University of Groningen

Harmonization of LC-MS/MS Measurements of Plasma Free Normetanephrine, Metanephrine, and 3-Methoxytyramine

Peitzsch, Mirko; Novos, Talia; Kaden, Denise; Kurlbaum, Max; van Herwaarden, Antonius E.; Mueller, Daniel; Adaway, Jo; Grouzmann, Eric; McWhinney, Brett; Hoad, Kirsten

Published in:
Clinical Chemistry

DOI:
[10.1093/clinchem/hvab060](https://doi.org/10.1093/clinchem/hvab060)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2021

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Peitzsch, M., Novos, T., Kaden, D., Kurlbaum, M., van Herwaarden, A. E., Mueller, D., Adaway, J., Grouzmann, E., McWhinney, B., Hoad, K., Woollard, G., Kema, I., Boot, C., Fassnacht, M., Sweep, F., Loh, T. P., Horvath, A. R., & Eisenhofer, G. (2021). Harmonization of LC-MS/MS Measurements of Plasma Free Normetanephrine, Metanephrine, and 3-Methoxytyramine. *Clinical Chemistry*, 67(8), 1098-1112. <https://doi.org/10.1093/clinchem/hvab060>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Harmonization of LC-MS/MS Measurements of Plasma Free Normetanephrine, Metanephrine, and 3-Methoxytyramine

Mirko Peitzsch^{a,*}, Talia Novos^b, Denise Kaden^a, Max Kurlbaum^c, Antonius E. van Herwaarden^d, Daniel Müller^e, Jo Adaway^f, Eric Grouzmann^g, Brett McWhinney^h, Kirsten Hoadⁱ, Gerald Woollard^j, Ido Kema^k, Christopher Boot^l, Martin Fassnacht^c, Fred Sweep^d, Tze P. Loh^m, Andrea R. Horvath^b, and Graeme Eisenhofer^{a,n}

BACKGROUND: Plasma-free normetanephrine and metanephrine (metanephrines) are the recommended biomarkers for testing of pheochromocytoma and paraganglioma (PPGL). This study evaluated the status of harmonization of liquid chromatography-tandem mass spectrometry-based measurements of plasma metanephrines and methoxytyramine and clinical interpretation of test results.

METHODS: 125 plasma samples from patients tested for PPGLs were analyzed in 12 laboratories. Analytical performance was also assessed from results of a proficiency-testing program. Agreement of test results from different laboratories was assessed by Passing-Bablok regression and Bland-Altman analysis. Agreement in clinical test interpretation based on laboratory specific reference intervals was also examined.

RESULTS: Comparisons of analytical test results by regression analysis revealed strong correlations for normetanephrine and metanephrine ($R \geq 0.95$) with mean slopes of 1.013 (range 0.975–1.078), and 1.019 (range 0.963–1.081), and intercepts of -0.584 (-53.736 to 54.790) and -3.194 (-17.152 to 5.933), respectively. The mean bias between methods was 1.2% (-11.6% to 16.0%) for metanephrine and 0.1% (-18.0% to 9.5%) for normetanephrine. Measurements of 3-methoxytyramine revealed suboptimal agreement between laboratories with biases ranging from -32.2% to 64.0% .

Interrater agreement in test interpretation was $>94\%$ for metanephrine and $>84\%$ for normetanephrine; improvements in interrater agreement were observed with use of harmonized reference intervals, including age-specific cut-offs for normetanephrine.

CONCLUSIONS: Analytical methods for metanephrines are well harmonized between laboratories. However, the 16% disagreement in test interpretation for normetanephrine suggests use of suboptimal method-dependent reference intervals for clinical decision-making for this metabolite. Improved analytical methods and reference interval harmonization are particularly required for 3-methoxytyramine.

Introduction

For initial biochemical screening for pheochromocytomas and paragangliomas (PPGLs), the Endocrine Society guideline recommends measurements of plasma-free or urinary fractionated normetanephrine and metanephrine, together referred to as metanephrines (1). Addition of 3-methoxytyramine, the O-methylated metabolite of dopamine, to the test panel provides additional utility for detection of rare dopamine-producing PPGLs (2). With liquid chromatography-tandem mass spectrometry (LC-MS/MS) combined with appropriately established reference intervals and preanalytical

^aInstitute of Clinical Chemistry and Laboratory Medicine, University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany; ^bDepartment of Clinical Chemistry, New South Wales Health Pathology, Prince of Wales Hospital, Sydney, Australia; ^cDepartment of Endocrinology and Diabetology, Julius-Maximilians-Universität Würzburg, Würzburg, Germany; ^dDepartment of Laboratory Medicine, Radboud University Medical Center, Nijmegen, The Netherlands; ^eInstitute for Clinical Chemistry, University Hospital Zurich, Zurich, Switzerland; ^fDepartment of Clinical Biochemistry, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK; ^gDépartement Médecine de Laboratoire et Pathologie, Laboratoire des Catécholamines et Peptides, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland; ^hDepartment of Chemical Pathology, Pathology Queensland, Royal Brisbane and Women's Hospital, Brisbane, Queensland, Australia; ⁱPathWest Laboratory, Fiona Stanley Hospital, Perth, Western Australia, Australia; ^jDepartment of

Pathology and Laboratory Medicine, Auckland City Hospital, Auckland, New Zealand; ^kDepartment of Laboratory Medicine, University Medical Centre Groningen, Groningen, The Netherlands; ^lDepartment of Blood Sciences, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK; ^mDepartment of Laboratory Medicine, National University Hospital Singapore, Singapore; ⁿDepartment of Internal Medicine III, University Hospital Carl Gustav Carus, Medical Faculty Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany.

*Address correspondence to this author at: Institute of Clinical Chemistry and Laboratory Medicine, University Hospital Carl Gustav Carus, Technische Universität Dresden, Fetscherstrasse 74, 01307 Dresden, Germany. Fax 0049-351-458-7346; e-mail mirko.peitzsch@uniklinikum-dresden.de.

Received December 27, 2020; accepted March 19, 2021.

DOI: 10.1093/clinchem/hvab060

precautions, measurements of plasma metabolites offer superior diagnostic accuracy over urine metabolites (3).

Although use of plasma-free metanephrines for diagnosis of PPGLs was established in 1995 (4), effective utilization of this test has been hampered by incorrectly applied preanalytical requirements, inappropriately established reference intervals, and inaccurate analytical methods (5–9). Poorly calibrated immunoassays that underestimate metabolite concentrations by more than 50%, in conjunction with reference intervals set too high, provide examples of how a potentially fatal tumor can be missed by incorrectly established methods (5, 10). The need for safe and effective diagnosis and management of patients with PPGLs calls for standardization, or at least harmonization, of the different laboratory-developed tests and commercial kit methods available for measurements of these metabolites.

External quality assurance (EQA) programs offer one approach for assessing interlaboratory comparability of test results to thereby inform the process of harmonization of laboratory tests and reference intervals (11–16). For immunoassays, this is made inherently difficult by the plethora of different commercial test kits, calibrators, and interferences (17, 18). In one report, harmonization of test results was termed a “mission impossible” (19). Although mass spectrometry-based methods are not immune to assay interferences (20–24), the high analytical specificity of these methods is an enabler of assay harmonization and standardization (25–27). The concept of harmonization should consider the total testing process, including preanalytical (matrix composition, sample collection, and processing), analytical (traceable calibration material, certified reference materials, a higher order reference measurement methods), and postanalytical procedures (units of reporting and reference intervals) (28).

For plasma-free metanephrines and 3-methoxytyramine, neither reference materials nor reference methods exist; therefore, standardization is not yet possible. The first step in such situations, according to a framework for test harmonization (29), is to assess the degree of measurement equivalence. As outlined earlier, participation in EQA programs (11) provides one means for comparing the trueness and comparability of results. However, such programs typically use artificially prepared, noncommutable materials that may not contain potentially interfering analytes inherently present in native patient samples. For example, interference by 3-O-methyldopa in measurements of methoxytyramine results in bias in native plasma but not in samples used in EQA programs (23). The present study therefore aimed to establish interlaboratory comparability of LC-MS/MS measurements of plasma-free metanephrines and 3-methoxytyramine in patient-derived samples. The study also compared test interpretations based on

the locally employed reference limits used by participating laboratories.

Materials and Methods

PARTICIPANTS AND ANALYTICAL METHODS

Twelve laboratories, including 2 in Germany (Dresden, Wuerzburg), three in Australia (Brisbane, Perth, Sydney), 2 in Switzerland (Lausanne, Zurich), 2 in the United Kingdom (Manchester, Newcastle), 2 in The Netherlands (Groningen, Nijmegen), and another in New Zealand (Auckland), participated in the study. All laboratories employed independently developed analytical settings (Supplemental Table 1). Laboratory-developed tests were used by 11 laboratories and another used a commercially available kit method (Chromsystems, Germany). A registration form was provided to all participants requesting relevant information on methods, including details of sample preparation, respective LC-MS/MS methods, and calibration materials (Supplemental Table 1).

SPECIMENS

The 125 plasma samples used for this study were derived from larger numbers of lithium heparin anticoagulated plasma specimens left over after routine analyses at 4 different participating centers. Specimens were collected between January 2018 to August 2018 after an overnight fast and at least 20 min of supine rest. All samples were from patients undergoing biochemical testing for PPGLs and in whom measurements of plasma metanephrines were available. Due to the limited volume of samples left after these measurements, there was a need to pool specimens from different patients for subsequent aliquoting and distribution to the 12 participating laboratories. For pooling, samples were first grouped according to similar concentrations and relative differences of concentrations for each of the 3 metabolites. Specimens with similar patterns of metabolite concentrations were then thawed and pooled together to achieve specimen volumes of at least 6.5 mL; in that process samples were de-identified to satisfy ethical committee requirements. A set of 125 plasma samples was thereby prepared that covered a range of concentrations from low normal to pathologic, and for the latter, variable increases commonly observed in patients with PPGLs. Each of the 125 samples was then divided into aliquots of a little over 0.5 mL and frozen at -80°C until distribution to laboratories. The specimen aliquots were distributed (on dry ice) in October 2018 to each participating laboratory and thereafter stored at -80°C until analysis within 4 months.

PROFICIENCY TESTING

Participation in the study required enrollment of laboratories in the international Quality Assurance Program

(QAP) of the Royal College of Pathologists of Australasia (RCPA) for plasma-free metanephrines (11). Method agreement within the RCPAQAP was examined according to results from lyophilized human serum samples assayed over the course of 2 6-month cycles of the program during 2018 (12 samples/cycle).

DATA ANALYSIS

Results for measurements of plasma metabolites were reported back to the principal investigators, compiled and shared with study participants for subsequent data review. Analysis of the data was carried out at Dresden and Sydney, with additional independent review at Singapore. Statistical analyses utilized the JMP® software package.

Lack of reference methods/reference materials for defining true target values for each measurand necessitated use of median concentrations of all laboratories for each patient sample as consensus target values (all-method medians). The RCPAQAP similarly uses all-participant medians as consensus target values. Analytical agreement between each participating laboratory and the all-laboratory medians for patient samples and all-participant medians for RCPAQAP materials was assessed using Passing-Bablok regression and Bland-Altman analyses.

Agreement in test interpretation was assessed by comparing the number of measurements falling within (nonpathological) and above (pathological) the upper reference limits (URLs) used by each participating laboratory (Supplemental Table 1). Plasma concentrations of normetanephrine increase with age (12). For this analyte, random patient ages were assigned to the study samples by the principal investigators. A result was defined “pathological” as any increase above the respective laboratory-derived URL. Interpretations were reported in conjunction with the analytical data and respective URLs used by the participants for clinical decision-making. Agreement between laboratories in the interpretation of nonpathological versus pathological results was assessed using the Fleiss-kappa inter-observer statistical test.

To evaluate the impact of method harmonization through calibration adjustment (i.e., bottom-up harmonization approach), the measurements of individual laboratories were adjusted to the regression equation derived from the all-method medians. The adjusted measurements were then compared against the original URLs of corresponding laboratories and reinterpreted. To evaluate the additive clinical impact of harmonized reference intervals, the adjusted measurements were further assessed and reinterpreted against URLs established by laboratory A according to clinical validation (3, 12).

The analytical performance specifications (APS) were adopted from the RCPAQAP (30). For

normetanephrine, metanephrine, and 3-methoxytyramine, those APS depend on metabolite concentrations, with ranges of ± 200 pmol/L, ± 75 pmol/L, and ± 36 pmol/L for observed concentrations up to 1000 pmol/L, 500 pmol/L, and 120 pmol/L, respectively, and $\pm 20\%$, $\pm 15\%$, and $\pm 30\%$ above these concentrations.

Results

INTER-LABORATORY COMPARISONS OF PATIENT DATA

Target concentrations for metabolites, defined by medians of the 12 laboratories, ranged from 171 to 37 281 pmol/L, 33 to 9116 pmol/L and 13 to 3612 pmol/L, for normetanephrine, metanephrine, and 3-methoxytyramine, respectively (Fig. 1, Supplemental Table 2). Results for normetanephrine were reported by all 12 laboratories for all 125 samples, whereas the full set of results for metanephrine were reported by 8 laboratories. The shortfall in reported results for metanephrine reflected concentrations below lower limits of quantification (LLOQ) for 19 and 36 samples in respective laboratories D and E, and 3 and 9 samples in laboratories F and M due to other analytical problems (Supplemental Table 1). Complete data for 3-methoxytyramine were reported by 5 laboratories (A, I, J, K, and L). For laboratories B, C, D, E, F, H, and M, results were reported as below the LLOQ in 83, 90, 90, 91, 5, 6, and 18 respective samples.

Variations in reported data (interquartile range (IQR)/median) ranged from 3.4% to 66.1% (median 12.6%, IQR 8.9%–17.8%) for normetanephrine, from 3.5% to 73.5% (median 13.6%, IQR 9.4%–21.3%) for metanephrine, and from 4.9% to 327.8% (median 47.5%, IQR 28.3%–79.9%) for 3-methoxytyramine (Fig. 1). Larger variations in reported data were observed in samples with lower concentrations (Supplemental Fig. 1).

Passing-Bablok regression analyses using the all-method medians as consensus target values revealed strong correlations for normetanephrine (mean $R = 0.994$; 95% CI 0.986–1.000), metanephrine (mean $R = 0.993$; 95% CI 0.985–1.000), and 3-methoxytyramine (mean $R = 0.975$; 95% CI 0.949–1.000) (Table 1, Fig. 2). Estimated slopes of regression lines were 1.013 (95% CI 0.996–1.035) for normetanephrine and 1.019 (95% CI 0.988–1.053) for metanephrine, indicating no significant proportional bias. However, for 3-methoxytyramine, three laboratories (A, B, M) presented with weaker correlations to the all-method medians ($R < 0.95$). One laboratory (M) showed a particularly high proportional bias indicated by a slope of the regression line of 3.542; this was subsequently clarified to reflect an unknown interference that was corrected after the present data became available.

Table 1. Differences in analytical test results for normetanephrine (NMN), metanephrine (MN), and 3-methoxytyramine (MTY) in native samples.

Passing-Bablok-Regression (Y = m*(all-method median) + b)										Bland-Altman-Analysis								
n ^a	slope	Lower CL (slope)	Upper CL (slope)	intercept	Lower CL (intercept)	Upper CL (intercept)	R	mean difference [%]	SD	mean + 1.96 SD	mean - 1.96 SD	slope	P	intercept	P	R	P	
NMN																		
Lab A	125	1.078	1.066	1.093	44.28	33.92	58.60	0.9982	16.0	18.5	52.3	-20.3	-0.004	0.0497	15.949	<0.0001	-0.1769	<0.0001
Lab B	125	0.988	0.974	1.002	-15.67	-26.74	-5.68	0.9988	-4.0	20.2	35.6	-43.6	0.0002		-4.490	0.0247	0.0681	
Lab C	125	0.994	0.985	1.003	-5.21	-10.45	-0.95	0.9997	-3.1	11.0	18.4	-24.5	0.0001		-3.351	0.0021	0.0642	0.0458
Lab D	125	1.035	1.006	1.057	-18.26	-32.47	-2.42	0.9953	-0.2	9.7	18.9	-19.3	0.0003		-0.830		0.1590	0.0221
Lab E	121	1.074	1.053	1.086	-12.57	-24.24	-0.62	0.9994	5.9	10.5	26.4	-14.6	-0.0001		6.079	<0.0001	-0.0387	0.0023
Lab F	120	1.028	1.000	1.071	-20.00	-43.12	-4.33	0.9504	-4.8	19.4	33.4	-42.9	-0.0003		-4.024	0.0413	-0.0836	0.0091
Lab H	125	0.996	0.980	1.025	54.79	34.05	80.24	0.9978	13.5	19.5	51.8	-24.8	-0.0010	0.0049	15.665	<0.0001	-0.2530	0.0009
Lab I	125	1.021	1.007	1.063	14.18	-19.19	27.87	0.9972	9.3	21.9	52.2	-33.6	-0.0004		10.137	<0.0001	-0.0912	
Lab J	125	0.978	0.967	0.986	-3.05	-9.93	3.37	0.9995	-4.5	8.3	11.6	-20.7	0.0001		-4.728	<0.0001	0.0512	
Lab K	125	0.979	0.960	0.996	15.65	2.93	26.86	0.9981	0.7	12.3	24.8	-23.3	-0.0002		1.180		-0.0800	
Lab L	125	0.975	0.965	0.994	-7.42	-21.16	-1.97	0.9980	-2.8	11.3	19.3	-25.0	0.0002		-3.326	0.0031	0.1038	0.0102
Lab M	123	1.010	0.986	1.039	-53.74	-67.42	-27.08	0.9958	-11.6	22.7	33.0	-56.1	0.0006		-12.891	<0.0001	0.1404	<0.0001
MN																		
Lab A	125	1.081	1.054	1.131	1.18	-5.13	5.57	0.9976	9.5	12.7	34.4	-15.5	0.0000		9.472	<0.0001	0.0033	
Lab B	125	0.981	0.941	1.040	-6.28	-15.65	2.98	0.9450	-5.0	28.0	49.8	-59.9	0.0024		-6.055	0.0247	0.0995	0.0181
Lab C	125	0.963	0.946	0.978	-0.75	-3.27	2.01	0.9996	-2.4	13.7	24.5	-29.2	-0.0002		-2.303		-0.0154	
Lab D	106	0.997	0.963	1.039	0.43	-5.97	6.89	0.9977	1.6	8.8	18.9	-15.6	-0.0002		1.688		-0.0233	
Lab E	89	1.022	0.986	1.067	5.93	-4.46	13.83	0.9993	7.0	8.4	23.4	-9.5	-0.0008		7.421	<0.0001	-0.1350	
Lab F	122	0.974	0.936	1.018	-17.15	-25.66	-12.37	0.9987	-18.0	18.7	18.7	-54.7	0.0027		-19.154	<0.0001	0.1710	<0.0001
Lab H	125	1.047	1.026	1.062	-2.83	-4.93	1.79	0.9979	5.7	13.8	32.8	-21.3	0.0001		5.674	<0.0001	0.0129	
Lab I	125	1.008	0.970	1.046	1.98	-3.94	7.63	0.9947	3.7	18.1	39.3	-31.8	-0.0003		3.890	0.0264	-0.0222	
Lab J	125	0.985	0.969	0.992	-2.76	-4.24	-0.15	0.9998	-5.7	10.9	15.6	-27.0	0.0007		-5.969	<0.0001	0.0759	
Lab K	125	1.071	1.044	1.099	0.57	-4.88	4.96	0.9996	7.4	13.2	33.3	-18.5	-0.0004		7.559	<0.0001	-0.0349	
Lab L	125	1.045	1.009	1.069	-10.35	-15.79	-5.03	0.9917	-2.8	14.1	24.8	-30.4	0.0017		-3.559	0.0080	0.1563	<0.0001
Continued																		

Continued

Table 1. (continued)																		
Passing-Bablok-Regression ($Y = m \cdot (\text{all-method median}) + b$)											Bland-Altman-Analysis							
	n ^a	slope	Lower CL (slope)	Upper CL (slope)	inter-cept	Lower CL (inter-cept)	Upper CL (inter-cept)	R	mean difference [%]	SD	mean + 1.96 SD	mean - 1.96 SD	slope	P	intercept	P	R	P
Lab M	123	1.055	1.013	1.094	-8.29	-14.83	-1.01	0.9981	-0.3	20.7	40.4	-41.0	0.0004		-0.467		0.0239	
MTY																		
Lab A	125	1.273	1.186	1.388	3.78	-2.25	7.51	0.9430	32.4	31.3	93.8	-29.0	0.0037		31.660	<0.0001	0.0664	
Lab B	42	0.842	0.711	0.956	-1.34	-14.69	11.80	0.9326	-12.9	54.9	94.7	-120.5	0.0091		-16.135		0.1156	
Lab C	35	0.976	0.878	1.042	18.25	3.63	37.49	0.9891	27.5	45.0	115.8	-60.8	-0.0115		32.957	0.0006	-0.2078	0.0064
Lab D	35	1.013	0.991	1.054	-2.74	-12.68	0.81	0.9986	-1.2	9.4	17.3	-19.7	0.0026		-2.468		0.2227	
Lab E	34	0.926	0.871	0.968	12.31	2.79	25.08	0.9986	-1.2	18.0	34.2	-36.5	-0.0042		1.190		-0.1925	0.0019
Lab F	120	0.977	0.958	1.007	-9.79	-11.44	-7.63	0.9994	-32.2	34.4	35.2	-99.5	0.0168	0.0135	-34.876	<0.0001	0.2250	<0.0001
Lab H	119	1.111	1.060	1.158	-10.11	-14.33	-7.67	0.9981	-17.2	42.3	65.8	-100.1	0.0145		-19.684	<0.0001	0.1661	<0.0001
Lab I	125	1.020	0.991	1.067	-6.55	-8.60	-4.78	0.9983	-13.5	40.8	66.2	-93.5	0.0091		-14.930	0.0002	0.1034	<0.0001
Lab J	125	0.988	0.968	1.018	-2.13	-3.93	-0.20	0.9995	-7.9	26.2	43.5	-59.4	0.0028		-8.397	0.0010	0.0487	0.0019
Lab K	125	1.051	1.004	1.148	13.74	9.87	17.91	0.9967	33.0	31.6	95.0	-29.0	-0.0235	0.0003	36.971	<0.0001	-0.3219	0.0025
Lab L	125	1.107	1.074	1.140	-7.52	-9.31	-5.90	0.9973	-11.8	24.9	36.9	-60.6	0.0148	0.0007	-14.392	<0.0001	0.2997	<0.0001
Lab M	109	3.542	2.126	5.724	-67.71	-165.5	-12.64	0.8491	64.0	72.4	206.0	-78.0	0.0096		61.069	<0.0001	0.0976	0.0004
*Missing reported results observed in respective samples are majorly caused by measurements <lower limits of quantification, or failures during the sample extraction (CL, confidence limit; SD, standard deviation).

Table 2. Differences in analytical test results for normetanephrine (NMN), metanephrine (MN), and 3-methoxytyramine (MTY) in samples derived from an external quality assurance scheme																		
Passing-Bablok-regression ($Y = m \cdot (\text{all-method median}) + b$)										Bland-Altman-analysis								
	n ^a	slope	Lower CL (slope)	Upper CL (slope)	inter-cept	Lower CL (inter-cept)	Upper CL (inter-cept)	R	mean difference [%]	SD	mean + 1.96SD	mean - 1.96SD	slope	P	intercept	P	R	P
NMN																		
Lab A	24	0.990	0.958	1.013	13.57	-16.93	58.73	0.9985	0.1	2.9	5.8	-5.6	-0.0008		1.364		-0.2111	
Lab B	22	1.022	0.983	1.070	-74.35	-151.81	-29.63	0.9940	-5.7	6.4	6.8	-18.2	0.0050	0.0087	-12.889	0.0001	0.5452	0.0039
Lab C	22	1.038	0.999	1.100	-21.02	-83.38	11.62	0.9963	1.6	4.3	10.0	-6.8	0.0019		-1.274		0.3247	
Lab D	24	1.028	0.991	1.074	-20.00	-77.91	19.15	0.9957	1.7	4.1	9.6	-6.3	0.0009		0.344		0.1678	
Lab E	12	1.153	1.074	1.227	-24.62	-129.86	51.09	0.9958	12.1	4.1	20.2	4.1	0.0011		10.412	0.0049	0.2192	
Lab F	24	1.096	0.951	1.335	-44.30	-337.00	90.36	0.9929	-2.1	16.8	30.9	-35.0	0.0000		2.472		0.0040	
Lab H	24	0.936	0.885	0.994	-12.30	-75.55	48.90	0.9931	-7.9	6.2	4.2	-20.0	-0.0007		-6.866	0.0272	-0.0843	
Lab I	24	1.048	0.973	1.171	-85.04	-265.58	18.75	0.9834	-2.3	10.6	18.6	-23.1	0.0026		-6.044		0.1855	
Lab J	24	0.967	0.886	1.020	-24.73	-69.56	42.17	0.9861	-6.2	7.0	7.6	-20.0	-0.0004		-5.581		-0.0419	
Lab K	24	0.984	0.942	1.045	11.52	-44.69	74.03	0.9944	0.3	4.8	9.6	-9.1	-0.0015		2.400		-0.2216	
Lab L	24	0.970	0.948	0.991	-12.62	-33.72	11.79	0.9990	-4.1	2.6	1.0	-9.2	0.0009		-5.391	0.0002	0.2498	
Lab M	24	0.985	0.902	1.072	-6.80	-98.93	6.94	0.9832	-1.4	9.9	18.0	-20.8	-0.0024		2.069		-0.1741	
MN																		
Lab A	24	0.932	0.893	0.963	10.69	-11.53	49.47	0.9955	-5.3	4.4	3.3	-13.8	-0.0046		-1.765		-0.3859	
Lab B	22	1.009	0.981	1.078	-12.36	-43.88	6.02	0.9956	-0.7	4.2	7.6	-9.0	0.0035		-3.465		0.3053	
Lab C	22	1.055	0.966	1.129	-9.87	-72.28	37.80	0.9862	2.2	8.3	18.4	-14.1	-0.0011		3.029		-0.0480	
Lab D	24	1.011	0.966	1.079	-17.95	-60.38	7.89	0.9925	-1.1	5.5	9.7	-11.9	0.0029		-3.381		0.2048	
Lab E	12	1.106	1.059	1.167	-12.26	-70.08	8.03	0.9985	6.6	3.3	13.1	0.1	0.0033		3.930		0.4134	
Lab F	24	1.033	1.003	1.067	-36.95	-48.97	-8.18	0.9977	-2.0	5.0	7.8	-11.7	0.0071	0.0047	-7.526	0.0009	0.5565	0.0053
Lab H	24	1.012	0.978	1.043	-10.79	-33.97	12.44	0.9963	-0.6	5.1	9.4	-10.6	-0.0002		-0.473		-0.0130	
Lab I	24	1.022	0.965	1.078	-10.61	-46.66	19.12	0.9937	-0.3	6.1	11.8	-12.3	0.0027		-2.423		0.1691	
Lab J	24	1.000	0.947	1.050	-35.80	-69.42	-4.74	0.9928	-7.2	6.7	6.0	-20.4	0.0073	0.0453	-12.715	0.0002	0.4122	0.0285
Lab K	24	1.041	1.001	1.094	-6.75	-51.09	14.99	0.9955	2.7	5.1	12.8	-7.4	0.0009		1.969		0.0697	
Lab L	24	0.981	0.951	1.019	-3.74	-36.71	20.16	0.9968	-2.5	5.2	7.8	-12.7	-0.0002		-2.332		-0.0145	
Continued																		

Table 2. (continued)

Passing-Bablok-regression (Y = m*(all-method median) + b)										Bland-Altman-analysis								
	n ^a	slope	Lower CL (slope)	Upper CL (slope)	intercept	Lower CL (intercept)	Upper CL (intercept)	R	mean difference [%]	SD	mean + 1.96SD	mean - 1.96SD	slope	P	intercept	P	R	P
Lab M	24	0.975	0.921	1.066	6.75	-53.10	44.99	0.9871	-1.8	7.7	13.4	-16.9	-0.0030		0.563		-0.1465	
MTY																		
Lab A	24	0.999	0.974	1.022	6.61	-6.03	21.25	0.9977	2.0	4.7	11.2	-7.1	-0.0049		5.304	0.0098	-0.3880	
Lab B	22	0.955	0.909	1.011	-8.23	-39.34	20.41	0.9960	-4.5	5.8	6.9	-16.0	-0.0024		-3.043		-0.1391	
Lab C	22	1.090	1.024	1.153	-14.66	-54.58	28.26	0.9924	4.9	5.9	16.5	-6.6	0.0034		2.637		0.2098	
Lab D	24	1.054	0.993	1.107	-7.27	-35.83	11.20	0.9950	3.1	4.7	12.3	-6.1	0.0021		1.733		0.1648	
Lab E	12	1.098	1.048	1.183	-5.49	-68.32	21.98	0.9972	6.9	6.6	19.7	-6.0	0.0033		4.571		0.2009	
Lab F	24	1.076	1.046	1.110	-5.18	-23.23	10.99	0.9978	5.1	5.2	15.3	-5.0	0.0052		1.634		0.3814	
Lab H	24	0.961	0.889	1.029	-8.05	-34.32	24.56	0.9874	-7.4	8.4	9.1	-23.8	0.0046		-10.278	0.0085	0.1953	
Lab I	24	0.941	0.867	1.043	-2.45	-55.58	27.26	0.9738	-6.2	10.2	13.8	-26.2	0.0048		-9.230	0.0429	0.1721	
Lab J																		
Lab K	23	1.054	0.983	1.129	10.83	-37.82	60.64	0.9915	7.8	7.7	22.8	-7.3	-0.0041		10.650	0.0066	0.1961	
Lab L	24	0.944	0.924	0.969	-1.16	-14.54	7.21	0.9975	-6.4	4.4	2.3	-15.0	0.0038		-8.797	<0.0001	0.3129	
Lab M	24	1.070	0.933	1.206	34.10	-19.85	99.48	0.9678	13.6	15.7	44.3	-17.1	-0.0174	0.0457	25.603	0.0006	-0.4116	
*Numbers lower than 24 caused by laboratory-specific participation in the external quality assurance program (CL, confidence limits; SD, standard deviation).																		

^aNumbers lower than 24 caused by laboratory-specific participation in the external quality assurance program (CL, confidence limits; SD, standard deviation).

Bland-Altman analyses of results from individual laboratories for normetanephrine compared to the all-methods median (Table 1) showed a mean difference less than that for the RCPAQAP APS of $\pm 20\%$ for all laboratories. For metanephrine, all laboratories met the RCPAQAP APS of $\pm 15\%$ except for Laboratory F (mean difference = -18%). For 3-methoxytyramine, most laboratories met the RCPAQAP APS requirements of $\pm 30\%$ except for laboratories A (32%), F (-32%), K (33%), and M (64%) (Table 1, Fig. 2).

INTER-LABORATORY COMPARISON OF PROFICIENCY TESTING DERIVED DATA

Median concentrations of metanephrines measured in RCPAQAP samples by the 12 laboratories ranged from 399 to 2562 pmol/L for normetanephrine and 228 to 1374 pmol/L for metanephrine (Fig. 1, Supplemental Table 2). For 3-methoxytyramine, median concentrations derived from the 11 laboratories that reported these measurements ranged from 122 to 1205 pmol/L. Variations in reported RCPAQAP data ranged from 2.8% to 16.6% (median 7.6%, IQR 6.6%–10.4%) for normetanephrine, from 2.0% to 13.2% (median 6.4%, IQR 5.5%–8.5%) for metanephrine, and from 6.2% to 15.6% (median 11.0%, IQR 8.3%–12.7%) for 3-methoxytyramine (Fig. 1). In contrast to results for patient samples, there were no differences in variation according to the target concentration ranges (online Supplemental Fig. 1).

Passing-Bablok regression analyses in RCPAQAP samples revealed excellent correlations ($R = 0.993$; $R = 0.994$; $R = 0.990$) for all 3 catecholamine metabolites with observed slopes of 1.018 (95% CI 0.958–1.093), 1.015 (95% CI 0.969–1.070), and 1.001 (95% CI 0.963–1.087) for normetanephrine, metanephrine, and 3-methoxytyramine, respectively (Fig. 3, Table 2). Bland-Altman analyses revealed mean observed biases for the respective metabolites of 1.2% (range -7.9% to 12.1%), -0.8% (range -7.2% to 6.6%) and 1.7% (range -7.4% to 13.6%) in comparison to target values (Fig. 3, Table 2). Therefore, all laboratories met RCPAQAP APS requirements for measurements of all 3 metabolites (Supplemental Fig. 2).

COMPARISONS OF NATIVE PATIENT VS PROFICIENCY TESTING SAMPLE DATA

Median concentrations of normetanephrine, metanephrine, and 3-methoxytyramine were 2.7-fold ($P = 0.006$), 4.7-fold and 14.5-fold ($P \leq 0.001$) higher in RCPAQAP samples than in samples from patients (Fig. 1). In contrast, median variations for the 3 metabolites were 1.7-fold (7.6% vs 12.6%), 2.1-fold (6.4% vs 13.6%), and 3.6-fold (11.0% vs 40.1%) ($P \leq 0.001$), respectively, lower in RCPAQAP samples than in patient samples.

AGREEMENT IN TEST INTERPRETATION

In order to assess agreement among laboratories in interpretation of test results, we examined data according to interpretations achieved according to the URLs used by each laboratory. Seven of 12 laboratories used age-dependent URLs for normetanephrine, whereas 5 laboratories used fixed URLs ranging from 710 pmol/L to 1180 pmol/L (Supplemental Table 1, Supplemental Table 2). Mean inter-rater agreement in test interpretation was 90.6% (113/125) (95% CI 89.2%–92.0%) and ranged from 83.7% to 93.3% (104/125 to 116/125) (Fig. 4).

After normalization of analytical data to the all-method medians, the agreement in test interpretation for normetanephrine increased slightly from 90.6% (113/125) to 91.2% (114/125) (95% CI 89.8%–92.6%), with a range of 84.9% to 93.6% (106/125 to 117/125). Further normalization of analytical data in conjunction with the use of the same but normalized URLs improved the agreement in test interpretations to 94.4% (118/125) (95% CI 93.5%–95.3%) ranging from 91.9% to 96.4% (115/125 to 120/125). This improvement was generally observed for all participating laboratories and particularly for laboratory C, which used the highest fixed URL of 1180 pmol/L.

For metanephrine, all laboratories used fixed URLs to distinguish nonpathological from pathological test results; these ranged widely from 299 pmol/L to 560 pmol/L (Supplemental Table 1, Supplemental Table 2). After applying each laboratories' own URLs first to the analytical raw data and then the normalized data, the observed agreement in test interpretation increased marginally from 97.0% (121/125) (95% CI 96.3%–97.7%) for the raw data to 97.2% (122/125) (95% CI 96.5%–97.9%) for the normalized data (Fig. 4). After application of the same but a normalized URL, the agreement in test interpretation improved further to 98.3% (123/125) (95% CI 97.9%–98.7%). This increase was largest for laboratories C, H, J, and L, which used particularly high (510 and 560 pmol/L) or low (299 and 280 pmol/L) URLs.

Discussion

Using patient plasma samples, this study demonstrates good agreement between 12 laboratories for measurements of normetanephrine and metanephrine with respective mean biases from target of 1.2% and 0.1%. All laboratories performed well within the APS of the international RCPAQAP. Agreement for 3-methoxytyramine was, however, suboptimal, with differences from target ranging from -32% to 64% in patient samples. Since 3-methoxytyramine circulates at lower concentrations than the metanephrines, inaccuracy and imprecision for this biomarker can be expected, but does indicate need

Fig. 4. Inter-rater agreement in interpretations of 125 test results for normetanephrine (A) and metanephrine (B) using i) the analyzed raw data and participants own upper reference limits (black circles), ii) normalized data and participants own upper reference limits (black filled triangles) and iii) normalized data in conjunction with normalized upper reference limits derived from laboratory A (black rectangles). Whiskers represent the lower and upper 95% confidence intervals. 19 and 36 results for metanephrine reported below the lower limit of quantification by laboratories D and E, respectively, are assumed nonpathologic, 3 and 9 missing results due to analytical problems from respective laboratories F and M are not considered in the analysis for those laboratories.

for improved analytical methods. Although analytical results for normetanephrine and metanephrine seem to be harmonized, differences exist in test interpretation based on laboratory-specific URLs; this variability in test interpretation may be improved by harmonized reference intervals.

Standardization of analytical methods requires reference materials, reference methods, and reference laboratories. Additionally, URLs should be transferable across methods for harmonized reporting (31). Some degree of harmonization can be achieved through participation in EQA programs that use commutable samples for proficiency testing (29). EQA programs, such as the RCPAQAP, are particularly valuable for measurements

of plasma-free metanephrines and 3-methoxytyramine, for which there is lack of commutable reference materials, reference methods, and reference laboratories. Through these programs, assessments of performance are defined according to quality specifications and in relation to APS (30).

According to those APS criteria, measurements of normetanephrine in all 12 laboratories are well harmonized and should, therefore, allow use of harmonized URLs for discrimination of pathological and nonpathological test results. Similar findings for metanephrine in all except one laboratory should also allow use of harmonized URLs for this metabolite. In contrast, for 3-methoxytyramine, one-third of all laboratories had a mean bias higher than $\pm 30\%$. In addition, due to differences in assay-specific LLOQs, some laboratories failed to detect 3-methoxytyramine in the clinically relevant range of patient samples (2). Nevertheless, in samples distributed by the RCPA program, all 3 metabolites met the APS criteria with minimal mean biases of less than $\pm 2\%$. In relation to the clinically relevant concentration range, it is important to highlight that the metabolite concentrations in RCPA samples were considerably higher than in patient samples. Variations observed in analytical test results for patient and RCPA samples were similar at concentration ranges observed in RCPA samples, but much higher for lower concentration ranges commonly observed in samples from patients without PPGLs (3). Therefore, the true clinical performance of these tests cannot be assessed purely on the basis of such an EQA program.

The finding of a positive relationship for differences in bias for measurements of 3-methoxytyramine in native plasma compared to RCPA samples indicates systematic discordance in the analysis of this metabolite among laboratories; therefore, participation in an EQA program remains informative and beneficial for improving measurement accuracy for 3-methoxytyramine. However, lack of relationships of bias for normetanephrine and metanephrine point to the presence of factors within patient specimens that influence patient results but that cannot be captured using the artificial sample matrix employed in the RCPAQAP program; this is, however, a common weakness of EQA observed with many laboratory tests. Altogether, the findings of this study call for a commutable QAP material that covers the clinically relevant concentration ranges and matrix composition, including potentially interfering endogenous substances that may not be present in artificial sample matrix.

Besides agreement in analytical test results, it is also important for patient care that results are interpreted among different laboratories with consistent diagnostic accuracy. Inter-rater analysis demonstrated generally good agreement in clinical interpretation for both

normetanephrine and metanephrine. However, diagnostic agreement was higher for metanephrine than for normetanephrine. The variation of literature-reported values for URLS of normetanephrine largely reflects higher concentrations during seated than supine sampling, with the latter recommended by the Endocrine Society guideline (1). Within the 12 laboratories of the present study, only one used URLS determined for seated blood sampling (10); accordingly, that laboratory showed the least agreement in test interpretation for normetanephrine. Differential adoption of fixed vs age-specific URLS may contribute to other disparities in interpretation of results for normetanephrine.

The laboratories with the least agreements for metanephrine had particularly low or high decision limits. The lower limits may reflect use of URLS for this metabolite determined by 97.5 percentiles for supine sampling, while higher values are associated with seated sampling or established to minimize false-positive results (32). Use of the latter has been justified by findings that most patients with PPGLs present with increases of normetanephrine, with or without increases of metanephrine (3); only occasionally do patients with PPGLs present with solitary increases of metanephrine. Thus, accurate diagnosis of PPGLs depends more on URLS used for normetanephrine than for metanephrine.

After data were normalized according to a single set of published URLS (32), agreement in test interpretation improved beyond that achieved by normalizing for bias, indicating that differences in URLS are more crucial for test interpretation than differences in analytical test performance. This strengthens the argument for common use of URLS as a next step towards harmonized reporting. Age-specific URLS for plasma-free normetanephrine have been proposed (12) and further validated in a comprehensive prospective diagnostic accuracy study involving more than 2000 patients including 236 confirmed PPGL cases (3). Age-specific URLS, with lower decision limits for younger patients and continuously increasing cut-offs with age, assist in reducing false-positive test results in older patients and false-negative test results in younger patients (12).

By demonstrating that the analytical methods for metanephrine and normetanephrine were reasonably harmonized among 12 laboratories, there can be confidence that already established age-specific reference limits (12, 32) are transferable to all laboratories of this study. In fact, the majority of laboratories within this study are already using age-specific URLS. Nevertheless, full transference of continuous age-dependent URLS into laboratory information management systems was reported to be difficult by several participants. This indicates a need to modernize existing systems to meet emerging demands for personalized interpretation of test results that will thereby achieve improved diagnostic accuracy.

For 3-methoxytyramine, the situation was quite different to that of normetanephrine and metanephrine, with highly variable URLS ranging from 36 to 200 pmol/L. Consistencies in bias according to patient and EQA materials point to systematic discordances in measurement methods. Interferences from endogenous substances (23) and cross-reaction with other metabolites, such as metanephrine when chromatographic separation is limited (33), may contribute to some of the discordance. Nevertheless, most of the discordance likely reflects imprecision of most methods for which the LLOQ was often close to the upper range of reference intervals. This points to the need for improved and more sensitive analytical methods, such as recently described by van Faassen et al. (34). Only through such methods is the full potential of plasma 3-methoxytyramine likely to be realized for applications involving patients with PPGLs or neuroblastoma (35–37).

Limitations of the study include lack of universal calibrators to exclude proportional bias due to differences in calibration, and lack of multiple aliquots of the same patient sample to establish variance within methods. Despite these limitations, the differences in agreement for normetanephrine and metanephrine were minor. Future harmonization studies should nevertheless include a universal set of calibrators and sample material replicates to assess precision. Another limitation involved the lack of true clinical diagnosis for the 125 pooled patient samples in order to better reconcile differences in test interpretation based on URLS. Relevant endogenous interfering substances, only present in some patient samples, may have also been diluted in the pooling process. Finally, some participating laboratories described LLOQs higher than required for precise measurement of low endogenous concentrations of 3-methoxytyramine and metanephrine. Therefore, such laboratories reported incomplete analytical test results that may have led to overestimated analytical accuracy. Nevertheless, LLOQs for metanephrine were always well below the applied URLS for clinical test interpretation and therefore not expected to have clinical relevance for this metabolite.

In conclusion, analytical results by LC-MS/MS and interpretation of results seem harmonizable for metanephrine and normetanephrine. However, differences exist in test interpretation based on laboratory URLS, particularly for normetanephrine. Adoption of common reference limits that have been validated in large studies and that can be easily verified locally using widely accepted criteria (38) may therefore improve clinical decision-making for the diagnosis of PPGLs by laboratories that show minimal bias in biochemical test results. Proficiency testing organizations may also consider a focus on distributing commutable QAP materials that cover concentration ranges relevant for clinical decision-making. In this way, laboratories can continuously better assess

their analytical quality in conjunction with the transferability and stability of their clinical decision limits. This should help ensure high diagnostic accuracy of laboratory services.

Supplemental Material

Supplemental material is available at *Clinical Chemistry* online.

Nonstandard Abbreviations: PPGL, pheochromocytoma and paraganglioma; LC-MS/MS, liquid chromatography-tandem mass spectrometry; EQA, external quality assurance; QAP, Quality Assurance Program; RCPA, Royal College of Pathologists of Australasia; URL, upper reference limit; APS, analytical performance specifications; LLOQ, lower limit of quantification.

Author Contributions: All authors confirmed they have contributed to the intellectual content of this paper and have met the following 4 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; (c) final approval of the published article; and (d) agreement to be accountable for all aspects of the article thus ensuring that questions related to the accuracy or integrity of any part of the article are appropriately investigated and resolved.

M. Peitzsch, financial support, administrative support, statistical analysis, provision of study material or patients; T. Novos, statistical analysis, provision of study material or patients; A.E. van Herwaarden, provision of study material or patients; D.M. Mueller, provision of study material or patients; M. Fassnacht, financial support, provision of study material or patients; F. Sweep, provision of study material or patients; A.R. Horvath, financial support, statistical analysis, administrative support, provision of study material or patients; G. Eisenhofer, financial support, statistical analysis, provision of study material or patients.

Authors' Disclosures or Potential Conflicts of Interest: Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest:

Employment or Leadership: None declared.

Consultant or Advisory Role: None declared.

Stock Ownership: None declared.

Honoraria: None declared.

Research Funding: This work was supported by the Deutsche Forschungsgemeinschaft (DFG), German Research Foundation Project number 314061271-CRC/TRR 205/1.

Expert Testimony: None declared.

Patents: None declared.

Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, preparation of manuscript, or final approval of manuscript.

References

- Lenders JW, Duh QY, Eisenhofer G, Gimenez-Roqueplo AP, Grebe SK, Murad MH, et al. Pheochromocytoma and paraganglioma: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab* 2014;99:1915-42.
- Rao D, Peitzsch M, Prejbisz A, Hanus K, Fassnacht M, Beuschlein F, et al. Plasma methoxytyramine: clinical utility with metanephrines for diagnosis of pheochromocytoma and paraganglioma. *Eur J Endocrinol* 2017;177:103-13.
- Eisenhofer G, Prejbisz A, Peitzsch M, Pamporaki C, Masjkur J, Rogowski-Lehmann N, et al. Biochemical diagnosis of chromaffin cell tumors in patients at high and low risk of disease: plasma versus urinary free or deconjugated o-methylated catecholamine metabolites. *Clin Chem* 2018;64:1646-56.
- Lenders JWM, Keiser HR, Goldstein DS, Willemsen JJ, Friberg P, Jacobs MC, et al. Plasma metanephrines in the diagnosis of pheochromocytoma. *Ann Intern Med* 1995;123:101-9.
- Weismann D, Peitzsch M, Raida A, Prejbisz A, Gosk M, Riestler A, et al. Measurements of plasma metanephrines by immunoassay vs liquid chromatography with tandem mass spectrometry for diagnosis of pheochromocytoma. *Eur J Endocrinol* 2015;172:251-60.
- Darr R, Pamporaki C, Peitzsch M, Miehle K, Prejbisz A, Peczkowska M, et al. Biochemical diagnosis of pheochromocytoma using plasma-free normetanephrine, metanephrine and methoxytyramine: importance of supine sampling under fasting conditions. *Clin Endocrinol (Oxf)* 2014;80:478-86.
- Griffin TP, Bogdanet D, Navin P, Callagy G, ÓShea PM, Bell M. The importance of standardisation of measurement and reference intervals for detection of pheochromocytoma and paraganglioma (PPGL). *Ir J Med Sci* 2018;187:993-8.
- Kim HJ, Lee JI, Cho YY, Lee SY, Kim JH, Jung BC, et al. Diagnostic accuracy of plasma free metanephrines in a seated position compared with 24-hour urinary metanephrines in the investigation of pheochromocytoma. *Endocr J* 2015;62:243-50.
- Boyd J, Leung AA, Sadrzadeh HS, Pamporaki C, Pacak K, Deutschbein T, et al. A high rate of modestly elevated plasma normetanephrine in a population referred for suspected PPGL when measured in a seated position. *Eur J Endocrinol* 2019;181:301-9.
- Peaston RT, Graham KS, Chambers E, van der Molen JC, Ball S. Performance of plasma free metanephrines measured by liquid chromatography-tandem mass spectrometry in the diagnosis of pheochromocytoma. *Clin Chim Acta* 2010;411:546-52.
- Pillai D, Callen S. Pilot quality assurance programme for plasma metanephrines. *Ann Clin Biochem* 2010;47:137-42.
- Eisenhofer G, Lattke P, Herberg M, Siegert G, Qin N, Darr R, et al. Reference intervals for plasma free metanephrines with an age adjustment for normetanephrine for optimized laboratory testing of pheochromocytoma. *Ann Clin Biochem* 2013;50:62-9.
- Clerico A, Ripoli A, Zucchielli GC, Plebani M. Harmonization protocols for thyroid stimulating hormone (TSH) immunoassays: different approaches based on the consensus mean value. *Clin Chem Lab Med* 2015;53:377-82.
- Clerico A, Ripoli A, Fortunato A, Alfano A, Carozza C, Corrales M, et al. Harmonization protocols for TSH immunoassays: a multicenter study in Italy. *Clin Chem Lab Med* 2017;55:1722-33.
- Grande Lac D, Van Uytanghe K, Reynders D, Das B, Faix JD, MacKenzie F, et al. Standardization of free thyroxine measurements allows the adoption of a more uniform reference interval. *Clin Chem* 2017;63:1642-52.
- Thienpont LM, Van Uytanghe K, De GL, Reynders D, Das B, Faix JD, et al. Harmonization of serum thyroid-stimulating hormone measurements paves the way for the adoption of a more uniform reference interval. *Clin Chem* 2017;63:1248-60.
- Nikooyeh B, Samiee SM, Farzami MR, Alavimajd H, Zahedrad M, Kalayi A, et al. Harmonization of serum 25-hydroxycalciferol assay results from high-performance liquid chromatography, enzyme immunoassay, radioimmunoassay, and immunochemiluminescence systems: a multicenter study. *J Clin Lab Anal* 2017;31:e22117.
- Bjerg LN, Halgreen JR, Hansen SH, Morris HA, Jorgensen NR. An evaluation of total 25-hydroxyvitamin D assay standardization: where are we today? *J Steroid Biochem Mol Biol* 2019;190:224-33.
- Iervasi G, Clerico A. Harmonization of free thyroid hormone tests: a mission impossible? *Clin Chem Lab Med* 2011;49:43-8.
- Vogeser M, Seger C. Pitfalls associated with the use of liquid chromatography-tandem mass spectrometry in the clinical laboratory. *Clin Chem* 2010;56:1234-44.
- Twentyman JM, Cradic KW, Singh RJ, Grebe SK. Ionic cross talk can lead to overestimation of 3-methoxytyramine during quantification of metanephrines by mass spectrometry. *Clin Chem* 2012;58:1156-8.
- Dunand M, Donzelli M, Rickli A, Hysek CM, Liechti ME, Grouzmann E. Analytical interference of 4-hydroxy-3-methoxymethamphetamine with the measurement of plasma free normetanephrine by ultra-high pressure liquid chromatography-tandem mass spectrometry. *Clin Biochem* 2014;47:1121-3.
- Peitzsch M, Adaway J, Eisenhofer G. Interference from 3-O-methyldopa with ultra-high performance LC-MS/MS measurements of plasma metanephrines: chromatographic separation remains important. *Clin Chem* 2015;61:993-6.
- Emms H, Farah G, Shine B, Boot C, Toole B, McFadden M, et al. Falsely elevated plasma metanephrine in patients taking midodrine. *Ann Clin Biochem* 2018;55:509-15.

25. Dirks NF, Vesper HW, van Herwaarden AE, van den Ouweland JM, Kema IP, Krabbe JG, Heijboer AC. Various calibration procedures result in optimal standardization of routinely used 25(OH)D ID-LC-MS/MS methods. *Clin Chim Acta* 2016;462:49–54.
26. Schotker B, Jansen EH, Haug U, Schomburg L, Kohrle J, Brenner H. Standardization of misleading immunoassay based 25-hydroxyvitamin D levels with liquid chromatography tandem-mass spectrometry in a large cohort study. *PLoS One* 2012;7:e48774.
27. Greaves RF, Ho CS, Hoad KE, Joseph J, McWhinney B, Gill JP, et al. Achievements and future directions of the APFCB mass spectrometry harmonisation project on serum testosterone. *Clin Biochem Rev* 2016;37:63–84.
28. Greaves RF. A guide to harmonisation and standardisation of measurands determined by liquid chromatography-tandem mass spectrometry in routine clinical biochemistry. *Clin Biochem Rev* 2012;33:123–32.
29. Greg MW, Myers GL, Lou GM, Kahn SE, Schonbrunner ER, Thienpont LM, et al. Roadmap for harmonization of clinical laboratory measurement procedures. *Clin Chem* 2011;57:1108–17.
30. Jones GR, Sikaris K, Gill J'. Allowable limits of performance' for external quality assurance programs - an approach to application of the Stockholm criteria by the RCPA quality assurance programs. *Clin Biochem Rev* 2012;33:133–9.
31. Panteghini M. Implementation of standardization in clinical practice: not always an easy task. *Clin Chem Lab Med* 2012;50:1237–41.
32. Eisenhofer G, Peitzsch M, Kaden D, Langton K, Mangelis A, Pamporaki C, et al. Reference intervals for LC-MS/MS measurements of plasma free, urinary free and urinary acid-hydrolyzed deconjugated normetanephrine, metanephrine and methoxytyramine. *Clin Chim Acta* 2019; 490:46–54.
33. Wright MJ, Thomas RL, Stanford PE, Horvath AR. Multiple reaction monitoring with multistage fragmentation (MRM3) detection enhances selectivity for liquid chromatography-tandem mass spectrometry analysis of plasma free metanephrines. *Clin Chem* 2015;61:505–13.
34. van Faassen M, Bischoff R, Eijkelenkamp K, De JW, van der Ley CP, Kema IP. In matrix derivatization combined with LC-MS/MS results in ultrasensitive quantification of plasma free metanephrines and catecholamines. *Anal Chem* 2020;92:9072–8.
35. Francini LC, Vazquez-Montes M, Budin T, Perera R, Dunand M, Grouzmann E, Beck-Popovic M. Pediatric reference intervals for plasma free and total metanephrines established with a parametric approach: relevance to the diagnosis of neuroblastoma. *Pediatr Blood Cancer* 2015;62:587–93.
36. Peitzsch M, Butch E, Lovorn E, Mangelis A, Furman W, Santana VM, et al. Biochemical testing for neuroblastoma using plasma free 3-O-methyldopa, 3-methoxytyramine, and normetanephrine. *Pediatr Blood Cancer* 2020;67:e28081.
37. Eisenhofer G, Deutschbein T, Constantinescu G, Langton K, Pamporaki C, Calsina B, et al. Plasma metanephrines and prospective prediction of tumor location, size and mutation type in patients with pheochromocytoma and paraganglioma. *Clin Chem Lab Med* 2020;59: 353–63.
38. CLSI. Defining, establishing, and verifying reference intervals in the clinical laboratory; approved guidelines. 3rd Ed: Wayne (PA): Clinical and Laboratory Standards Institute; 2010.