Between-sample CV (%)</th>
<th>% of total variance</th>
<th>Between-day CV (%)</th>
<th>% of total variance</th>
<th>Analytical variation SD</th>
<th>CV (%)</th>
<th>% of total variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urine</td>
<td>I</td>
<td>2.9</td>
<td>4.1</td>
<td>139.9</td>
<td>93.4</td>
<td>&lt;0.1</td>
<td>&lt;0.1</td>
<td>37.3</td>
<td>6.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>44.4</td>
<td>66.1</td>
<td>148.8</td>
<td>95.9</td>
<td>4.9</td>
<td>11.0</td>
<td>5.0</td>
<td>28.9</td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>13.4</td>
<td>16.9</td>
<td>125.9</td>
<td>93.7</td>
<td>1.8</td>
<td>13.3</td>
<td>1.1</td>
<td>4.0</td>
<td>5.2</td>
</tr>
<tr>
<td></td>
<td>IV</td>
<td>35.1</td>
<td>52.3</td>
<td>148.9</td>
<td>96.3</td>
<td>&lt;0.1</td>
<td>&lt;0.1</td>
<td>10.2</td>
<td>3.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>V</td>
<td>427.1</td>
<td>587.8</td>
<td>137.6</td>
<td>85.4</td>
<td>102.8</td>
<td>24.1</td>
<td>2.6</td>
<td>220.6</td>
<td>12.0</td>
</tr>
<tr>
<td></td>
<td>VI</td>
<td>7.9</td>
<td>5.7</td>
<td>72.2</td>
<td>97.2</td>
<td>0.6</td>
<td>7.2</td>
<td>1.0</td>
<td>8.9</td>
<td>1.8</td>
</tr>
<tr>
<td>Plasma</td>
<td>I</td>
<td>#11.4</td>
<td>#8.4</td>
<td>73.6</td>
<td>81.3</td>
<td>#1.6</td>
<td>14.4</td>
<td>#3.7</td>
<td>32.2</td>
<td>15.6</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>13.0</td>
<td>10.3</td>
<td>79.1</td>
<td>96.0</td>
<td>0.4</td>
<td>3.2</td>
<td>0.2</td>
<td>2.1</td>
<td>3.9</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>27.4</td>
<td>21.3</td>
<td>77.9</td>
<td>92.1</td>
<td>0.7</td>
<td>2.6</td>
<td>0.1</td>
<td>6.2</td>
<td>7.6</td>
</tr>
<tr>
<td></td>
<td>IV</td>
<td>16.4</td>
<td>13.6</td>
<td>83.2</td>
<td>99.5</td>
<td>0.2</td>
<td>1.2</td>
<td>0.1</td>
<td>1.0</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>V</td>
<td>124.6</td>
<td>107.0</td>
<td>85.8</td>
<td>89.4</td>
<td>26.1</td>
<td>20.9</td>
<td>5.3</td>
<td>26.0</td>
<td>5.3</td>
</tr>
<tr>
<td></td>
<td>VI</td>
<td>17.3</td>
<td>10.8</td>
<td>62.3</td>
<td>98.8</td>
<td>0.1</td>
<td>0.6</td>
<td>&lt;0.1</td>
<td>1.2</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>VII</td>
<td>12.4</td>
<td>11.9</td>
<td>96.3</td>
<td>94.4</td>
<td>&lt;0.1</td>
<td>&lt;0.1</td>
<td>&lt;0.1</td>
<td>2.9</td>
<td>5.6</td>
</tr>
<tr>
<td></td>
<td>VIII</td>
<td>41.5</td>
<td>16.7</td>
<td>40.2</td>
<td>90.9</td>
<td>3.1</td>
<td>7.5</td>
<td>3.1</td>
<td>4.3</td>
<td>10.4</td>
</tr>
</tbody>
</table>

Between-sample, segment due to variation between samples; analytical variation, segment due to repeated measurements; SD, standard deviation, i.e. absolute error; CV, coefficient of variation, i.e. relative error; % of total variance, % of variance-segment to the total variance; #, mean and SD in Mint/mmol creatinine and nmol/L for urine and plasma, respectively, unless otherwise stated; ¶, mean and SD in Mint/mmol creatinine and

Mass spectrometry methods are highlighted in black.

Most methods are similar in both analytical variation and between-sample variation (Table 3). Of note is that the between-sample CV is lower for plasma than for urine hepcidin. This might indicate that the difference between urine and plasma is not due to the method, but more to biological mechanisms, e.g. the hepcidin excretion pathways. MS methodologies II, IV and VII, exploiting an internal standard, show slightly lower contribution of the analytical variation to the total variance compared to the other MS methods I and III. These findings corroborate the assumption that the use of an internal standard decreases the analytical variation of the MS-techniques. IC-method V shows a high between-sample CV in combination with a relatively high analytical CV. Furthermore, IC-methods VI and VIII express the lowest between-sample CV compared to all other methods. However, both methods also display low analytical variation. This relatively low between-sample variation of both the latter IC-methods illustrates the difficulties in the generation of specific antibodies for hepcidin.

Importantly, the contribution of the analytical variation to the total variance is low for all methods (Table 3), which indicates the potential suitability of all methods to distinguish hepcidin levels of different samples. However, of note is that the higher the analytical variation of a method the lower the probability that populations with only small differences in hepcidin levels can be distinguished, e.g. hepcidin levels of healthy controls from that of patients with low-grade inflammation due to the presence of features of the metabolic syndrome.

Spearman correlations between the individual sample mean hepcidin values obtained by most methods were generally high (range 0.62-1.00), except for correlations with method VIII that are somewhat disappointing (range 0.04-0.18). This should be interpreted with caution due to the small number of samples analyzed (n=7 and n=8 for plasma and urine, respectively).

In summary, hepcidin levels reported by the various methods vary considerably but analytical variation is generally low and similar for all methods. We recommend further harmonization of the various hepcidin assays by: (i) introducing an internal standard for all the MS-based methods used for clinical studies; (ii) reaching consensus on level assignment and level adjustment of the calibrators used in every procedure; (iii) production of a calibrator that mimics patient sera and (iv) regular testing of shared samples and/or calibrators that are commutable and have been value assigned for quality control.

**Authorship and Disclosures**

JJCK and EHJMK designed research, performed statistical analysis, contributed analytical tools, interpreted data and wrote the paper. SSB, MB, NC, DG, RCH, VK, AM, GO, NT, CT, DGW and TG contributed analytical tools and edited the paper. JCMH, designed and performed statistical analysis, interpreted data and co-wrote the paper. DWS designed research and statistical analysis, interpreted the data, wrote the paper and co-ordinated the study. DWS and HT steer the www.hepcidinanalysis.com initiative that serves the scientific community with hepcidin analysis in body fluids.

VK and AM have patents applications to disclose (ELISA hepcidin); GO and TG are employees of Intrinsic Life Sciences and have ownership of stocks to disclose. The other authors have no conflict of interests to declare.
References