An isotope dilution-liquid chromatography-tandem mass spectrometry (ID-LC-MS/MS)-based candidate reference measurement procedure (RMP) for the quantification of methotrexate in human serum and plasma

https://doi.org/10.1515/cclm-2022-1001
Received October 5, 2022; accepted February 3, 2023; published online February 15, 2023

Abstract

Objectives: To develop an isotope dilution-liquid chromatography-tandem mass spectrometry-(ID-LC-MS/MS)-based candidate reference measurement procedure (RMP) for quantification of methotrexate in human serum and plasma.

Methods: Quantitative nuclear magnetic resonance (qNMR) was used to determine absolute methotrexate content in the standard. Separation was achieved on a biphenyl reversed-phase analytical column with mobile phases based on water and acetonitrile, both containing 0.1% formic acid. Sample preparation included protein precipitation in combination with high sample dilution, and method validation according to current guidelines. The following were assessed: selectivity (using analyte-spiked samples, and relevant structural-related compounds and interferences); specificity and matrix effects (via post-column infusion and comparison of human matrix vs. neat samples); precision and accuracy (in a five-day validation analysis). RMP results were compared between two independent laboratories. Measurement uncertainty was evaluated according to current guidelines.

Results: The RMP separated methotrexate from potentially interfering compounds and enabled measurement over a calibration range of 7.200–5,700 ng/mL (0.01584–12.54 μmol/L), with no evidence of matrix effects. All pre-defined acceptance criteria were met; intermediate precision was ≤4.3% and repeatability 1.5–2.1% for all analyte concentrations. Bias was −3.0 to 2.1% for samples within the measuring range and 0.8–4.5% for diluted samples, independent of the sample matrix. RMP results equivalence was demonstrated between two independent laboratories (Pearson correlation coefficient 0.997). Expanded measurement uncertainty of target value-assigned samples was ≤3.4%.

Conclusions: This ID-LC-MS/MS-based approach provides a candidate RMP for methotrexate quantification. Traceability of methotrexate standard and the LC-MS/MS platform were assured by qNMR assessment and extensive method validation.

Keywords: isotope dilution LC-MS/MS; methotrexate; qNMR; reference measurement procedure; SI units; traceability.

Introduction

Methotrexate (MTX) is a folic acid antagonist that competes with folate-related enzymes involved in nucleotide synthesis, and thereby inhibits cell division [1, 2]. In addition, MTX has anti-inflammatory effects that appear to be independent of its effects on cellular division [2].

MTX is hydrophilic, displays poor solubility in acidic solutions and depends on active membrane transport for distribution inside the body [1, 3]. Protein binding in plasma is approximately 60% [1]. Important metabolites are MTX polyglutamate, 7-hydroxymethotrexate (7-OH-MTX) and...
2,4-diamino-N10-methylpterico acid (DAMPA) [1, 4]. The predominant route of elimination is renal while smaller amounts are excreted via bile and feces [1, 3, 5].

In low doses (7.5–25 mg/week), MTX is widely used in the treatment of inflammatory autoimmune diseases, such as rheumatoid and juvenile idiopathic arthritis and psoriasis [2, 6]. Higher doses (1–5 g/week) are used in the treatment of hematologic malignancies and solid organ tumors [1, 2, 6]. The drug is used to treat both adults and children [1], and may be combined with other anti-inflammatory, immuno-suppressive, or anti-cancer drugs [2, 3]. Various drug interactions have been described that hamper renal excretion of the agent (e.g., non-steroidal anti-inflammatory drugs, salicylates, proton-pump inhibitors, aminoglycosides, penicillin, and others) [1]. MTX has a narrow therapeutic range and can cause serious and potentially fatal side effects, including nephrotoxicity caused by precipitation of MTX in the renal tubules in acidic urine [1, 2, 6]. This in turn can cause further accumulation of MTX, which can induce nonrenal events, including hepatotoxicity, mucositis, immuno- and myelosuppression and others; these adverse effects may constitute a hurdle for the continuation of a planned anticancer therapy [1].

As the distribution and action of MTX depend on specific transporters and enzymes, genetic polymorphisms and or variable expression levels of these proteins may cause high variability of both MTX pharmacokinetics and pharmacodynamics [2, 3, 7]. As a consequence of the unpredictability of individual patient responses to standard doses and due to the narrow therapeutic range of MTX, therapeutic drug monitoring (TDM) is required to limit toxic and improve patients’ outcomes in high-dose MTX therapy [1, 6, 8, 9]. Together with surveillance for acute kidney injury (AKI, serum creatinine, urine output) and other measures, TDM of MTX influences clinical decision-making and therapeutic consequences [1]. Severe side effects including AKI can be prevented or mitigated by means of intensified hydration, intensified urinary alkalization (additional sodium bicarbonate), increased dosing of folinic acid (i.e., leucovorin rescue) and application of carboxypeptidase G2 (enzymatic cleavage of MTX to non-toxic DAMPA and glutamate), and/or hemodialysis in very severe cases [1]. Of note, hyperhydration, urinary alkalization and leucovorin rescue are an integral part of high-dose MTX-treatment even if no elevated MTX levels or adverse effects occur [1]. Sampling regimens for TDM of MTX differ depending on the specific treatment protocols [1]. However, some threshold levels have been established to assist toxicity risks, to guide dosing of leucovorin and to manage further supportive and therapeutic activities [1, 10].

Over the last 20 years, liquid chromatography-mass spectrometry (LC-MS) has been increasingly employed for quantitative assays in clinical laboratories [11–13]. This includes many quantitative LC-MS methods for quantification of MTX and its metabolites in biological matrices [14–21]. Despite these developments, individual methods may exhibit limitations, including laborious sample preparation (e.g., solid phase extraction) [6, 15], limited calibration range, and high method variability (inter- and intra-day precisions) [4, 6, 22, 23]. In addition, the traceability of clinical laboratory measurements is fundamental to ensure that results of such methods are comparable between laboratories and to reduce between-method variability [24, 25]. A lack of traceability and standardization arguably poses a risk for patients, since it can lead to misinterpretation if results are compared with reference ranges or therapeutic ranges that have been established with other methods [26]. Central to the concept of traceability is the availability of defined reference materials and reference measurement procedures (RMPs) [27]. However, the Joint Committee for Traceability in Laboratory Medicine (JCTLM), which was established to achieve assay standardization and the global harmonization of clinical laboratory test results [28], currently has no reference materials or RMP listed in their database for MTX [29]. Thus, there remains an unmet need for a standardized approach to accurately and precisely measure MTX levels in patients to improve TDM and to help inform treatment decisions.

The current study sought to develop and validate an isotope dilution-liquid chromatography-tandem mass spectrometry (ID-LC-MS/MS)-based candidate RMP for MTX quantification in human serum and plasma. The reference material was characterized using quantitative nuclear magnetic resonance (qNMR), a technique that can be used to assess the absolute quantity of a standard substance, thereby enabling traceability to SI units – an acceptable approach for National Metrological Institutes (NMIs) that develop and maintain measurement standards [25, 30, 31].

Materials and methods
A detailed account of the methods, including a full list of materials and equipment used, can be found in Supplementary Material 1.

Materials

Chemicals and reagents: LC-MS-grade reagents including acetonitrile, methanol, isopropanol, and formic acid (all Biosolve, Valkenswaard, The Netherlands), and sodium hydroxide (NaOH, ACS grade; all Sigma-Aldrich, Taufkirchen, Germany) were used. Water that was purified using a Millipore Milli-Q® Direct 16 system (Merck KgaA,
DAMPA, leucovorin, dihydrofolic acid, tetrahydrofolic acid, and LC-MS, 13C2H3-MTX (C360, Lot No. TMALS-12-134-B1), and 7-OH-MTX were stored in darkened laboratories and indoor lighting.

General requirements for laboratory equipment: A detailed list of the equipment used, as well as a description of the method presented, is provided in Supplementary Material 1.

Due to the light sensitivity of MTX (i.e., ultraviolet [UV] light >290 nm), all MTX solutions were prepared and stored in brown Falcon tubes (SafeSeal reaction tube, PP, Sarstedt, Nymbrecht, Germany) or vials (15/50 mL centrifuge tubes, polypropylene [PP], VWR), reaction tubes (NMR tubes [5 mm diameter, 8 inch long]) were purchased from Sigma-Aldrich and/or BCBW3152) were obtained from Sigma-Aldrich. NMR tubes (5 mm diameter, 8 inch long) were purchased from Sigma-Aldrich and/or Euroisotop (a subsidiary of Cambridge Isotope Laboratories, Inc, Hadfield, United Kingdom).

qNMR for determination of the purity of the standard material: qNMR measurements were performed on a Jeol 600 MHz NMR spectrometer (Jeol Ltd, Tokyo, Japan) with a He-cooled cryoprobe. Single-Pulse 1HNMR measurements were performed on a Jeol 600 MHz NMR spectrometer operating in positive electrospray ionization mode (ESI+). The widespread use of full NMR sequences (Supplementary Material, Figures 3 and 4) were utilized such as J-res, E_COSY Phase and TOCSY in order to exclusively rule out any ambiguities of chemical shifts belonging to methotrexate-like molecules/additional organic impurities, etc. Additional details about NMR acquisition and FID processing parameters can be found in Supplementary Material 2.

Preparation of calibrators and quality control (QC) samples: Two individual primary stock solutions were prepared for the matrix-based calibrators; these were diluted to provide the working and spike solutions. Stock solutions were prepared by weighing 10 mg of MTX on an ultra-microbalance and dissolving in 10 mL of methanolic 0.1 mol/L NaOH in a 50 mL Falcon tube. The concentration of each primary stock solution was calculated based on the purity of the reference material (0.91 ± 0.3%, determined by qNMR), and the weighed amount.

Each of the two primary stock solutions were diluted with 15% methanol (v/v in Biosolve-water) to provide working solutions with a concentration of 20 and 40 μg/mL. Stock and working solutions were used to prepare eight calibrator spike solutions of increasing concentrations ("calibrator levels"). The final matrix-based calibrators (concentration range: 7.200–5,700 ng/mL [0.01584–12.54 μmol/L]) were prepared using a 1 + 19 dilution (v + v) in commercial analyte-free human serum.

Four levels (24, 420, 3,000 and 4,200 ng/mL) of matrix-based QC material were prepared in the same way as described for the calibrator levels, using a third primary stock solution. To monitor for systematic drifts a native serum-based sample, pooled from left-over commercial external quality assessment (EQA) materials, with an MTX concentration in the middle of the calibration range (approximately 1,570 ng/mL) was used to generate a control chart; results were considered acceptable if they were within two standard deviations (SD) of the initial control chart measurement.

The spiked material was stored at −20 °C for a maximum period of 20 weeks. Native matrices were stored at −80 °C. Samples were thawed and brought to room temperature before use.

Preparation of the ISTD solution: An ISTD stock solution was prepared by weighing 2.0 mg of 13C2H3-MTX on an ultra-microbalance and dissolving in 20 mL of methanolic 0.1 mol/L NaOH in a 50 mL Falcon tube. A final ISTD solution (concentration of 900 ng/mL) was prepared by diluting 90 μL of the stock solution in 9,910 μL water.

Sample preparation: Serum and plasma (Li-heparin plasma, K3-EDTA plasma, and K2-EDTA plasma) were used as sample matrices. Samples used for validation were generated by spiking appropriate analyte amounts into commercial analyte-free serum and plasma matrices. Native samples for the method comparison study were exclusively anonymized leftover samples.

Samples were prepared in 1.5 mL reaction tubes by adding 100 μL of sample specimen (native/spiked/calibrator/QC samples) and 200 μL ISTD solution. Samples were vortexed and equilibrated on a shaker (Eppendorf Thermomixer® C) at 300 rpm at room temperature for 30 min. Twenty-five micro litre of the mixed samples were transferred to a new reaction tube and proteins were precipitated by adding 100 μL methanol. Samples were vortexed and kept for 30 min at −20 °C, before being diluted with 875 μL 0.1% formic acid in water and centrifuged for 30 min at 32,000 × g at 4 °C. Afterwards, 150 μL of the supernatants were transferred to a HPLC-vial with micro-insert for measurement.

Liquid chromatography-mass spectrometry (LC-MS): Chromatographic separation was performed using an Agilent 1290 Infinity II LC system (Santa Clara, California, USA). Analytes were detected using an AB Sciex Q-Trap® 6500+ mass spectrometer (Framingham, Massachusetts, USA) with a Turbo V ion source.

Details regarding LC and MS parameters and system setup are provided in Supplementary Material 1.

Chromatographic separation was achieved in reversed-phase mode on a biphenyl analytical column (Restek Raptor Bipheryl 2.7 μm, 100 × 2.1 mm), fitted with a Restek Raptor EXP Guard Column (2.7 μm, 5 × 2.1 mm), with water and acetonitrile, both containing 0.1% formic acid, as eluents. Briefly, 5 μL of the prepared sample were injected and MTX was separated using a flow rate of 0.4 mL/min at a column temperature of 30 °C in a 10 min gradient program. The analyte eluted at a retention time of 3.36 ± 0.20 min.

MTX was detected by multiple reaction monitoring with the mass spectrometer operating in positive electrospray ionization mode (ESI+ mode).

As quantifier, the transition of mass-to-charge ratio (m/z) 455.1 → 308.1 was chosen and associated with the corresponding ISTD transition (13C2H3-MTX m/z 459.0 → 312.2). Additional qualifier transitions (MTX m/z 455.1 → 175.2 and 13C2H3-MTX m/z 459.0 → 175.1) facilitated an assessment
(A) SST2: Separation of 7-OH-MTX:

![Chromatogram of SST2 sample](image)

(B) Cal1:

![Chromatogram of Calibration Sample](image)

(C) Native Sample (c = 65.03 ng/mL = 0.1413 μM):

![Chromatogram of Native Sample](image)

Figure 1: Methotrexate LC-MS/MS derived analytical readouts. (A) Chromatogram of SST2 sample with 5,700 ng/mL MTX and 12 μg/mL 7-OH-MTX spiked in neat solution (left) and internal standard (right); (B) chromatogram showing the lowest calibrator peak with a concentration of 7,200 ng/mL MTX spiked in serum (left) and internal standard (right); (C) chromatogram of MTX at concentration of 65.03 ng/mL (0.1413 μM) in a native serum patient sample (left) and internal standard (right). 7-OH-MTX, 7-hydroxymethotrexate; Cal1, calibration 1; cps, counts per second; ISTD, internal standard; MTX, methotrexate; SST2, system suitability test 2.

for potential interferences. Furthermore, transition of 7-OH-MTX (m/z 471.1 → 324.2) as the main metabolite was monitored in the system suitability test to ensure complete chromatographic separation (see example chromatograms in Figure 1).

**System suitability test (SST):** To assess system performance and ensure the long-term stability of the method, an SST was performed, examining sensitivity, carryover, and chromatographic resolution before every sequence. Two levels (SST1 and SST2), corresponding to the analyte...
concentration within the processed calibrator levels 1 and 8, were prepared in 10% methanol. Additionally, SST2 contained 100 ng/mL 7-OH-MTX.

Results of the SST are considered acceptable if the signal-to-noise ratio of the quantifier transition is $\geq 10$ for SST1. In addition, SST2 must show a MTX retention time of 3.36 min ($\pm 0.2$ min), and MTX must be baseline-separated from 7-OH-MTX with a chromatographic resolution ($R$) $\geq 2.0$.

To assess for potential carryover, the SST2 injection was followed by two solvent blank injections. For an acceptable SST result, the analyte peak area of the first blank must be $\leq 20\%$ of the analyte peak area of SST1.

**Calibration, structure of analytical series and data processing:** Calibration of the system was performed using the calibrators as described in detail in Preparation of calibrators and quality control (QC) samples. The calibrators were prepared once and injected in increasing concentration within the processed calibrator levels 1 and 8, prepared in 10% methanol. Additionally, SST2 contained 100 ng/mL 7-OH-MTX.

Results of the SST are considered acceptable if the signal-to-noise ratio of the quantifier transition is $\geq 10$ for SST1. In addition, SST2 must show a MTX retention time of 3.36 min ($\pm 0.2$ min), and MTX must be baseline-separated from 7-OH-MTX with a chromatographic resolution ($R$) $\geq 2.0$.

To assess for potential carryover, the SST2 injection was followed by two solvent blank injections. For an acceptable SST result, the analyte peak area of the first blank must be $\leq 20\%$ of the analyte peak area of SST1.

**Method validation**

The assay was validated and measurement uncertainty was determined as previously described by Taibon et al. [27], and in accordance with the following guidelines: the Clinical & Laboratory Standard Institute Guidelines C62A Liquid Chromatography-Mass Spectrometry Methods [32], the International Conference on Harmonization guidance document Harmonised Tripartite Guideline Validation of Analytical Procedures: Text and Methodology Q2 (R) [33] and the Guide to the expression of uncertainty in measurement [34].

**Selectivity:** To assess selectivity, the relevant structurally related compounds 7-OH-MTX, DAMPA, folic acid, leucovorin, dihydrofolic acid, and 5-methyl tetrahydrofolic acid were spiked in commercial analyte-free native human serum pool and the MTX and ISTD quantifier traces were checked for interference from these substances. Furthermore, these substances were spiked in neat solution (i.e., 10% methanol) together with MTX and the ISTD to demonstrate their baseline separation from the analyte and ISTD.

To test for potential interfering matrix signals in the analyte quantifier and qualifier transition, three different native human serum pools were checked at the expected retention time window. Additionally, analyte-free human serum was spiked with ISTD only to check for residual unlabelled analyte within the stable isotope-labeled ISTD.

**Specificity/matrix effects:** Potential matrix effects were investigated using a qualitative post-column infusion experiment. Quantitative analysis based on the comparison of absolute areas of analyte and ISTD in different sample sets was also performed.

In the post-column infusion experiment, a neat solution of MTX and ISTD (in 0.1% formic acid in 10% methanol) was infused at a flow rate of 10 $\mu$L/min via a T-piece into the HPLC column effluent, prior to entering the MS/MS system, to generate a stable analyte background MS signal. The change in this background signal was then measured after injecting a processed blank matrix sample (serum, Li-heparin, K2-EDTA plasma, and K2-EDTA plasma). Changes in the analyte signal indicated a matrix component-mediated effect on the degree of ionization of the analyte.

In the quantitative analysis, which was based on Matuszewski et al. [35], two sample sets were prepared at four different concentration levels (QC1–QC4, see Preparation of calibrators and quality control (QC) samples) in neat solution (set1) and in native human serum pool (set2). Set1 was prepared by spiking the appropriate amount of analyte in neat solution (0.1% formic acid in 10% methanol) and then diluting to the final concentration of processed samples. Set2 was prepared by processing matrix samples and spiking with analyte in the final dilution step after sample extraction.

The matrix effect (ME) was evaluated by comparison of analyte area, internal standard area, and area ratio as follows: $\text{ME} [\%]=\frac{\text{set2/ set1}}{\times 100}.

Recoveries were reported as the percentage of recovery of the measured concentration relative to the nominal concentration.
Engel et al.: An ID-LC-MS/MS-based candidate RMP for methotrexate quantification

Linarity: The preferred regression model for calculation was determined based on three independently prepared sets of MTX calibrators as outlined in Preparation of calibrators and quality control (QC) samples. The calibration range was extended by ±20%, by including two additional calibrators (final concentration of 5,800 and 6,840 ng/mL MTX). The peak area ratio of analyte vs. ISTD was plotted against the respective analyte concentration (ng/mL), and the correlation coefficient and residuals were determined for each curve.

Furthermore, linearity was demonstrated based on the recovery of serially diluted native samples. Therefore, a low concentrated (level 1) and a high concentrated (level 11) native sample were used to prepare the following nine mixtures: 9 + 1, 8 + 2, 7 + 3, 6 + 4, 5 + 4, 6 + 3, 7 + 2 + 8, and 1 + 9 v/v.

Lower limit of measuring interval (LLMI) and limit of detection (LoD): Precision and accuracy at LLMI were determined using six independently prepared replicate samples at the lowest calibrator level (7,200 ng/mL) and had to meet the performance specifications of the RMP. In addition, the signal-to-noise ratio was evaluated.

The limit of detection (LoD) was estimated by determining the mean and SD of blank matrix samples (maximum signal intensity of baseline in the peak region of the analyte, 10 independent samples from the precision experiment) and calculating LoD as the mean +3 SD with the mean peak height of calibrator 1 (n=10 samples) serving as quantification reference.

Precision and accuracy: Precision and accuracy were evaluated as described by Taibon et al. [27]. Accuracy was estimated using four concentration levels (QC1–QC4, see Preparation of calibrators and quality control (QC) samples) covering the measuring range, which were spiked in human serum, native Li-heparin plasma, K_{2}EDTA plasma, and K_{2}EDTA plasma. Precision was assessed with spiked serum (QC1–QC4, see Preparation of calibrators and quality control (QC) samples) and two native patient samples (approximately 150.0 and 3,750 ng/mL MTX), encompassing the medical decision points (MDP) for initiation and cessation of leucovorin and carbboxypeptidase rescues [1].

In brief, precision was assessed daily in a five-day validation analysis using two individual calibrator preparations for two measurement sequences (Part A and Part B). Samples were prepared in triplicate for each part and injected twice (n=12 measurements per level and day and n=60 measurements per five days). Repeatability assessments involved within-injection and within-preparation variability, while intermediate precision included between-calibration and between-day variability. Repeatability and intermediate precision were expressed as standard deviation (SD) and coefficient of variation (CV). Variability was determined using an ANOVA-based variance-components analysis.

Accuracy was assessed in a two-part, single-day experiment similar to precision (i.e., n=12 measurements per level). In addition, dilution integrity was shown using two spiked samples at concentration levels of 20,000 and 100,000 ng/mL in the serum and plasma matrices. After 1 + 9 v/v dilution with analyte-free serum, triplicate samples were prepared on a single day (n=3 per level). Accuracy was reported as (a) the percentage of recovery of the measured concentration relative to the final concentration of the spiked analyte in the individual sample, and (b) as mean bias per level.

Sample stability: Stability of processed samples on the autosampler was investigated for six days at 8 °C, with four calibrator concentration levels (QC1–QC4, see Preparation of calibrators and quality control (QC) samples, n=2 sample preparations per level) being re-measured daily. Stability of spike solutions stored at −20 °C was evaluated at three concentration levels (24, 420, and 4,200 ng/mL, n=2 sample preparations per level) over an 11-week period (re-measured weekly). Stability of matrix-based spiked calibrator and control materials stored at −20 °C was evaluated at four concentration levels (QC1–QC4, see Preparation of calibrators and quality control (QC) samples, n=3 sample preparations per level) over a 21-week period (re-measured weekly until week 12, bi-weekly afterwards). The closeness of agreement between the measured value for the stored samples and the value from freshly prepared samples of MTX was reported. The total error was used as an acceptance criterion, and calculated based on the results from precision.
and trueness experiment, resulting in a TE of ±10%. Stability can be ensured for a measurement interval of 2–28 days for y = 1 day, and for a measurement interval of >4 weeks for y = 1 week.

**Equivalence of results between independent laboratories:** To assess the equivalence of RMP results between two independent laboratories (Labor Berlin – Charité Vivantes Services GmbH, Berlin [Site/Laboratory 1]; Roche Diagnostics GmbH, Penzberg [Site/Laboratory 2]), a comparative study was performed using 194 native anonymized residual patient samples (130 native serum samples and 64 native plasma samples) as described in the Supplementary Material 1. In addition, a three-day precision experiment was performed at Laboratory 2 based on the experimental design described above. All samples were provided by Laboratory 1. The method was applied as described within Supplementary Material 1.

Results from the Labor Berlin site were also compared with multiple proficiency testing schemes from six different commercially available providers. From August 2020 to May 2022, 154 external quality-assessment samples were analyzed up to bi-weekly and compared with over 100 different clinical labs worldwide using their routine methods (mostly automated immunoassays with turbidimetric, fluorescence, chemiluminescence or photometric detection, partially enzyme-assisted; ≤10 other LC-MS methods).

**Uncertainty of measurements:** The uncertainty of measurements was assessed according to the GUM [34] for the following parameters: (a) qNMR target value assignment of the primary reference material; (b) preparation of calibrator materials; and (c) LC-MS/MS method. A detailed description is provided in Supplementary Material 3.

**Results**

**qNMR for determination of the purity of the standard material**

Six individual qNMR-experiments (Supplementary Material 2, Figure 2), involving six individual weighings of the analyte (MTX, DRE-C15056900, LGC/Dr. Ehrenstorfer) and methyl 3,5-dinitrobenzoate as standard, yielded a final content value of 89.1 ± 0.3% (k=1). The remaining 10.9% can, therefore, be attributed mainly to water and inorganic salts. Since we have utilized a polar aprotic NMR solvent with a high dielectric constant, these inorganic impurities render themselves readily soluble and allow the exact, true and absolute quantification of methotrexate in this material and from any other commercial source. Traceability to the SI-unit kilogram was established by using qNMR ISTDs that are directly traceable to qNMR-certified reference material (NIST PSI) [36].

**Selectivity**

Use of a biphenyl reversed-phase column in combination with the employed mobile phases minimized matrix effects and isobaric interferences. There was successful baseline separation of MTX from potentially interfering compounds (e.g., 7-OH-MTX as the most critical compound was separated with an average resolution of 3.6). This was evidenced by a lack of interfering matrix signals in the MTX-quantifier and qualifier transition or in the ISTD transition at the retention time of the analyte (3.37 min; Figure 2). The mean peak area of the remaining unlabeled MTX derived from the 13C2H3-MTX ISTD was well below 10% of the area of the analyte at the limit of quantitation.

**Specificity/matrix effects**

In the post-column infusion analysis investigating potential matrix effects, none of the matrices tested demonstrated ion suppression or enhancement in the region of the retention time of MTX and the ISTD.

In the quantitative investigation of matrix effects based on Matuszewski et al. [35], no matrix effect was observed. Mean matrix effects for the peak areas in serum vs. neat were 99–105% for the analyte, 99–105% for the ISTD, and area ratios were 99–101%. The results demonstrate the compensatory effect of the ISTD and prove that no matrix effect was present.

**Linearity**

Linearity was demonstrated in native serum-based calibration curves that exhibited random and equal distribution of residuals in a linear and a quadratic regression model (Figure 3). Thus, the simpler linear regression with $1/x^2$ weighting was chosen for assay calibration. Correlation coefficients were $\geq 0.999$ for each calibration curve.

Linearity of the method was also confirmed using serially diluted samples, with results demonstrating a linear dependence and a correlation coefficient of $\geq 0.999$. The deviation of measured concentration vs. calculated calibration curve ranged from −2.7 to 2.6%.

**Lower limit of measuring interval (LLMI) and limit of detection (LoD)**

Using six replicate spiked samples at the concentration of the lowest calibrator level, bias and precision were −2.1 and 2.3%, respectively. Signal-to-noise ratio was $\geq 20$ for all samples (LLMI=7.200 ng/mL). The LoD was estimated from blank matrix samples and found to be 1.128 ng/mL.
Precision and accuracy

Assessment of intermediate precision, including variances as between-day, -calibration, -preparation and -injection, demonstrated CVs of 1.6−4.3%. Repeatability CV range was 1.5−2.1% over all concentration levels (Table 1).

In the absence of certified secondary reference materials, accuracy was assessed using four levels of spiked serum and plasma samples. The bias for all concentration levels (n=12 for spiked QC samples) ranged from 0.1 to 2.1% in serum, −0.5 to 0.2% in native Li-heparin plasma, −3.0 to 1.2% in K$_2$-EDTA plasma, and −0.5 to 0.8% in K$_3$-EDTA plasma. Bias for two levels of diluted samples in serum and plasma matrices was 0.8−4.5% (Table 2). The confidence intervals of the bias within the different matrices overlap for all values except the lowest value in the K$_3$ EDTA plasma matrix, which has a slightly larger bias.

Sample stability

The stability of processed samples on the autosampler was demonstrated for five days at 8 °C, with recoveries of 98−106% compared with freshly prepared samples. In terms of stability at −20 °C, methanolic spike solutions were stable for 10 weeks, with recoveries of 97−106% compared with freshly prepared solutions. Spiked serum control samples were stable for 20 weeks, with recoveries of 95−107% compared with freshly prepared samples.

---

**Table 1:** Detailed precision performance obtained by VCA analysis in serum samples (n=60 measurements).

<table>
<thead>
<tr>
<th>Variance source</th>
<th>Nominal concentration</th>
<th>Native patient sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>24.00 ng/mL</td>
<td>420.0 ng/mL</td>
</tr>
<tr>
<td>Intermediate</td>
<td>3.6</td>
<td>1.7</td>
</tr>
<tr>
<td>Between-day</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Between-calibration</td>
<td>2.9</td>
<td>0.4</td>
</tr>
<tr>
<td>Repeatability</td>
<td>2.1</td>
<td>1.7</td>
</tr>
<tr>
<td>Between-preparation</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>Between-injection</td>
<td>2.0</td>
<td>1.7</td>
</tr>
</tbody>
</table>

CV, coefficient of variation; VCA, variance component analysis. The coefficients of variation for repeatability and intermediate precision, which were determined from the individual variances, are printed in bold.

**Table 2:** Bias and 95% CI of QC levels and samples in native serum or lithium heparin plasma, K$_2$-EDTA plasma or K$_3$-EDTA plasma (n=12 measurements for each sample; n=6 sample preparations, n=2 injections).

<table>
<thead>
<tr>
<th>Bias evaluation results</th>
<th>Mean bias, %</th>
<th>SD, %</th>
<th>95% CI, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human native serum</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 1</td>
<td>2.1</td>
<td>2.5</td>
<td>0.7 to 3.5</td>
</tr>
<tr>
<td>Level 2</td>
<td>0.3</td>
<td>1.5</td>
<td>−0.6 to 1.2</td>
</tr>
<tr>
<td>Level 3</td>
<td>0.1</td>
<td>1.3</td>
<td>−0.6 to 0.9</td>
</tr>
<tr>
<td>Level 4</td>
<td>0.6</td>
<td>1.5</td>
<td>−0.3 to 1.4</td>
</tr>
<tr>
<td>Dilution 1</td>
<td>4.0</td>
<td>0.5</td>
<td>3.5 to 4.5</td>
</tr>
<tr>
<td>Dilution 2</td>
<td>4.5</td>
<td>2.1</td>
<td>2.1 to 6.9</td>
</tr>
<tr>
<td>Human native lithium heparin plasma</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 1</td>
<td>−0.5</td>
<td>1.5</td>
<td>−1.3 to 0.4</td>
</tr>
<tr>
<td>Level 2</td>
<td>−0.5</td>
<td>2.1</td>
<td>−1.7 to 0.7</td>
</tr>
<tr>
<td>Level 3</td>
<td>−0.2</td>
<td>1.4</td>
<td>−1.0 to 0.6</td>
</tr>
<tr>
<td>Level 4</td>
<td>0.2</td>
<td>1.4</td>
<td>−0.6 to −1.0</td>
</tr>
<tr>
<td>Dilution 1</td>
<td>0.8</td>
<td>1.1</td>
<td>−0.4 to 2.0</td>
</tr>
<tr>
<td>Dilution 2</td>
<td>2.0</td>
<td>1.7</td>
<td>0.1 to 3.9</td>
</tr>
<tr>
<td>Human native K$_2$-EDTA plasma</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 1</td>
<td>−3.0</td>
<td>2.5</td>
<td>−4.4 to −1.6</td>
</tr>
<tr>
<td>Level 2</td>
<td>0.7</td>
<td>2.0</td>
<td>−0.4 to 1.9</td>
</tr>
<tr>
<td>Level 3</td>
<td>1.1</td>
<td>1.8</td>
<td>0.1 to 2.1</td>
</tr>
<tr>
<td>Level 4</td>
<td>1.2</td>
<td>1.2</td>
<td>0.5 to 1.9</td>
</tr>
<tr>
<td>Dilution 1</td>
<td>1.9</td>
<td>2.2</td>
<td>−0.6 to 4.3</td>
</tr>
<tr>
<td>Dilution 2</td>
<td>2.4</td>
<td>1.4</td>
<td>0.8 to 4.0</td>
</tr>
<tr>
<td>Human native K$_3$-EDTA plasma</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 1</td>
<td>−0.5</td>
<td>1.3</td>
<td>−1.2 to 0.2</td>
</tr>
<tr>
<td>Level 2</td>
<td>0.2</td>
<td>1.7</td>
<td>−0.7 to 1.2</td>
</tr>
<tr>
<td>Level 3</td>
<td>0.8</td>
<td>1.8</td>
<td>−0.2 to 1.8</td>
</tr>
<tr>
<td>Level 4</td>
<td>−0.2</td>
<td>1.6</td>
<td>−1.1 to 0.7</td>
</tr>
<tr>
<td>Dilution 1</td>
<td>1.4</td>
<td>2.9</td>
<td>−1.8 to 4.7</td>
</tr>
<tr>
<td>Dilution 2</td>
<td>1.0</td>
<td>1.1</td>
<td>−0.2 to 2.3</td>
</tr>
</tbody>
</table>

CI, confidence interval; SD, standard deviation.
Equivalence of results between independent laboratories

The equivalence of RMP results between two independent laboratories was demonstrated in a scatter plot with regression fit. Of the 194 native anonymized residual patient samples analyzed, 23 were excluded, including 21 that were below the LLMI and two samples that were highlighted as outliers using the LORELIA (local reliability) outlier test [37]. Passing–Bablok regression analysis demonstrated very good agreement between the two laboratories, yielding a regression equation excluding outliers with a slope of 1.02 (95% CI 1.00–1.03), an intercept of 2.37 (95% CI 0.29–13.46), and Pearson correlation coefficient of 0.997 (Figure 4).

Bland–Altman analysis also showed good agreement between the two laboratories, with the data scatter being ≤2 shown comparable CVs (n=36) for repeatability ≤3.9% and intermediate precision ≤4.1% as achieved within Laboratory 1.

Of the 154 external quality assessment samples, 21 samples from one provider had to be excluded from bias calculation due to insufficient participants (n=3) in the respective testing scheme. A further six samples from another provider were excluded, due to an inter-lab CV of >20.0% assuming the consensus value may be compromised in this case. These six samples were associated with concentrations below 0.1 μmol/L and a notably strong bias of the external quality assessment scheme (EQAS) consensus value vs. the RMP of 15–50%. The inter-lab CV of these six samples exceeds the 2σ interval and was detected as outliers by the ROUT method using GraphPad Prism 9.3.1 (GraphPad Software, San Diego, USA). Thus, 127 external quality assessment samples from five different proficiency schemes were evaluated for a potential bias vs. the reported RMP and demonstrated fair agreement with the method. Deviation of EQAS consensus values from the RMP values ranged from −15.6 to 19.6%, mean deviation was 1.1%. Mean inter-lab CV of all proficiency testing schemes was 10.7%. However, the scheme with the widest concentration range including samples below 0.3 μmol/L exhibited an inter-lab CV of up to 52.5% (Table 3) and a bias of the consensus value vs. the RMP up to 49.7%.

Uncertainty of results

The total uncertainties for single MTX measurements, represented by the combined uncertainty of calibrator preparation and uncertainty of the precision experiment, were CV ≤4.5% regardless of the concentration level and the type of sample (Table 4). The derived total uncertainty was multiplied by a coverage factor of k=2 to obtain an expanded uncertainty of ≤9.0%, which corresponds to an approximate confidence level of 95%, assuming a normal distribution.

To further reduce measurement uncertainty, the preparation and measurement of calibrators and samples were repeated. Measurement of samples in three replicates on two different days (n=6) saw measurement uncertainty reduced to an expanded uncertainty of ≤3.4% (k=2) (Table 5).

Table 3: Overview of proficiency testing schemes with enrolment of the candidate RMP from August 2020 to May 2022.

<table>
<thead>
<tr>
<th>EQAS provider</th>
<th>Mean number of laboratories</th>
<th>Number of samples analyzed</th>
<th>MTX range, μg/mL</th>
<th>Inter-lab CV, %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>Range</td>
</tr>
<tr>
<td>1. Germany</td>
<td>33</td>
<td>16</td>
<td>0.54–3.1</td>
<td>11.3</td>
</tr>
<tr>
<td>2. Germany</td>
<td>121</td>
<td>14</td>
<td>0.23–1.5</td>
<td>12.0</td>
</tr>
<tr>
<td>3. UK</td>
<td>61</td>
<td>39</td>
<td>0.17–4.1</td>
<td>7.5</td>
</tr>
<tr>
<td>4. USA</td>
<td>60</td>
<td>20</td>
<td>0.16–5.1</td>
<td>9.5</td>
</tr>
<tr>
<td>5. UK</td>
<td>27</td>
<td>44/38*</td>
<td>0.038–36.3</td>
<td>13.6</td>
</tr>
<tr>
<td>6. UK</td>
<td>2</td>
<td>21</td>
<td>0.024–0.88</td>
<td>n/a</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>154</td>
<td>Overall mean</td>
<td>10.7</td>
</tr>
</tbody>
</table>

*Six samples with inter-lab CV ≥20% were excluded from calculation of bias vs. RMP. CV, coefficient of variation; EQAS, external quality assurance services; n/a, not applicable; MTX, methotrexate; RMP, reference measurement procedure.

Discussion

This paper describes a candidate RMP for the quantification of MTX in human serum and plasma, with the method allowing measurement of MTX over a calibration range of 7.200–5,700 ng/mL (0.01584–12.54 μmol/L).
The ID-LC-MS/MS-based method was extensively validated, and the results of performance assessments indicate its potential for evaluating and standardizing routine assays, in addition to assessing patient samples to ensure traceability of individual patient results – a fundamental consideration for an RMP [27].

The method is highly selective and allows determination of MTX to very low concentrations without interference of the main metabolites 7-OH-MTX and DAMPA, other related substances, or the matrix. It is thus superior to other methods that cannot sufficiently separate the metabolites or suffer from cross-reactivity and overestimate MTX concentrations like immunoassays [16, 38, 39]. The described method can therefore also be used after carboxypeptidase intervention, particularly when declining MTX concentrations have to be monitored in the context of very high intervention, particularly when declining MTX concentrations like immunoassays [16, 38, 39]. The described method can therefore also be used after carboxypeptidase intervention, particularly when declining MTX concentrations have to be monitored in the context of very high intervention, particularly when declining MTX concentrations like immunoassays [16, 38, 39]. The described method can therefore also be used after carboxypeptidase intervention, particularly when declining MTX concentrations have to be monitored in the context of very high intervention, particularly when declining MTX concentrations have to be monitored in the context of very high intervention, particularly when declining MTX concentrations have to be monitored in the context of very high intervention, particularly when declining MTX concentrations have to be monitored in the context of very high intervention, particularly when declining MTX concentrations have to be monitored in the context of very high intervention, particularly when declining MTX concentrations have to be monitored in the context of very high intervention, particularly when declining MTX concentrations have to be monitored in the context of very high intervention, particularly when declining MTX concentrations have to be monitored in the context of very high intervention, particularly when declining MTX concentrations have to be monitored in the context of very high intervention, particularly when declining MTX concentrations have to be monitored in the context of very high intervention, particularly when declining MTX concentrations have to be monitored in the context of very high intervention, particularly when declining MTX concentrations have to be monitored in the context of very high intervention, particularly when declining MTX concentrations have to be monitored in the context of very high intervention, particularly when declining MTX concentrations have to be monitored in the context of very high concentration ranges than previously described [22]. Dilution integrity was demonstrated which can further extend this range by 1:100

### Table 4: Total measurement uncertainty for MTX (single measurement).

<table>
<thead>
<tr>
<th>CV, %</th>
<th>Calibrator levels, MTX concentration</th>
<th>Native patient sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>24.00 ng/mL 420.0 ng/mL 3,000 ng/mL 4,200 ng/mL</td>
<td>1 2</td>
</tr>
<tr>
<td>Type B uncertainty</td>
<td>Characterization of reference material</td>
<td>0.25 0.25 0.25 0.25</td>
</tr>
<tr>
<td>Calibrator preparation</td>
<td>Preparation of:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stock solution</td>
<td>0.32 0.32 0.32 0.32</td>
</tr>
<tr>
<td></td>
<td>Working solution</td>
<td>0.60 0.60 n/a n/a</td>
</tr>
<tr>
<td></td>
<td>Spike solution</td>
<td>1.38 1.28 0.59 0.59</td>
</tr>
<tr>
<td></td>
<td>Matrix-based calibrator</td>
<td>1.47 1.38 0.77 0.77</td>
</tr>
<tr>
<td>Type A uncertainty</td>
<td>Intermediate precision</td>
<td>3.6 1.7 1.7 1.6</td>
</tr>
<tr>
<td></td>
<td>Total measurement uncertainty</td>
<td>3.9 2.2 1.9 1.8</td>
</tr>
<tr>
<td>Single measurement</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CV, coefficient of variation; MTX, methotrexate. The total measurement uncertainty of the whole approach for a single measurement estimated as a combination of the uncertainty of calibrator preparation (type B uncertainty) and uncertainty of the precision experiment (type A uncertainty) are given in bold.

### Table 5: Total measurement uncertainty for target value assignment (n=6).

<table>
<thead>
<tr>
<th>CV, %</th>
<th>Nominal concentration</th>
<th>Native patient sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>24.00 ng/mL 420.0 ng/mL 3,000 ng/mL 4,200 ng/mL</td>
<td>1 2</td>
</tr>
<tr>
<td>Type B uncertainty</td>
<td>Characterization of reference material</td>
<td>0.25 0.25 0.25 0.25</td>
</tr>
<tr>
<td>Calibrator preparation</td>
<td>Preparation of:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stock solution</td>
<td>0.32 0.32 0.32 0.32</td>
</tr>
<tr>
<td></td>
<td>Working solution</td>
<td>0.60 0.60 n/a n/a</td>
</tr>
<tr>
<td></td>
<td>Spike solution</td>
<td>1.38 1.28 0.59 0.59</td>
</tr>
<tr>
<td></td>
<td>Matrix-based calibrator</td>
<td>1.47 1.38 0.77 0.77</td>
</tr>
<tr>
<td>Type A uncertainty</td>
<td>Mean of measurement results, CV %</td>
<td>0.5 0.6 0.5 0.4</td>
</tr>
<tr>
<td></td>
<td>Total measurement uncertainty</td>
<td>1.5 1.5 0.9 0.9</td>
</tr>
<tr>
<td>Target value assignment, CV %</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CV, coefficient of variation. The total measurement uncertainty of the whole approach for target value assignment estimated as a combination of the uncertainty of calibrator preparation (type B uncertainty) and uncertainty of the precision experiment (type A uncertainty) are given in bold.
dilution to 570 μg/mL (i.e., 1,254 μmol/L), if necessary. Thus, the method covers the common medical decision points for the management of high-dose MTX therapy [1, 8, 10].

Despite the wide calibration range, the method yielded highly accurate and precise results even at the LLMI, producing less variability than other previously described methods [22]. There are no traceable reference materials or higher-level reference methods available for MTX. In the current study, performance was evaluated relative to existing routine applications. For this purpose, inter-lab CVs provided by proficiency testing schemes were used. A mean inter-lab CV of 10.7% was observed across five different schemes evaluated over a period of almost two years (Table 3). The combined uncertainty of ≤4.5% (k=1) for single measurements of the described method meets the general requirement of CV_{RMP} ≤ 0.5 × CV_{Routine} for an RMP [40]. This would be applicable to TDM. For target value assignment, the protocol stipulates a triplicate measurement on two independent days. This yielded a combined uncertainty of ≤3.4% (k=2) for n=6 and underlines the method’s superior performance (Table 5).

Transferability of the method between two independent laboratories proved the method was capable of measuring native patient samples with suitably low deviation. Comparability with other methods, including mostly automated immunoassays, was demonstrated in proficiency testing schemes. However, with samples ≤0.1 μmol/L immunoassay methods yielded an overestimation of approximately 27% compared with this method and mean inter-lab CV was 10 times the CV of this method at the LLMI (i.e., 25 vs. 2.3%).

The ID-LC-MS/MS-based method described offers several advantages over previously described LC-MS approaches [6, 22]. The method incorporates a quick and simple sample preparation with methanol-based protein precipitation, a short run-time of 10 min, and reliable quantification of MTX in both serum and plasma matrices, while using the same calibrators. These attributes make this ID-LC-MS/MS-based method suitable to be used not only for target value assignment but also for method comparison studies or complaint sample management.

Use of quantitative MS is often restricted by the availability of commercial calibrators, and use of in-house preparations can result in heterogeneous analyte measurements between LC-MS/MS laboratories [13]. A strength of the method described was using qNMR – a method that is increasingly acceptable to NMIs – to determine the absolute content of MTX in the standard and enable unequivocal traceability to SI units [25]. Aside from exhibiting a performance suitable for an RMP, the traceability of the current method is a key distinguishing feature and is pivotal in determining the absolute quantity of a measurand [24, 25].

In summary, the described candidate RMP is suitable for TDM and also for assigning target values to secondary reference materials to aid in assay standardization (e.g., in the in vitro diagnostics industry). This could also help to improve external quality assessment schemes and facilitate comparison and standardization of the large number of different methods developed and applied in clinical laboratories [41]. Standardized, comparable methods might also facilitate the establishment of consistent target ranges for TDM, independent of assay type and treatment scheme [26].

Conclusions

The current paper describes a candidate RMP for MTX, representing to our knowledge the first RMP for this analyte. Traceability of the MTX reference material and the LC-MS/MS platform were assured by incorporating qNMR to assess the absolute content of analyte in the reference material, and extensive method validation. The method is suitable not only for measurement of patients’ specimens but also for assay standardization and target value assignment for various applications. It can thus help to reduce method-derived variability of MTX measurements at different stages and aid in TDM to manage sufficient drug exposure and limit toxicity for better patient outcomes.

Acknowledgments: Third-party medical writing assistance, under the direction of the authors, was provided by Aisling O’Keeffe, PhD, and Liz Southey, MSc, The Salve Health Ltd, UK, and was funded by Roche Diagnostics GmbH (Penzberg, Germany). Technical assistance in the lab during method development and validation was provided by Niklas Müller and Tanja Heise, Labor Berlin – Charité Vivantes Services GmbH (Berlin, Germany).

Research funding: Labor Berlin – Charité Vivantes Services GmbH is a funded cooperation partner of Roche Diagnostics GmbH.

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission. All authors confirmed they have contributed to the intellectual content of this paper and have met the following four 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.
Competing interests: All authors have no conflict of interest to declare. Lena Ruhe, Anett Engel, and Johannes Kolja Hegel are all employees of Labor Berlin – Charité Vivantes Services GmbH, Berlin, Germany. Annika I. Ostermann and Tamara Sumalowitsch were employees of Labor Berlin – Charité Vivantes Services GmbH, Berlin, Germany at the time the study was conducted. Vincent J.T. Schweinsberg is an employee of Labor Berlin – Charité Vivantes Services GmbH, Department of Laboratory Medicine and Toxicology, Sylder Str. 2, 13353 Berlin, Germany. Judith Taibon, Neeraj Singh, Jo Anne Wright, Franziska Liesch, Andrea Geistanger, and Christian Geletneky are all employees of Roche Diagnostics GmbH. Friederike Bauland is an employee of Chrestos Concept GmbH & Co. KG, (Girardetstr. 1–5, 45131 Essen, Germany) and did the work on behalf of Roche Diagnostics GmbH. Roche employees holding Roche non-voting equity securities (Genussscheine): Judith Taibon, Christian Geletneky, Andrea Geistanger.

Informed consent: Not applicable.

Ethical approval: All procedures were in accordance with the Helsinki Declaration. All samples used were exclusively anonymized leftover samples; therefore, ethical approval was not required.

References


Supplementary Material: This article contains supplementary material (https://doi.org/10.1515/cclm-2022-1001).