PAPER

External quality assurance schemes (EQUASs) and interlaboratory comparison investigations (ICIs) for the human biomonitoring of aromatic amines in urine as part of the quality assurance programme under HBM4EU†

Stefanie Nübler,[a] Therese Burkhardt,[b] Moritz Schäfer,[a] Johannes Müller,[a] Karin Haji-Abbas-Zarrabi,[a] Nikola Pluym, [b] Max Scherer,[b] Gerhard Scherer,[b] Marta Esteban-López,[c] Argelia Castaño,[c] Hans G. J. Mol,[d] Holger M. Koch,[e] Jean-Philippe Antignac,[f] Jana Hajslova,[g] Katrin Vorkamp [h] and Thomas Göen *[a]

Exposure to aromatic amines may occur *via* tobacco smoke, hair dyes or tattoo inks, but also in the workplace during certain manufacturing processes. As some aromatic amines are known or suspected carcinogens, human biomonitoring (HBM) is essential to assess their exposure. Aromatic amines were among the selected chemicals in HBM4EU, a European-wide project to harmonise and advance HBM within 30 European countries. For this purpose, the analytical comparability and accuracy of participating laboratories were assessed by a QA/QC programme comprising interlaboratory comparison investigations (ICIs) and external quality assurance schemes (EQUASs). This paper presents the evaluation process and discusses the results of three ICI/EQUAS rounds for the determination of aromatic amines in urine conducted in 2019 and 2020. The final evaluation included ten participants which analysed the following six targeted aromatic amines over three rounds: aniline, *ortho*-toluidine (TOL), 4,4′-methylenedianiline (MDA), 4,4′-methylenebis(2-chloroaniline) (MOCA), 2,4-diaminotoluene (2,4-TDA), and 2,6-diaminotoluene (2,6-TDA). Most participants achieved satisfactory and highly comparable results, although low quantification limits were required to quantify the parameters at the level of exposure in the general population. Hydrolysis of the sample followed by liquid–liquid extraction and subsequent analysis of the derivatised analytes by means of GC-MS/MS were preferred for the sensitive and precise determination of aromatic amines in urine. This QA/QC programme succeeded in establishing a network of laboratories with high analytical comparability and accuracy for the analysis of aromatic amines in Europe.

1 Introduction

Exposure of the environment and the human population to aromatic amines is mainly a result of anthropogenic activities.[1] Aromatic amines encompass a large group of compounds consisting of at least one aromatic ring and an amine group, with aniline being one of the simplest structures. They have been used in many agricultural and industrial processes including the manufacture of many consumer-oriented products and their applications, *e.g.* in the production of rubber,

[a]*Friedrich-Alexander-Universität Erlangen-Nürnberg, Institute and Outpatient Clinic of Occupational, Social, and Environmental Medicine, Henkestraße 9-11, 91054 Erlangen, Germany. E-mail: thomas.goeen@fau.de*

[b]*ABF Analytisch-Biologisches Labor, ABF GmbH, Semmelweisstr. 5, 82152 Planegg, Germany*

[c]*National Center for Environmental Health, Instituto de Salud Carlos III, Ctra. Majadahonda – Pozuelo km 2.2, 28220 Madrid, Spain*

[d]*Wageningen Food Safety Research, Wageningen University and Research, Akkermaalsbos 2, Wageningen, 6708 WB, Netherlands*

[e]*Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr-University Bochum (IPA), Bürkle-de-la-Camp-Platz 1, 44789 Bochum, Germany*

[f]*Oniris, INRAE, UMR 1329 Laboratoire d'Etude des Résidus et Contaminants dans les Aliments (LABERCA), F-44307 Nantes, France*

[g]*Department of Food Analysis and Nutrition, University of Chemistry and Technology Prague, Technicka 5, 16000 Prague, Czech Republic*

[h]*Department of Environmental Science, Aarhus University, Frederiksborgvej 399, 4000 Roskilde, Denmark*

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d4ay01309c

cutting oils, plastics, pesticides and pharmaceuticals; as intermediates and degradation products of azo dyes; as well as in hair and tattoo dyes.[2,3] Moreover, some aromatic amines are present in tobacco smoke, leading to higher exposure levels in active and passive smokers compared to non-smokers.[2,4,5] Internal exposure to aromatic amines can also be caused by contact to nitroarenes, azo dyes and aromatic diisocyanates, which are metabolised to aromatic amines in the human body.[6,7]

High acute exposure to aromatic amines can result in the production of methaemoglobin and thus may cause methaemoglobinemia.[8] The most significant health concern is, however, the development of cancer caused by chronic exposure to aromatic amines.[9]

Due to the different exposure routes, human biomonitoring (HBM) is a powerful tool for the holistic assessment of chemical exposure and risk.[10] In case of exposure to aromatic amines and their precursors, the determination of aromatic amines (free or conjugated) in urine is a well-established HBM approach.[11]

Aromatic amines were among the chemicals selected as first-priority substances in the European HBM Initiative HBM4EU.[12] This European-wide project was a joint effort of 30 countries and European Commission authorities, with the aims to harmonise and advance HBM in Europe and to improve risk management for chemicals in support of policymaking.[13–15] A major objective of the HBM4EU project was to establish a network of analytical laboratories across Europe that would produce high-quality and comparable HBM data for the prioritised substances.[16] A comprehensive Quality Assurance/Quality Control (QA/QC) scheme was designed consisting of several rounds of interlaboratory comparison investigations (ICIs) and external quality assurance schemes (EQUASs) for all prioritised substances. ICIs investigated the comparability of results between laboratories participating in the HBM4EU QA/QC programme. EQUASs were based on the comparison of the participants' results with those of selected expert laboratories that applied validated analytical methods and had experience with HBM of aromatic amines in population studies.[16]

This paper presents the QA/QC programme developed in HBM4EU for a range of aromatic amines, including the evaluation process, the results obtained, and the main challenges encountered. Originally, nine biomarkers of aromatic amines were included in the programme. However, due to the low participation in the first round, three biomarkers (p-aminophenol, N-acetyl-4-aminophenol, and p-phenylenediamine) were excluded. The final biomarkers were: aniline, ortho-toluidine (TOL), 4,4′-methylenedianiline (MDA), 4,4′-methylenebis(2-chloroaniline) (MOCA), 2,4-diaminotoluene (2,4-TDA), and 2,6-diaminotoluene (2,6-TDA) (Table 1). The structures of these six aromatic amines analysed in the ICI/EQUAS are shown in ESI Fig 1.† These compounds were identified as suitable biomarkers in HBM4EU, including substances with substantial (MDA, MOCA) and insufficient HBM information.[12] Aniline and TOL are widely present in the environment and general biomarkers of exposure to aromatic amines, while MDA, MOCA, 2,4-TDA and 2,6-TDA are mainly associated with exposure to plastic products.[11,12]

2 Materials and methods

2.1 Design of the HBM4EU QA/QC programme

The HBM4EU QA/QC programme offered three rounds of customised exercises between May 2019 and August 2020. In these exercises, two different assessment schemes were applied. An evaluation as EQUAS was preferred, but when the requirements to evaluate an individual biomarker in an EQUAS were not fulfilled, an assessment as ICI was conducted. The overall ICI/EQUAS design for all substance groups in HBM4EU is described in Esteban López et al.[16] and is depicted in Fig. 1. Two control materials (CMs), spiked with different concentrations of all target biomarkers (Table 1), i.e. CM_{low} and CM_{high}, had to be analysed in each round. A round was considered "passed" if the absolute value of the Z-score was ≤2 for the specific biomarker in both CMs. To be approved for the analysis of HBM4EU samples, a laboratory had to pass at least two ICI/EQUAS rounds.[17]

In response to two calls for laboratories to perform analysis of aromatic amines in HBM4EU, 18 laboratories expressed their interest, of which eleven laboratories (61%) from four countries (ESI Table 1†) finally registered for participation in the ICI/EQUAS programme. Participation was possible for the whole set of aromatic amine biomarkers (six) or for less. Laboratories were asked to report the measured concentrations of the CMs alongside the respective limits of quantification (LOQs).

For EQUAS evaluation, five expert laboratories had been selected prior to the ICI/EQUAS rounds according to selection criteria predetermined by the HBM4EU Quality Assurance Unit (QAU). The selection of these experts was based on the fact that these laboratories had experience in the determination of aromatic amines which was documented in peer-reviewed publications. In addition, the following selection criteria were considered: number of years of experience in the analysis of aromatic amines in urine, application of highly sensitive and selective analytical techniques with sufficiently low LOQs, application of isotope-labelled standards for quantification, availability of in-house validation reports, data on on-going intra-laboratory performance, ISO 17025 accreditation and success rate in round robin tests or comparative results in HBM studies. Four of the selected expert laboratories were from Germany and one was from the UK. All expert laboratories were also included as participants in the respective ICI/EQUAS rounds.

2.2 Control materials

2.2.1 Preparation of control materials. Control materials (CMs) were freshly prepared before each ICI/EQUAS round by ABF (Analytisch-Biologisches Forschungslabor; ABF GmbH, Planegg, Germany). For the preparation of the CMs, urine samples of several donors were pooled. All donors gave their informed consent for the use of their samples in the preparation of the CMs. The low and high spiking concentrations were selected with regard to the expected exposure levels in the general population and workers with specific exposure to amines or precursors, respectively,[2,7,18] representing the target populations in HBM4EU.

Table 1 Urinary biomarkers for aromatic amines included in the QA/QC scheme for HBM4EU

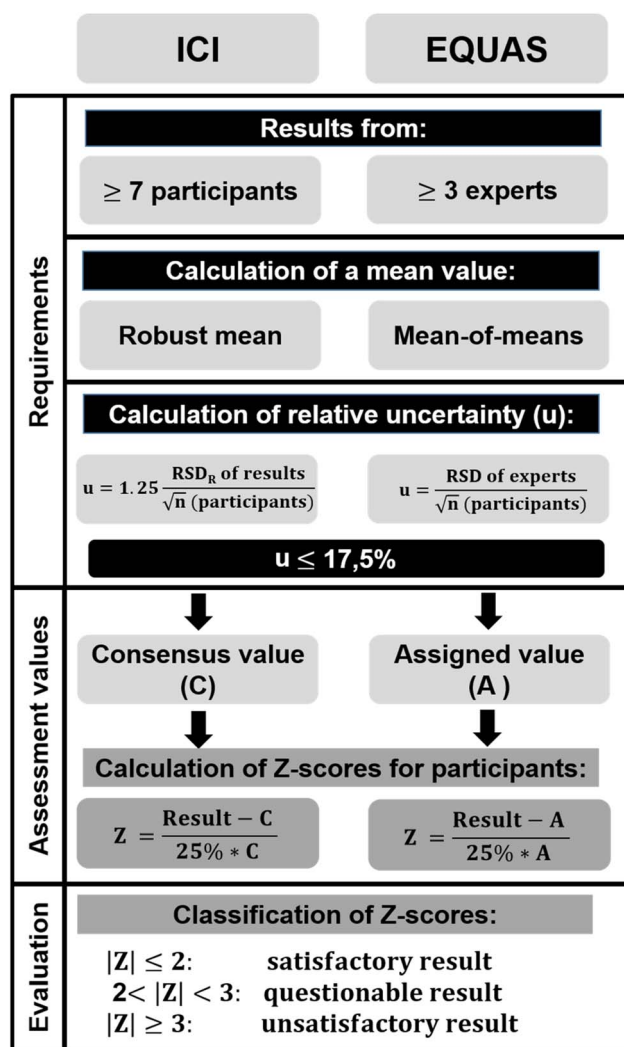| Abbreviation	Target biomarker	CAS no.	Formula
TOL	*ortho*-Toluidine[a]	95-53-4	C_7H_9N
Aniline	Aniline	62-53-3	C_6H_7N
MOCA	4,4′-Methylenebis(2-chloroaniline)	101-14-4	$C_{13}H_{12}Cl_2N_2$
2,4-TDA	2,4-Diaminotoluene	95-80-7	$C_7H_{10}N_2$
2,6-TDA	2,6-Diaminotoluene	823-40-5	$C_7H_{10}N_2$
MDA	4,4′-methylenedianiline	101-77-9	$CH_2(C_6H_4NH_2)_2$

[a] Other name: 2-methylaniline (2-MA).

Fig. 1 Evaluation of laboratory performance in ICI and EQUAS.

Accordingly, the final concentrations of the CMs were in a range between 0.3 ng mL^{-1} (TOL) and 52 ng mL^{-1} (2,4-TDA) for CM$_{low}$ as well as between 1.5 ng mL^{-1} (TOL) and 220 ng mL^{-1} (2,6-TDA) for CM$_{high}$ (ESI Table 2†).

Two different pools of non-smoker urine were mixed, adjusted to pH 4.0, and spiked with a stock solution to obtain the analytes in the final concentrations shown in ESI Table 2.† The stock solution was prepared from individual standards of different suppliers. As the stability of the free aromatic amines,

except for TOL, is poor, aniline, MOCA, 2,4-TDA, 2,6-TDA, and MDA were spiked as diacetyl-conjugates (synthesised and provided by Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany) and the participants were instructed to include a hydrolysis step that is typically carried out in the measurement of real samples.[19] For spiking of aniline, free aniline (no. 51788-1ML-F, Merck KGaA, Darmstadt, Germany) was applied in the first round while acetanilide (no. 397237, Merck KGaA, Darmstadt, Germany) was used in the second and the third round. TOL was considered stable at pH 4.0 and was therefore not spiked as a conjugate (no. 397237, Merck KGaA, Darmstadt, Germany). Each CM was aliquoted to a volume of 8 mL in 15 mL falcon tubes (polypropylene, Greiner AG, Kremsmünster, Austria) and stored at ≤ -20 °C until transport.

2.2.2 Characterisation of CMs. Analytical methods used to confirm homogeneity and stability of the CMs are given in detail in the supplementary material (ESI page 4 and ESI Tables 3 and 4†).

For homogeneity testing, ten tubes of CM$_{low}$ and CM$_{high}$, respectively, were randomly selected from the freezer (-20 °C) shortly after CM preparation. Samples were thawed, homogenised by vortex shaking, and analysed in duplicate. Results obtained were evaluated according to Fearn and Thompson[20] as well as Thompson *et al.*[21] A relative standard deviation (RSD) of 25% was chosen to be acceptable for homogeneity for CM$_{low}$ and CM$_{high}$, respectively. This threshold value was regarded as fit-for-purpose taking into account what is technically feasible and realistic in current routine practises with respect to the analytical method and concentration levels analysed.

For stability testing, three samples of each CM level were randomly selected from the freezer (-20 °C) and analyzed at $t = 0$ days (day of CM preparation) and on the day on which results were submitted ($t = 40$ days, 46 days or 60 days). Stability was assessed by comparing the means of the triplicates. Results were evaluated according to ISO 13528 (ref. 22) and the International Harmonised Protocol for the Proficiency Testing of Analytical Laboratories.[21]

2.3 Distribution of CMs

CMs were sent on dry ice. In the first and second rounds, three samples each of CM$_{low}$ and CM$_{high}$ were sent to participants. In the third round, each participant received one sample of CM$_{low}$ and CM$_{high}$. The selected expert laboratories always received six samples of each CM. At the time of shipment, participants

received a letter with instructions on sample handling, a sample-receipt form, a result-submission form and a method-information form. They were asked to perform a single analysis of each sample using the same procedure as for the analysis of the respective HBM4EU study samples. The deadline for submitting results was around five weeks (on average) after sample shipment. Participants were informed that some of the analytes in the CMs were spiked as conjugates and that a hydrolysis step should be applied. Furthermore, it was noted that dilution of the CMs by at least threefold should still allow for analysis within the following LOQs from literature: 1 ng mL^{-1} for 2,4-TDA; 2,6-TDA; MDA, and MOCA; 0.050 ng mL^{-1} for TOL; and 0.50 ng mL^{-1} for aniline.

2.4 Assessment of laboratory performance

The HBM4EU QA/QC programme followed specific requirements defined in four standard operating procedures for organisation (including selection criteria for experts), preparation of CMs, evaluation, and reporting of results, which were all available to the participants in the online library of the HBM4EU website. Details of the evaluation process in ICI/EQUAS were the same for different biomarker groups in HBM4EU and have been described in previous publications.[16,23] In addition to submitting the results of the analyses, the participants were also asked to provide information about the methods they used in each round. For this purpose, they were given a form that requested information on sample preparation, extraction, cleanup, derivatisation, instrumental analysis, use of an internal standard and calibration.

A schematic overview of the requirements, assessment values and evaluation in an ICI or EQUAS is shown in Fig. 1. For an ICI, the robust mean of participants' results was taken as a consensus value (C) if at least seven quantitative results from participating laboratories were available and the uncertainty of the consensus was within certain requirements (details see ref. 16). Robust statistics (details see ref. 24 and 25) were applied to reduce the influence of outliers on C. For an EQUAS, a minimum of three designated expert laboratories analysed six samples of each CM; their means were used to calculate the assigned value (A) as a mean-of-means when the uncertainty of A (u), calculated as relative standard deviation (RSD$_{experts}$) divided by the square root of the number of expert laboratories, did not exceed 17.5%. Due to the low number of quantitative results, classical statistics were used in EQUASs. EQUAS evaluations had the advantage that participants' performance could also be assessed when the number of participants was too small or their results were too heterogeneous for ICI evaluation.

Z-scores of the participants' results (x) were calculated as a measure of proficiency using C or A and the target RSD (σ_T) of 25% (see equation in Fig. 1). The value for the highest variability ($\sigma_T = 25\%$) considered acceptable for participants' results was set according to expert judgement, taking into account what is technically feasible in current routine practice.[16]

Only the absolute values of Z-scores were relevant for the assessment of laboratory performance and were categorised as follows:

- $|Z| \leq 2 \Rightarrow$ satisfactory result
- $2 < |Z| < 3 \Rightarrow$ questionable result
- $|Z| \geq 3 \Rightarrow$ unsatisfactory result.

For each participant, a round was considered "passed" for the specific biomarker if only satisfactory Z-scores were obtained in both CMs.

2.5 Statistical assessment

Statistical analyses were conducted using Excel 2013 (Microsoft Office Professional Plus 2013) for testing the CMs; for statistical analyses to compare the participants' results, SPSS was applied (IBM Corp. Released 2023. IBM SPSS Statistics for Windows, Version 29.0.2.0 Armonk, NY: IBM Corp).

3 Results and discussion

3.1 Homogeneity and stability of CMs

For each of the three ICI/EQUAS rounds, homogeneity and stability of the CMs were tested. The obtained values of the homogeneity measurements are shown in ESI Table 5.† The assessment of homogeneity as described in Section 2.2.2 demonstrated that the prepared CMs were sufficiently homogeneous for the application in the respective ICI/EQUAS rounds. The RSDs of the measured concentrations were highest in the first round for aniline (CM$_{low}$, RSD = 14.8%) and MOCA (CM$_{low}$/CM$_{high}$, RSD = 21.6%/17.5%), but still below the maximum acceptable RSD of 25%, and were clearly reduced by the third round. For the remaining four biomarkers, homogeneity was very high across all rounds (RSD = 2.0–9.1%). ESI Table 6† summarises the obtained values of the stability measurements. The mean relative difference of the measured concentrations on the day of freezing and after a period of 40–60 days at −20 °C was highest for MOCA in CM$_{low}$ (16.3–37.7%) and for aniline in CM$_{low}$ in the first round (14.3%) and ranged for the other aromatic amines between 0 and 14.3% across all rounds and CMs. For aniline, stability was improved from the second round on by spiking with acetanilide instead of free aniline. The stability test according to Section 2.2.2 showed sufficient stability of all biomarkers except for TOL. The fact that TOL failed the stability test in the first and the third round was mainly attributed to the low final concentrations in the CMs and the inherent analytical imprecision, which was not considered in the stability test. However, this fact was taken into account when calculating Z-scores. In previous experiments, TOL was stable in spiked urine samples over 3.5 years if the samples were adjusted to pH 4.0 (data not shown). For this reason, the CM for the ICI/EQUAS had been adjusted to pH 4.0 in the preparation phase.

3.2 Participation and evaluation results

In all rounds, 18 laboratories were invited to participate in the ICI/EQUAS for aromatic amines in urine and the participation rate of the laboratories slightly increased over the rounds. In the first round, nine laboratories registered and eight submitted results. In the second round, nine laboratories registered and eight participated, while in the last round all ten registered

laboratories from four countries submitted results. Table 2 shows the maximum number of participants who submitted results for the respective parameter across all rounds. Overall, MDA was quantified by all ICI/EQUAS participants ($n = 10$), followed by MOCA ($n = 9$) and the TDA isomers ($n = 8$), while the number of laboratories reporting results for TOL ($n = 6$) and aniline ($n = 5$) was the lowest, based on maximum participation over three rounds. In total, four participants reported results for all six biomarkers (ESI Fig. 2†), while one laboratory only participated in ICI/EQUAS for two aromatic amines (MOCA, MDA). Four experts participated in the first round, three experts in the second, and five in the final round. One possible reason for the higher participation rate in the ICI/EQUAS for MDA, 2,4-TDA and 2,6-TDA compared to the other biomarkers is that they had been selected for the occupational study on diisocyanate exposure in the HBM4EU project.[26] Laboratories wishing to participate in the diisocyanate study needed to meet the QA/QC requirements in HBM4EU.

The participants' results are summarised in Table 2 and presented in detail in ESI Table 7,† the latter containing the evaluation schemes, the evaluation results and the percentage of satisfactory, questionable, and unsatisfactory Z-scores in each round. Overall, the participants' results in each of the three ICI/EQUAS rounds were predominantly satisfactory. The percentage average of the number of satisfactory results (with $|Z| \leq 2$, see Section 2.4) over all evaluable aromatic amines ranged from 62.5% to 100%, and increased slightly from round 1 to round 3 for CM$_{low}$ and CM$_{high}$ (ESI Fig. 3†). Although the average percentage of satisfactory results for all parameters was almost the same across all rounds (88%), the increase in satisfactory results was higher for CMs with higher levels of aromatic amines (14.5% more satisfactory results in round 3 than in round 1) than for CMs with lower levels (8.4% increase). The training effect was therefore more pronounced at the higher concentrations, although the absolute values of the Z-scores did not significantly differ between CM$_{low}$ and CM$_{high}$ (Mann–Whitney U test, $U = 6102$, $Z = -1.112$, $p = 0.267$).

Over all rounds, the participants achieved the most comparable results for TOL and MDA, as the average of the robust relative standard deviation of the participants' results (study RSD$_R$) was below 20% for both aromatic amines (TOL: 15.3%, MDA: 16.6%). For 2,4-TDA and aniline (only one round), the mean study RSD$_R$ over all rounds was <23%, while it was clearly higher for MOCA (43.4%) and 2,6-TDA (53.3%). The determination of the lower levels of aromatic amines resulted in a higher mean study RSD$_R$ (34.0%) compared to the participants' results for the higher levels (mean study RSD$_R$ = 25.4%), which was to be expected given higher uncertainties close to the LOQs. Interestingly, the results of the participants varied the most in the 3rd round (mean study RSD$_R$ = 37.6%), while the mean study RSD$_R$ was lower in the 1st round (31.3%) and lowest in the 2nd round (19.5%). The third round might have been the most challenging one for the following two reasons. First, both CMs contained on average lower final concentrations of aromatic amines in the third round (mean = 30.2 ng mL^{-1}) than in the first (58.3 ng mL^{-1}) and the second round (62.6 ng mL^{-1}). Second, only one sample of each CM was sent to the participants in the last round, so that repetition variances were not balanced, as was possible in the 1st and the 2nd round. Nevertheless, this challenging last round was passed with the best Z-scores of all rounds by most participants, which could reflect a certain training effect. The high variation of results in the third round was mainly caused by strongly deviating results from different individual participants for the biomarkers TOL, MOCA and 2,6-TDA. In contrast, the variation of the participants' results in the first round was not influenced by strongly deviating individual results. In the second round, only one participant reported very different results for MDA. However, by applying robust statistics, as described in Section 2.4, the influence of individual outliers on the evaluation could be reduced.

The relative standard deviation of the expert results (RSD$_{experts}$, 11.7%) was on average less than half the mean study RSD$_R$ (29.7%) across all rounds and CMs, which illustrates the

Table 2 Overview of maximum participation, comparability of results and Z-scores in all three rounds[a]

Aromatic amine	CM	Maximum number of participants (of which experts)	Comparability of results (study RSD$_R$ in %)	Z-score (%)		
				Satisfactory	Questionable	Unsatisfactory
TOL	Low	6 (3)	17.7	94.4%	0%	5.6%
	High	6 (3)	12.9	94.4%	5.6%	0%
Aniline	Low	5 (3)	na	na	na	na
	High	5 (3)	22.5	100%	0%	0%
MOCA	Low	9 (5)	52.4	83.8%	8.3%	7.8%
	High	9 (5)	34.4	83.8%	8.3%	7.8%
2,4-TDA	Low	8 (5)	22.1	86.5%	4.7%	8.8%
	High	8 (5)	20.4	86.5%	4.7%	8.8%
2,6-TDA	Low	8 (5)	61.4	77.8%	4.2%	18.0%
	High	8 (5)	45.1	82.3%	4.7%	13.0%
MDA	Low	10 (5)	16.7	96.3%	0%	3.7%
	High	10 (5)	16.6	92.2%	4.2%	3.7%

[a] CM = control material; na = not applicable; study RSD$_R$ = robust relative standard deviation of participants' results.

expertise of the expert laboratories in measuring the respective aromatic amines, but might also be influenced by the fact that the experts could report mean values of six samples. However, not for all measurements, the uncertainty of the experts was low enough to enable an EQUAS evaluation (see ESI Table 7†).

The LOQs reported by the participants (including experts) ranged from 0.010 to 25.0 ng mL^{-1}, being lowest for TOL and highest for 2,4- and 2,6-TDA, with a substantial range for 2,4- and 2,6-TDA (ESI Table 8†). Over the three rounds, most laboratories reported constant LOQs, two participants reported higher LOQs in successive rounds (for aniline or for MOCA, 2,4-TDA, 2,6-TDA, and MDA), while three participants could improve their LOQ for one biomarker each (TOL, MOCA, or MDA). The recommended LOQs according to the literature (0.05 ng mL^{-1} for TOL, 0.5 ng mL^{-1} for aniline, and 1 ng mL^{-1} for 2,4-TDA, 2,6-TDA, MDA and MOCA) were not achieved by all participants. For MDA, all participants met the requirements, and for 2,4-TDA and 2,6-TDA, 83% of the measurements achieved the required LOQs. In contrast, only around 50% of the reported LOQs for MOCA, aniline, and TOL were in line with the given requirements.

3.3 Method features of participants and experts

The main features of the methods applied by the participants and the expert laboratories, which were also included as participants, are illustrated in ESI Fig. 4 and 5.† All laboratories hydrolysed the samples, but not all provided information on the method of hydrolysis. Two participants each used 10–11 M NaOH for 2–3 h at 95 °C, concentrated HCl for 2 h/10 h, and concentrated/3 M H$_2$SO$_4$ for 1.5 h at 90 °C/2 h, respectively. After pH adjustment, all participants and experts carried out a liquid–liquid extraction. The amount of the extracted sample ranged from 0.25 mL to 5 mL and was 2.5 mL on average. Derivatisation with subsequent GC-MS analysis was used by all experts, but only by two participants, while the application of heptafluorobutyric anhydride (HFBA) and pentafluoropropionic anhydride (PFPA) as derivatising agents was quite equally distributed among experts (three used HBFA, two applied PFPA) and participants (one used HBFA, one PFPA). In the GC-MS group, three experts and two participants carried out a single-quadrupole detection and two experts applied a triple-quadrupole detection. HPLC with triple-quadrupole mass-spectrometric detection was conducted by three participants. In addition, one expert used HPLC with triple-quadrupole mass-spectrometric detection for the determination of MOCA, 2,4-TDA, 2,6-TDA, and MDA, while aniline and TOL were analysed by GC-MS. Four experts and two participants prepared a matrix-matched calibration curve, whereas two participants used solvent standards. Two laboratories did not provide any information on this item. All except one laboratory stated that they carried out a multi-level calibration ($n = 9$). Predominantly, isotope-labelled internal standards (ISTDs) were used ($n = 8$) and the response was normalised to the respective ISTD ($n = 7$). Only one participant applied no ISTD and another participant used a different ISTD. In general, the use of a deuterium-labelled ISTD is also possible, but an unintended D–H exchange can occur under strongly acidic/basic conditions

during hydrolysis.[27] Since a deuterium-labelled ISTD was applied in the homogeneity and stability testing of TOL and MDA, possible D–H exchanges were monitored for each measurement but not observed. Altogether, despite a relatively small number of analytes and laboratories, a variety of methods was applied. While some features are generally established, such as the hydrolysis step, the liquid–liquid extraction, the use of ISTDs and matrix-matched calibrations, more variety existed in the instrumentation and the use of derivatisation agents.

Significant differences were found in the absolute values of the Z-scores for different methods applied. The Z-scores of the HPLC measurements were significantly higher ($p < 0.01$) than those of the GC measurements (Fig. 2a). This is consistent with the finding that participants using GC to detect aromatic amines also reported significantly lower ($p < 0.01$) LOQs than participants using HPLC (Fig. 2b). In the GC measurements, no statistical difference was observed between the Z-scores obtained with single-quadrupole and triple-quadrupole detection (Mann–Whitney U-test, $U = 2872$, $Z = -1.361$, $p = 0.175$) or between the Z-scores obtained with PFPA or HBFA as a derivatisation reagent (Mann–Whitney U-test, $U = 3036$, $Z = -1.426$, $p = 0.154$). Overall, GC proved to be more sensitive and reliable than HPLC for detecting the six aromatic amines investigated in the ICI/EQUAS.

3.4 Challenges in the EQUAS and ICI evaluation

Evaluating the results from the participating laboratories was challenging due to the small number of participants ($n \leq 10$) and experts ($n \leq 5$). Originally, it was planned to assess the participants' measurement results in all three rounds using the experts' results in an EQUAS. For most EQUAS evaluations, the mean results of three expert laboratories (minimum number of required experts) were applied to calculate the mean-of-means as A (see Section 2.4). In the first round, results of four expert labs were applied for the assessment of the participants' results for MDA in CM$_{low}$ and in the third round, the EQUAS evaluations of 2,4-TDA, 2,6-TDA, and MDA were based on the results of five expert laboratories. However, in the first two rounds, the uncertainty of the experts' consensus value was too high in some cases (>17.5%), so that four biomarkers in CM$_{low}$ (MOCA, 2,4-TDA, 2,6-TDA, MDA) and three biomarkers in CM$_{high}$ (2,4-TDA, 2,6-TDA, MDA) had to be evaluated as ICIs on the basis of the consensus value of the participants' results (C). In the first round, 2,6-TDA$_{high}$ was evaluated as ICI, although quantitative results from only five participants (at least seven participants required according to protocol) were available. The evaluation of the results for aniline was only possible in CM$_{high}$ of the third round using the EQUAS evaluation scheme. Overall, the first round was mainly and the third round was exclusively evaluated as EQUASs, while the results of the second round were mostly assessed as an ICI evaluation. Of a total 36 evaluations (six aromatic amines × two CMs × three rounds), 21 evaluations were feasible as EQUASs and ten as ICIs, while in five cases (aniline) no evaluation was possible (Table 2).

In order to compare the two different evaluation schemes, the consensus values of the participants' results were also calculated for cases in which an assessment as EQUAS was possible. The
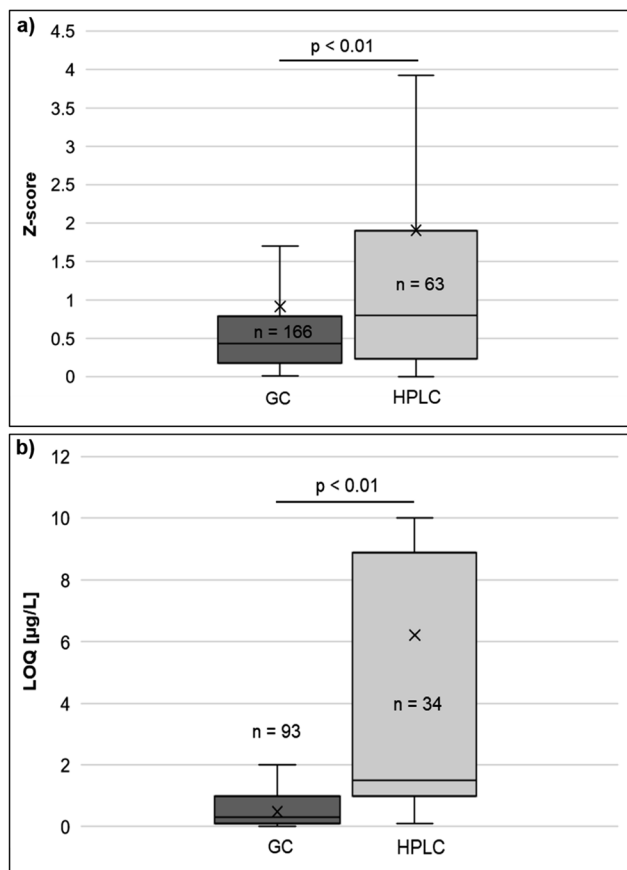
Fig. 2 Significantly different Z-scores (a) and LOQs (b) obtained over three ICI/EQUAS rounds for aromatic amines by measurement with GC compared to HPLC. A Mann–Whitney U-test was calculated to determine (a) if the Z-scores and if the LOQs (b) differed depending on the use of the instrument (GC vs. HPLC). The distributions differed between both groups (Kolmogorov–Smirnov $p < 0.05$). (a) There was a statistically significant difference in the Z-scores between GC users ($M_{rank} = 107.9$) and HPLC users ($M_{rank} = 137.6$), $U = 3931$, $Z = -3.009$, $p < 0.005$, using the exact sampling distribution of U. The effect size was however rather weak (determination coefficient $R^2 = 0.04$). In the boxplot for GC, two extreme outliers were excluded (Z-scores of 1823 and 79), although significance criteria ($p < 0.01$) would have still be fulfilled. (b) There was a statistically significant difference in the LOQs between GC users ($M_{rank} = 53.67$) and HPLC users ($M_{rank} = 92.26$), $U = 620$, $Z = -5.276$, $p < 0.001$, using the exact sampling distribution of U. The effect size was middle to strong (determination coefficient $R^2 = 0.22$).

relative difference of the mean values obtained by ICI (C) and EQUAS (A) calculated as $[(C - A)/A] \times 100\%$ was predominantly $\leq 10\%$ and, interestingly, it was greater in the high-level CMs than in the lower ones (Table 2). On average, the relative difference between A and C was only 7.5%, which shows that ICI and EQUAS generally lead to comparable Z-scores and enabled an equivalent evaluation of participants' performance for the measurement of the respective aromatic amines.

4 Conclusions

As part of the HBM4EU project, the designed and applied QA/QC programme assessed the European interlaboratory comparability

and accuracy of the analysis of six aromatic amines (aniline, TOL, MDA, MOCA, 2,4-TDA, and 2,6-TDA) in urine. In spite of two pan-European calls and the open-access character of the programme, only ten laboratories from three different European countries participated in the programme, while, for example, 29 laboratories participated in the ICI/EQUAS of the HBM4EU project for another group of carcinogenic substances, the polycyclic aromatic hydrocarbons.[28] Almost all of the participants in the QA/QC programme for aromatic amines achieved overall satisfactory and highly comparable results for the biomarkers in which they participated, although low quantification limits were required to quantify the biomarkers at the exposure levels of the general population. Thus, the QA/QC programme succeeded in establishing a network of laboratories with high analytical comparability and accuracy for the HBM analysis of aromatic amines in Europe. However, more efforts are needed to expand the number of qualified participants, possibly including capacity building measures. The current state of the art also revealed some challenges with certain biomarkers, for example high study RSD_Rs amongst the laboratories for MOCA (43.4%) and 2,6-TDA (53.3%). Furthermore, the results indicated that the analysis of derivatised analytes by GC-MS or GC-MS/MS was the method of choice in terms of low LOQs and high precision for the determination of aromatic amines in urine at the level of exposure of the general population.

Abbreviations

A	Assigned value
C	Consensus value
CM	Control material
EQUAS	External quality assurance scheme
HFBA	Heptafluorobutyric anhydride
HBM	Human biomonitoring
HBM4EU	European human biomonitoring initiative
ICI	Interlaboratory comparison investigation
ISTD	Internal standard
LOQ	Limit of quantification
MDA	4,4′-Methylenedianiline
MOCA	4,4′-Methylenebis(2-chloroaniline)
TOL	ortho-Toluidine
PFPA	Pentafluoropropionic anhydride
RSD	Relative standard deviation
study RSD_R	Robust relative standard deviation of participants' results
2,4-TDA	2,4-Diaminotoluene
2,6-TDA	2,6-Diaminotoluene

Data availability

The data supporting this article have been included as part of the ESI.†

Author contributions

Conceptualization: A. Castaño, M. Esteban-López, T. Göen. Methodology: H. Mol, J. P. Antignac, T. Burkhardt, A. Castaño,

D. Dvorakova, M. Esteban-López, T. Göen, J. Hajslova, H. M. Koch, M. Scherer, G. Scherer, K. Vorkamp. Formal analysis: T. Göen, S. Nübler, M. Schäfer, K. Haji-Abbas-Zarrabi. Investigation: T. Burkhardt, N. Pluym, S. Nübler, M. Schäfer, J. Müller, K. Haji-Abbas-Zarrabi. Resources: T. Göen, J. Müller. Writing – orig. draft: S. Nübler, T. Burkhardt, T. Göen. Writing – rev. & edit: K. Vorkamp, N. Pluym, M. Scherer, G. Scherer, D. Dvorakova, M. Esteban-López, H. Mol, H. M. Koch, J. P. Antignac, J. Hajslova, A. Castaño. Supervision: T. Göen. Project administration: A. Castaño, M. Esteban-López, T. Göen. Funding acquisition: A. Castaño, M. Esteban-López, T. Göen, H. Mol, H. M. Koch, J. P. Antignac, J. Hajslova, K. Vorkamp.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

References

1 IARC (International Agency for Research on Cancer), *Some Aromatic Amines and Related Compounds*, WHO Press, Lyon, France, 2022, vol. 127.

2 B. Kütting, T. Göen, U. Schwegler, H. Fromme, W. Uter, J. Angerer and H. Drexler, *Int. J. Hyg. Environ. Health*, 2009, **212**, 298–309.

3 G. Talaska, *J. Environ. Sci. Health, Part C: Environ. Carcinog. Ecotoxicol. Rev.*, 2003, **21**, 29–43.

4 F. Luceri, G. Pieraccini, G. Moneti and P. Dolara, *Toxicol. Ind. Health*, 1993, **9**, 405–413.

5 K. Riedel, G. Scherer, J. Engl, H.-W. Hagedorn and A. R. Tricker, *J. Anal. Toxicol.*, 2006, **30**, 187–195.

6 C. S. Eskilsson, R. Davidsson and L. Mathiasson, *J. Chromatogr. A*, 2002, **955**, 215–227.

7 C. Keen, M. Coldwell, K. McNally, P. Baldwin, J. McAlinden and J. Cocker, *Toxicol. Lett.*, 2012, **213**, 3–8.

8 K. V. Pathak, T.-L. Chiu, E. A. Amin and R. J. Turesky, *Chem. Res. Toxicol.*, 2016, **29**, 255–269.

9 P. Vineis and R. Pirastu, *Cancer Causes Control*, 1997, **8**, 346–355.

10 J. Angerer, U. Ewers and M. Wilhelm, *Int. J. Hyg. Environ. Health*, 2007, **210**, 201–228.

11 G. Sabbioni, *Chem. Res. Toxicol.*, 2017, **30**, 1733–1766.

12 K. Vorkamp, A. Castaño, J.-P. Antignac, L. D. Boada, E. Cequier, A. Covaci, M. E. López, L. S. Haug, M. Kasper-Sonnenberg, H. M. Koch, O. P. Luzardo, A. Osīte, L. Rambaud, M.-T. Pinorini, G. Sabbioni and C. Thomsen, *Environ. Int.*, 2021, **146**, 106082.

13 C. Ganzleben, J.-P. Antignac, R. Barouki, A. Castaño, U. Fiddicke, J. Klánová, E. Lebret, N. Olea, D. Sarigiannis, G. R. Schoeters, O. Sepai, H. Tolonen and M. Kolossa-Gehring, *Int. J. Hyg. Environ. Health*, 2017, **220**, 94–97.

14 J. L. Vicente, C. Ganzleben, R. Gasol, I. Marnane, L. Gilles, J. Buekers, J. Bessems, A. Colles, A. Gerofke, M. David, R. Barouki, M. Uhl, O. Sepai, I. Loots, A. Crabbé, D. Coertjens, M. Kolossa-Gehring and G. Schoeters, *Int. J. Hyg. Environ. Health*, 2023, **248**, 114111.

15 T. Santonen, S. Mahiout, P. Alvito, P. Apel, J. Bessems, W. Bil, T. Borges, S. Bose-O'Reilly, J. Buekers, A. I. C. Portilla, A. C. Calvo, M. de Alba-González, N. Domínguez-Morueco, M. E. López, I. Falnoga, A. Gerofke, M. Carmen González-Caballero, M. Horvat, P. Huuskonen, N. Kadikis, M. Kolossa-Gehring, R. Lange, H. Louro, C. Martins, M. Meslin, L. Niemann, S. P. Díaz, V. Plichta, S. P. Porras, C. Rousselle, B. Scholten, M. J. Silva, Z. Šlejkovec, J. S. Tratnik, A. Š. Joksić, J. V. Tarazona, M. Uhl, A. V. Nieuwenhuyse, S. Viegas, A. M. Vinggaard, M. Woutersen and G. Schoeters, *Int. J. Hyg. Environ. Health*, 2023, **249**, 114139.

16 M. Esteban López, T. Göen, H. Mol, S. Nübler, K. Haji-Abbas-Zarrabi, H. M. Koch, M. Kasper-Sonnenberg, D. Dvorakova, J. Hajslova, J.-P. Antignac, V. Vaccher, I. Elbers, C. Thomsen, K. Vorkamp, S. Pedraza-Díaz, M. Kolossa-Gehring and A. Castaño, *Int. J. Hyg. Environ. Health*, 2021, **234**, 113740.

17 L. Gilles, E. Govarts, L. Rambaud, N. Vogel, A. Castaño, M. Esteban López, L. Rodriguez Martin, G. Koppen, S. Remy, M. Vrijheid, P. Montazeri, L. Birks, O. Sepai, L. Stewart, U. Fiddicke, I. Loots, L. E. Knudsen, M. Kolossa-Gehring and G. Schoeters, *Int. J. Hyg. Environ. Health*, 2021, **237**, 113809.

18 K. Jones, P. D. Johnson, P. E. J. Baldwin, M. Coldwell, J. Cooke, C. Keen, A.-H. Harding, D. Smith and J. Cocker, *Ann. Work Exposures Health*, 2017, **61**, 383–393.

19 J. Lewalter and P. Biedermann, *The MAK-Collection for Occupational Health and Safety*, John Wiley & Sons, Ltd, 2012, pp. 67–105.

20 T. Fearn and M. Thompson, *Analyst*, 2001, **126**, 1414–1417.

21 M. Thompson, S. L. R. Ellison and R. Wood, *Pure Appl. Chem.*, 2006, **78**, 145–196.

22 ISO (International Organization for Standardization), ISO 13528, https://www.iso.org/obp/ui/#iso:std:iso:13528:ed-2:v2:en, accessed 9 April 2024.

23 S. Nübler, M. Esteban López, A. Castaño, H. G. J. Mol, K. Haji-Abbas-Zarrabi, M. Schäfer, J. Müller, J. Hajslova,

D. Dvorakova, J.-P. Antignac, H. M. Koch, L. S. Haug, K. Vorkamp and T. Göen, *Sci. Total Environ.*, 2022, **847**, 157481.

24 Analytical Committee, *Analyst*, 1989, **114**, 1693–1697.

25 Analytical Committee, *Analyst*, 1989, **114**, 1699–1704.

26 K. Jones, K. S. Galea, B. Scholten, M. Loikala, S. P. Porras, R. Bousoumah, S. Ndaw, E. Leese, H. Louro, M. J. Silva, S. Viegas, L. Godderis, J. Verdonck, K. Poels, T. Göen, R.-C. Duca, T. Santonen and HBM4EU Diisocyanates Study Team, *Int. J. Environ. Res. Public Health*, 2022, **19**, 8811.

27 P. Grocholska and R. Bąchor, *Molecules*, 2021, **26**, 2989.

28 S. Nübler, M. Esteban López, A. Castaño, H. G. J. Mol, J. Müller, M. Schäfer, K. Haji-Abbas-Zarrabi, J. Hajslova, J. Pulkrabova, D. Dvorakova, K. Urbancova, H. M. Koch, J.-P. Antignac, A. K. Saki, K. Vorkamp, T. Burkhardt, G. Scherer and T. Göen, *Int. J. Hyg. Environ. Health*, 2023, **250**, 114169.