Scientific Reports

https://doi.org/10.1038/s41598-026-34985-w

Article in Press

Comparison of human metabolome changes identified in a placebo-controlled amphetamine administration study versus those using forensic toxicology routine data

Received: 20 June 2025

Accepted: 1 January 2026

Published online: 06 January 2026

Cite this article as: Bovens A., Leu C., Brockbals L. *et al.* Comparison of human metabolome changes identified in a placebo-controlled amphetamine administration study versus those using forensic toxicology routine data. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-34985-w

Annina Bovens, Claudio Leu, Lana Brockbals, Friederike Holze, Matthias E. Liechti, Thomas Kraemer & Andrea E. Steuer

Comparison of human metabolome changes identified in a placebo-controlled amphetamine administration study versus those using forensic toxicology routine data

Annina Bovens[1], Claudio Leu[1], Lana Brockbals[1], Friederike Holze[2], Matthias E. Liechti[2], Thomas Kraemer[1], *Andrea E. Steuer[1]

[1]Department of Forensic Pharmacology and Toxicology, Zurich Institute of Forensic Medicine, University of Zurich, Zurich, Switzerland

[2]Department of Biomedicine and Department of Clinical Research, Division of Clinical Pharmacology and Toxicology, University Hospital Basel, University of Basel, 4056 Basel, Switzerland

Corresponding Author

Prof. Andrea E. Steuer
University of Zurich
Zurich Institute of Forensic Medicine (ZIFM)
Department of Forensic Pharmacology & Toxicology
Winterthurerstrasse 190/52
CH-8057 Zurich
Switzerland
Tel.: 0041 446355679; fax: 0041 446356852
E-mail address: andrea.steuer@irm.uzh.ch

Abstract

Metabolome studies in forensic toxicology focus on the search for endogenous biomarkers changed by, e.g., drugs of abuse. However, placebo-controlled studies, the ideal study design, in humans are scarce for ethical reasons. Thus, the idea of using routine samples became popular, although confounding factors cannot be controlled.

To systematically evaluate the use of routine samples for metabolomics, a comparison between a placebo-controlled amphetamine study in humans (A, n_{pos}=18, n_{neg}=18) to routine samples either positive or negative for amphetamine, prepared and analyzed over six months (re-evaluated, B, n_{pos}=28, n_{neg}=35) and prepared and analyzed within a single analytical batch (re-extracted, C) was performed. Samples were analyzed using untargeted liquid chromatography-tandem-mass-spectrometry. Comparison was conducted on feature level and based on significance (p- and fold-change-values).

Only 3 features were significant in A, B, and C, and 2 were identified as amphetamine-(fragments). All 31 significant features from A were present in B and C; however, only 11 (36%) and 4 (13%) of them were significant mainly because of higher variation. Still, other significant features were found in routine samples (B/C).

In conclusion, routine samples are generally suitable for detecting differences in the metabolome, even if they do not match those of a controlled study.

Keywords

Amphetamine, Forensic Toxicology, LC-HRMS, Metabolomics, Retrospective

1. Introduction

Metabolome studies focus on the measurable change in (endogenous) metabolites triggered by a particular stimulus. The metabolites of interest are low molecular weight compounds (MW < 1000 Da), like amino acids, lipids, sugars, etc. While metabolomics, in the strict sense, often includes only endogenous compounds, the analytical techniques applied also detect small exogenous molecules (e.g., drugs and their metabolites, ingredients of food, plants, etc.). The field of metabolomics has evolved into a valuable tool in different fields, including precision medicine[1-3], biomarker research[4-9], and drug discovery[10,11]. Lately, its application has also been used in forensic medicine and forensic toxicology (FT)[12-18]. One area of interest, particularly related to FT, is the search for analytical biomarkers for xenobiotics like drugs of abuse (DoA) or new psychoactive substances (NPS) that might improve drug detection, case interpretation, or elucidate underlying pharmacological mechanisms. Placebo-controlled drug administration studies, ideally in a crossover design, represent the gold standard for metabolome investigations but are rare for DoA in humans for ethical reasons. If such studies exist, they are often not designed specifically for forensic purposes but to answer pharmacological or pharmacokinetic questions[14]. However, FT laboratories usually have access to a large cohort of authentic case samples and corresponding data files from the routine work they perform for the authorities. Given the legal regulations and quality management required in forensic environments, samples and analytical results are often controlled regarding sample collection tubes, storage, and laboratory handling[19]. With regular use of liquid chromatography-high resolution mass spectrometry (LC-HRMS) in FT routine work – the technique that is also most often used in (un)targeted metabolome analysis – a re-evaluation of acquired data files for other than the detection and quantification of DoA became feasible. Following these developments, the use of routine forensic data for metabolome studies became of interest over the last few years[19-28]. Thereby, two different strategies can be applied. First, stored samples can be re-extracted and re-analyzed to investigate a metabolome question. This bears the advantage that, similarly to classically designed metabolome experiments, all samples can be measured within one analytical run (batch), and standard metabolome quality control measures (e.g., pool samples, pool dilutions) can be applied[28]. Second, already acquired HRMS data files can be re-processed with metabolome processing workflows[13,19,20,23]. Since this type of data is usually collected for up to several years, their quantities significantly exceed existing controlled DoA studies. However, independent of the chosen strategy, several challenges must be considered when working with routine data. For routine samples, drug doses and time of ingestion remain unknown, and confounding factors like diet, smoking, and drug co-consumption cannot be controlled, leading to (some) variability. Additional inter-batch differences are introduced from preparation and data acquisition when re-processing data files. This is particularly important as untargeted metabolome comparisons are only made based on peak area differences (endogenous metabolites of interest are unknown before analysis, omitting quantification).

Several studies already exist that show the general applicability of routine forensic (toxicological) data. The initial proof-of-principle was done by Nielsen *et al.*[19] who looked for endogenous MDMA markers using routine data re-processing. Mollerup *et al.*[22] then identified new analytical markers amenable to positive electrospray ionization (ESI) for valproate, which can only be measured directly after negative ionization. Lately, Ward *et al.*[24] compiled a large sample cohort from HRMS data files for different

metabolome studies related to postmortem questions, e.g., the determination of the cause of death or the severity of an oxycodone intoxication. Finally, Wang *et al.*[20] could even confirm some new markers for gamma-hydroxybutyric acid (GHB) with their retrospective routine data analysis, which were initially proposed in a placebo-controlled crossover GHB administration study[29].

Even though all these studies re-purposing routine forensic (toxicological) data highlight the potential to use routine data for metabolome experiments, no systematic investigations or validations in comparison to controlled studies in humans are available yet. Thus, we aim to systematically evaluate the challenges and chances of FT routine data by comparing detectable changes in the blood metabolome from a placebo-controlled crossover amphetamine administration study to random routine data positive and negative for amphetamine.

2. Materials and Methods

2.1 Chemicals and Reagents

The chemicals, reagents and blood collection tubes used were already described in detail by Steuer *et al.*[30]. As deuterated internal standards (IS) D3-7-aminoflunitrazepam, D3-benzoylecgonine, D3-clomipramine, D3-cocaine, D3-diphenhydramine, D3-duloxetine, D3-ecgoninemethylester, D3-flunitrazepam, D3-hydromorphone, D3-mirtazapine, D3-morphine, D3-oxymorphone, D3-sertraline, D3-trimipramine, D4-7-aminoclonazepam, D4-a-hydroxy-midazolam, D4-bromazepam, D4-buprenorphine, D4-clozapine, D4-haloperidol, D4-ketamine, D4-midazolam, D4-Ndesalkylflurazepam, D4-risperidone, D4-zopiclone, D5-alprazolam, D5-amisulpride, D5-diazepam, D5-fentanyl, D5-MDA, D5-MDEA, D5-MDMA, D5-nitrazepam, D5-nordazepam, D5-oxazepam, D5-temazepam, D6-amphetamine, D6-chlorpheniramine, D6-chlorprothixene, D6-citaloprame, D6-oxycodone, D6-paroxetine, D6-pregabaline, D6-trazodone, D6-venlafaxine, D6-zolpidem, D8-quetiapine, D9-methadone and D9-methamphetamine were obtained either from AdipoGen (Liestal, Switzerland), Lipomed (Arlesheim, Switzerland), LGC (Wesel, Germany), or Cerilliant (delivered by Sigma-Aldrich, Buchs, Switzerland). The concentration of each IS can be found in the supplementary information **Table S1**.

Methanol (MeOH) and acetonitrile (ACN) were obtained from Fisher Chemical (Switzerland), both in optima LC/MS grade. Water, 0.2 μm filtered, was from VWR (Switzerland), and sodium hydroxide, 98%, was from Honeywell (Germany). Formic acid, ammonium formate, and 2-propanol were from Sigma Aldrich (Switzerland), all in LC/MS grade. All other chemicals used were from Sigma Aldrich and of the highest grade available.

2.2 Study Design for Comparative Analysis Between Study Types

By using an untargeted LC-HRMS workflow (see 2.3), differences in (endogenous) small molecules resulting from amphetamine administration/intake were compared between three different study types as detailed below. The first study type is a placebo-controlled, crossover amphetamine administration study in humans ("controlled-administration", A). For the second study type, routine samples from living individuals submitted to the Zurich Institute of Forensic Medicine (ZIFM) were continuously prepared and analyzed within the first days after submission to the laboratory. Data files from samples either positive or negative for amphetamine within six months were selected for re-processing to find differences in response to amphetamine ("re-processed", B). The third cohort contains the same selected routine samples positive or negative for amphetamine, but here they were re-extracted and measured within a single analytical batch ("re-extracted", C).

2.2.1 Study Type A (controlled-administration)

Plasma samples (lithium heparin stabilized) derived from a placebo-controlled, crossover-designed clinical study conducted by Holze *et al.*[31]. The study was registered at ClinicalTrials.gov (NCT03019822) and was in full accordance with the Declaration of Helsinki, as well as approved by the Ethics Committee Northwest Switzerland (EKNZ). Briefly, 28 participants received d-amphetamine (a single oral dose of 30 mg) or a placebo on different study days at the same time in the morning. The washout periods between amphetamine administration and placebo were at least 10 days. From the 28 participants whose samples were initially collected for pharmacokinetic measurements over 11.5 h after

administration, those collected at 3.5 h after amphetamine or placebo intake from 18 participants were used for the current metabolome experiment (**Table S2)**. The samples were stored at -80°C for five years and thawed once prior to analysis. Sample preparation and analysis were performed as described under 2.3. All measurements were run within one day and one analytical batch.

2.2.2 Study Type B (re-processed authentic routine samples)

Whole blood samples (potassium fluoride stabilized, 0.5%) sent to the ZIFM by police or state attorneys for routine FT analysis, related to driving under the influence or impairment during criminal acts, were continuously analyzed for routine purposes in the first days upon submission to the laboratory. From the saved HRMS datafiles, n=28 amphetamine-positive (amphetamine concentrations 1.0-390 ng/mL, determined by routine LC-MS/MS analysis[30], and n=35 amphetamine-negative samples, matching in age and drug co-administration, were selected for metabolome-re-processing from a six-month period. An overview of the sample cohort is provided in **Table S3**. To account for varying blood collection time points and intervals between event and blood collection (hours between events, e.g., driving under the influence and blood collection), the samples were further divided into different subgroups as follows: collection time (CT) group 1 5 a.m. – 11 a.m., CT2 11 a.m. – 5 p.m., CT3 5 p.m. – 11 p.m., and CT4 11 p.m. – 5 a.m.; time interval (TI) group 1 $\leq$ 3 h, TI2 $>$ 3 h.

Initial sample preparation and analysis were performed as described under 2.3 and ran over several months in multiple analytical batches. Re-processing of the saved data files and further untargeted metabolome analysis of authentic samples after anonymization were in full conformance with Swiss ethical laws, particularly those covering the use of human material in research. A waiver and a declaration of no objection for ethical approval of the Cantonal Ethics Board of the Canton of Zurich were obtained (KEK waiver no. 42.2005 and BASEC-Nr. Req2017-00946).

2.2.3 Study Type C (re-extracted authentic routine samples)

The same whole blood samples as for study type B (re-processed) have been re-extracted for study type C after storage at -20°C for less than one year. The samples (n_{pos}=28, n_{neg}=35, **Table S3**) selected from metabolome re-processing in study type B, were thawed, freshly extracted, and measured as described in 2.3 within one day and one analytical batch.

2.2.4 Random Batch

In addition to the three study types described above, saved data files from 500 random, routine authentic whole blood samples (measured continuously during routine analysis as described under 2.3) were evaluated. These 500 samples did not have a common stimulus to trigger metabolome changes, like, e.g., amphetamine, and were assigned to two random groups for statistical analysis. This random assignment was performed three times, which created three different random groups undergoing the sample data processing workflow as described under 2.4.

2.3 LC-HRMS Analysis

2.3.1 Sample Preparation

For all the above-described study types, the following protein precipitation was performed[30]. Briefly, 200 μL blood or plasma were mixed with 50 μL IS-mix, followed by 50 μL MeOH and 400 μL ACN in an Eppendorf reaction tube. After vortexing, the tubes were shaken on a thermomixer for 10 min at 1'400 rounds per minute (rpm) at room temperature (RTemp), followed by centrifugation for 10 min at 10'000 rpm. 250 μL supernatant were transferred into an LC autosampler vial containing 20 μL 20% formic acid and evaporated to dryness under a gentle stream of nitrogen at RTemp (ca. 1 h), and the residue was reconstituted in a 300 μL eluent mixture (A/B, 95/5, v/v). A pooled sample, once containing all samples from study A (controlled-administration), and once containing all samples from study C (re-extracted), was created by combining 50 μL of each sample from the respective study type. Five pool dilutions (1%, 10%, 20%, 50%, 100%) were prepared through dilution with the eluent mixture.

2.3.2 LC-HRMS Analysis

The LC-HRMS system consisted of a Bruker HPLC system (Elute, Bruker, Switzerland) coupled to an Impact II QTOF (Bruker, Switzerland), controlled by the following software: Compass HyStar (4.1, Bruker, Switzerland), OtofControl (1.0.17, Bruker, Switzerland), and Compass DataAnalysis (2.0, Bruker, Switzerland).

The LC separation was done using a reversed phase (RP) IntensitySolo HPLC 1.8, C18-2 column 100 x 2.1 mm (Bruker, Switzerland) heated at 40°C and applying the following gradient with eluents A (water/MeOH (99:1) with 5mM ammonium formate and 0.01% formic acid) and B (MeOH with 5 mM ammonium formate and 0.01% formic acid): 4% B at 0 min and holding for 10 sec, 18.3% B at 1 min, 50% B at 2.5 min, 99.9% B at 14 min and holding for 2 min; 4% B at 16.10 min and holding until 20 min. The autosampler temperature was set to 5°C with an injection volume of 5 μL. MS measurements were conducted in ESI positive mode, using 500 V for the end plate offset, 2500 V capillary voltage, 2.0 bar nebulizer gas, 8 L/min dry gas, and 200°C dry temperature. The MS was operated in data-independent acquisition mode (DIA) with a mass range from 30-1000 *m/z* in both full scan (MS1) and broadband-activated collision-induced dissociation (bbCID, MS2) modes. The collision energy of the MS1 was set to 6.0 eV with a sample time of 0.5 sec, and a rolling average of 3. The MS2 uses a collision energy of 30.0 eV. 1 mM sodium formate/acetate in 2-propanol (50:50) with 0.2% acid solution was used for the internal mass calibration within each run.

2.4 Data Processing

2.4.1 Peak Picking and -Alignment

The paramounter[32] scripts were used on nine representative raw data files (three from each study type A, B, and C) to estimate the optimal parameters for peak picking and alignment in MS-DIAL[33](version 4.92). After additional manual evaluation, the following parameters were used: MS1 and MS2 tolerance for data collection 0.05 Da; no restriction of retention time and MS1/MS2 mass ranges; minimum peak height for peak detection 1'000 amplitude; mass slice width 0.05 Da; linear weighted moving average smoothing 5 scans; minimum peak width 5 scans; $[M+H]^+$, $[M+NH4]^+$, $[M+Na]^+$ and $[M+H-H_2O]^+$ as possible adducts; alignment reference file pooled sample 20%, alignment retention time tolerance 0.052

min; alignment MS1 tolerance 0.018 Da; gap filling enabled; and blank filtering removing features (defined by a distinct retention time and *m/z* combination) if their fold change (fc) was lower than two compared to the mean blank signal (n=5). The resulting aligned peaks were further handled as features. All three study types were analyzed within the same MS-DIAL run. The random batch was processed in a separate MS-DIAL run.

2.4.2 Data Normalization

The peak area feature table exported from MS-DIAL, resulting from all three study types, was transferred to R[34]. In an initial data cleaning step, features with a signal-to-noise (S/N) ≤ 3 and a peak area ≤ 500 occurring in more than 50% of the measured samples across all three study types were removed. Features showing a linearity $R^2 \leq 0.75$ (linearity of each feature in diluted pool samples, see 2.3.1) were also excluded from further analyses. Probabilistic quotient normalization (PQN)[35] was done in MetaboAnalyst[36] (version 5.0) using the pool sample at 20% as reference.

2.4.3 Data Filtering/Statistical Analysis

In R[34], statistical analysis and data filtering for statistically significant features were performed separately for each study type, A, B, and C. Since all three studies did not follow a normal distribution[37], hypothesis testing of differences between amphetamine-positive and -negative groups was done using a Wilcoxon rank-sum test (Mann-Whitney-U test, $p \leq 0.05$). The fc of each feature was determined via the median between groups. Features with the following combined characteristics were considered statistically significant and potentially interesting for follow-up analysis: $p \leq 0.05$ and $0.5 \leq fc \geq 1.9$.

2.4.4 Identification

Preliminary identification was done for selected features using Bruker's built-in TASQ® software (version 2023b).

2.5 Comparison Between Study Types A, B, and C

For description and comparison of the chosen study cohorts, first, a principle component analysis (PCA) for amphetamine-positive samples was performed in MetaboAnalyst[36] (version 5.0) between different blood CT and TI groups of cohort B/C. In addition, a Wilcoxon rank sum test (Mann-Whitney-U test, $p \leq 0.05$) was applied to check for significant differences between amphetamine concentrations in cohorts A and B/C.

To compare metabolome findings between the three study types, the resulting numbers of features fulfilling the described statistical filtering criteria (2.4.3) were compared between studies A (controlled-administration), B (re-processed), and C (re-extracted). A Venn diagram was used to evaluate the general overlap between studies. Significant features from study A were also manually searched in studies B and C.

2.6 Quality Control

The ISs were used as quality control (QC), calculating relative standard deviations (RSD) of the peak areas before and after PQN and retention time variation. These calculations were done for each study type A, B, and C. An RSD of ≤ 30% was considered acceptable. In addition, pooled QC samples (mix of all samples per study type) were prepared for study A (controlled-administration) and C (re-extracted) as described under sample preparation (2.3.1).

3. Results

3.1 Quality Control

The ISs were chosen as QC, as the study design did not allow for the typically applied metabolomics QC, the preparation of consistent, pooled, authentic samples, for all three study types (limited by continuous analysis of study cohort B over several months). The results for the peak area deviation (RSD, %) for each study type individually and the retention time variation, calculated for study B which covers the whole acquisition period, are summarized in **Table S1**. 12% of the raw IS areas did not fulfill the criteria with an RSD > 30%. However, as expected, after PQN, peak area RSDs were significantly narrowed (**Table S1**), and 93% of the IS passed this quality criteria filter. Regarding retention time, RSDs did not exceed 4% and were stable even over longer analysis times (study B). As expected, higher analytical variability in peak area was observed in study B, measured over multiple analytical batches, compared to the analytical one-batches A (controlled-administration) and C (re-extracted).

3.2 Study Cohorts and Included Samples

An overview of the study cohorts is provided in **Tables S2** and **S3**. Amphetamine concentrations were between 70-130 ng/mL for study A, and 1.0-390 ng/mL for studies B/C, and were significantly different ($p \leq 0.05$) between A and B/C. 10 participants of study A (crossover design) were male and eight female, and all but seven in cohort B/C were male. The age was significantly different ($p<0.05$, Mann-Whitney test) between cohorts A (mean age 27 ± 2 years) and B/C (mean age 36 ± 10 years, amphetamine-positive). The negative cohort in B/C was selected to match the positive groups regarding gender, age, and drug co-administration. In contrast to A, where amphetamine intake was always at 9 a.m. with blood samples collected at 12:30 p.m. (CT2), blood samples for B/C were collected at various CT and TI groups. Only three amphetamine-positive blood samples of cohort B/C were collected within the same time frame as the study A samples. PCA analysis between samples from different CT groups was done for studies B/C and did not reveal any outliers due to blood collection time (**Figure S1**). Most of the authentic samples (B/C) were also collected within 3 h after the event. However, the time of drug intake was not available and could have been much earlier (longer time intervals). PCA analysis between the TI groups showed one sample outlier regarding PC 2. However, as PC 2 only contributes 8.6% to the total variation, it is considered negligible (**Figure S2**).

3.3 Comparison of Significant Changes Induced by Amphetamine Between Study Types

In total, 10'082 features resulted after peak picking and data cleaning (S/N ≥ 3, raw peak area ≥ 500, linearity $R^2 \geq 0.75$) for studies A, B, and C. From these, 31, 130, 75 features were considered as

significantly different ($p \leq 0.05$ and $0.5 \leq fc \geq 1.9$) between amphetamine positive and amphetamine-negative groups per study type A (controlled-administration), B (re-processed), and C (re-extracted), respectively. The results are visualized in the Venn diagram in **Figure 1**. For study A, 15 features increased and 16 decreased, respectively, while for studies B and C, 71 and 14 features revealed feature increases, and 59 and 61 feature decreases. Study A is considered the gold standard since it is a placebo-controlled crossover study with the best possible control of confounding factors. Therefore, the observed changes are more confidently induced by amphetamine, and the 31 significant features found in study A were used to compare the three study types. Only 3 features (4.11_91.0242, 4.10_136.1114, and 10.62_494.3137) were found that consistently result in significant differences between amphetamine-positive and amphetamine-negative blood samples (**Figure 2a, b, c**). Despite these features being statistically significant in all studies, the boxplots visualize differences, mainly in peak area intensities and their variability. The calculated corresponding fc- and p-values are summarized in **Table S4**. Feature 4.10_136.1114 was identified as amphetamine, based on the Bruker Impact II TASQ® database criteria (retention time 4.1 ± 0.5 min, area threshold ≥ 3000, height threshold ≥ 750, ppm ± 7.35, mDa ± 1.0). Feature 4.11_91.0242 represents the main amphetamine fragment (retention time 4.1 ± 0.5 min, area threshold ≥ 1500, height threshold ≥ 400, ppm ± 10.98, mDa ± 1.0). Feature 10.62_494.3137 could not be identified yet and is, in contrast to the two amphetamine-derived features, decreased between the amphetamine-positive and negative-groups. The low overlap in significant metabolome changes among the study types warrants more detailed investigation, as provided in the following chapters.

3.3.1 Comparison Between Studies A (controlled-administration) and C (re-extracted routine samples)

Since study A (controlled-administration) and C (re-extracted) were each analyzed within one analytical batch, they are expected to be analytically similar and should show less analytical (inter-batch) variability. All 31 significant ($p \leq 0.05$ and $0.5 \leq fc \geq 1.9$) study A features were present in study C, but only 4 (13%) – including the 3 generally overlapping features - were significantly different between the amphetamine-positive and -negative groups in study C. A boxplot of feature 14.98_401.3914, as the only feature additionally matching between studies A and C, is depicted in **Figure 2d**.

The 27 features that are significantly different between the amphetamine and placebo group in study A, but do not show a significant difference between amphetamine-positive and amphetamine-negative samples in study C, are exemplified in **Figure 3a, b** for two characteristic features. In general, features in study A showed less variability between samples than study C and were less prone to outliers. The median differences observed for these (endogenous) features in study C usually showed only marginal changes between amphetamine-positive and amphetamine-negative samples, meaning fc-values close to 1 (± 0.3) (**Table S4**). This applied to 18 features (67%).

Looking at the features that significantly distinguish amphetamine-positive from amphetamine-negative samples in study C, 71 of these did not show a significant p-value between the amphetamine administration and placebo group in study A, as exemplified for feature 9.30_241.2163 in **Figure 3c**. These were often features with higher peak areas in study C, compared to A (54 features, 76%), such

as, e.g., feature 9.30_241.2163. Other examples (65 features, 92%), such as 13.12_498.2279, showed similar trends in both groups, but did not reach significant fc-changes in A (**Figure 3d**).

3.3.2 Comparison Between Studies A (controlled-administration) and B (re-processed routine samples)

In addition to the uncontrolled conditions of routine samples in general, study B introduced additional analytical variations with measurement over multiple batches over a long time interval. Interestingly, despite an expected higher variation, more overlapping features between studies A and B were detected than between studies A and C.

All 31 significant study A features were present in study B, but only 11 (36%) showed significant differences ($p \leq 0.05$ and $0.5 \leq fc \geq 1.9$) between amphetamine-positive and amphetamine-negative samples in study B. These included the 3 features (4.10_136.1114, 4.11_91.0242, and 10.62_494.3137), significantly changed in all study types (see **Figure 2a, b, c**). In **Figure 4,** boxplots of 2 other features are depicted as examples. Feature 2.67_295.0679 thereby resulted in much higher peak areas in amphetamine-positive samples of study B. Feature 10.57_472.3042 showed a similar trend in A and B/C, but only reached a significant fc-value in studies A and B. The same was true for its linked features 10.57_414.3009 (fragment minus *m/z* 58), 10.57_494.2858 (Na-adduct), and 10.57_510.2595 (K-adduct). All features, their fc- and p-values are given in **Table S4**.

The 2 features in **Figure 3a and b**, which stand exemplarily for the 19 significantly different features in A but not C, also did not change significantly in B. Similarly, features in study B had higher variability than study A, were more prone to outliers, and showed only minimal median differences in peak area per condition.

Study B provided the highest number of significant features, but with no overlap with study A. **Figure 3e, f** shows two of these 119 features that are significant in B but not in A. Even though large variation and outliers are observed for features in B, they are still statistically significant, in contrast to the non-significant ones observed in studies A (e.g., feature 7.57_549.1089, representative for 102 features, 86 %). Also, higher peak areas compared to A (49 features, 41%), such as for feature 1.24_420.7904 were observed.

3.3.3 Comparison Between Studies B (re-processed routine samples) and C (re-extracted routine samples)

Studies B and C used the same samples, once after continuous analysis over a time period of several months (B), but immediately upon arrival at ZIFM, and once re-extracted (longer storage period), but analyzed within one analytical batch (C). Thus, a high overlap between significant features would have been expected. Surprisingly, even though all 75 significant study C features were present in B, only 16 (21%) of them revealed significant differences ($p \leq 0.05$ and $0.5 \leq fc \geq 1.9$) between amphetamine-positive and amphetamine-negative samples in study B. Four representative examples of the 16 overlapping features are shown in **Figure 5.**

No consistent trend was observed regarding peak area variation. Some features (e.g., 3.85_211.0571, 10 features in total, 63%) showed higher variation in study B, while others had higher variation in C (e.g., 14.96_175.1481, 6 features in total, 38%). Surprisingly large peak area differences were observed

between B and C despite PQN, e.g., exemplified in features 12.92_88.2017, while other features were overall comparable in terms of fc- and p-values e.g. 15.21_764.5564.

3.4 Random Batch

To exclude random significant findings, the data processing workflow was applied to 500 (other) authentic routine samples, which were randomly divided three times into two groups for statistical comparison. In total, 34'104 features resulted after peak picking and data cleaning. The random assignment of these authentic routine samples into two groups, without a common stimulus, such as amphetamine consumption, showed a mean of only 2 significant ($p \leq 0.05$ and $0.5 \leq fc \geq 1.9$) features (range 1-3).

4. Discussion

FT routine samples for research purposes, including metabolomics, have become increasingly popular during the last few years[19-28]. Here, we systematically compare, for the first time, results obtained from DoA metabolomics using human FT routine samples with those from a placebo-controlled administration study.

Amphetamine was chosen as the example compound because plasma samples from a placebo-controlled administration study were available. It does not undergo extensive drug metabolism, which would result in numerous increased features of non-endogenous nature, and metabolome changes induced by amphetamine and amphetamine-like drugs have already been described[38].

4.1 Data Acquisition and Evaluation

Data acquisition in our study was done using DIA, employing the drug screening method routinely used at the ZIFM. While DIA offers several advantages, such as seamless acquisition of MS/MS data, data analysis of large DIA batches also poses challenges. The large amount of generated data requires a suitable IT infrastructure for acquisition, storage, and processing within a reasonable time frame. As untargeted metabolomics comparisons are based solely on peak areas, which can considerably fluctuate from day to day, depending on the daily instrument performance, a robust data normalization strategy was needed. PQN was chosen due to the highly varied character of endogenous compounds. Normalization by a single IS would not adequately capture this variation. PQN considers each sample and each feature individually, resulting in a more robust normalization as indicated by the reduced RSDs in our QC data set, even for data acquired over several months (**Table S1**). Still, the influence of analytical batches on area shifts is well described[39] and visible in our data with analytical variation generally lower for samples analyzed in a single batch (studies A and C) than for samples measured across multiple batches (study B).

Sample acquisition for study B (re-processed) took place over several months and analytical batches. Thus, using pooled samples and pool dilutions as typical QC measures in untargeted metabolomics was not feasible for all study types evaluated. It has to be considered that both the PQN reference and the linearity filter were based solely on features being present in study A (controlled-administration). Nevertheless, as study A was considered the most reliable study type and the entire comparison was

based on study A features, this appeared as the best choice, although potentially interesting features in studies B/C, missing the linearity filter of A features, may have been excluded.

The final comparison was based on features rather than identified metabolites. Feature annotation remains the bottleneck in untargeted metabolomics investigations[14] and would have massively reduced the number of comparable features. Therefore, further feature identification was omitted. The combined peak picking and alignment procedure of all study types has allowed a smooth comparison at the feature level.

Untargeted metabolomics data processing can be done not only by univariate feature-based hypothesis testing but also using multivariate approaches and/or machine learning. However, these more sophisticated statistics require specified experience in bioinformatics, which is often lacking in routine FT laboratories. Also, these more complex and global approaches ultimately lead to a selection of relevant features discriminating amphetamine-positive from amphetamine-negative samples that need closer (manual) evaluation.

4.2 Comparison of (Significant) Features Among Study Types

Only 3 features (4.10_136.1114, 4.11_91.0242, and 10.62_494.3137) revealed significant changes present in all three study types A (controlled-administration), B (re-processed), and C (re-extracted). Two of these were identified as amphetamine and its main fragment ion, respectively. The presence of amphetamine serves as proof of concept for the presented workflow, at least if feature changes are high enough, as was expected for amphetamine. The fact that amphetamine was apparently integrated by MS-DIAL even in placebo/amphetamine-negative samples (**Figure 2a, b**) could be explained by known interferences occurring with similar *m/z values* as amphetamine and its rather uncharacteristic fragment ion *m/z* 91. As the chosen DIA method does not link the precursor to the respective fragment ions, each mass alone remains less specific. Except for the expected amphetamine, the lack of overlap between study types remained disappointing and needed further investigation. Based on the different comparison results between each study type, three main observations were made for features being significant in only one study type. First, differences in peak areas were observed, with often much higher peak areas in routine samples, which can best be explained by the higher doses in typical abuse samples in forensics compared to the controlled study. Second, smaller median differences were present mostly for routine samples, indicated mainly by lower fc-values and also non-significant p-values. Third, higher variations were observed, again mainly but not exclusively in routine samples.

Several aspects may be responsible for these effects, besides the ingested dose, the blood sampling time point, the time difference between drug intake and blood collection, the analyzed matrix, sample stability, or simply coincidence. While the following points are likely to have a very significant impact on the observed metabolome changes, they also represent the reality of routine samples that cannot be circumvented (except for the matrix used). The following discussion of these points also illustrates why highly controlled studies with as few confounding factors as possible represent the gold standard in metabolome research.

Dose/stereoselectivity: The individual doses of amphetamine were controlled in study A, but were unknown for studies B/C. Higher and/or multiple doses might have been consumed recreationally, leading to significantly higher amphetamine concentrations in studies B/C (**Tables S2, S3**). It can

therefore be assumed that routine samples are likely to show more intense or different metabolome changes compared to controlled single-dose administration. Furthermore, only the pharmacologically active d-amphetamine was administered in study A, while, recreationally, most likely racemic amphetamine (mixture of d- and l-amphetamine) was consumed. L-amphetamine may exhibit different effects on the metabolome; however, to the best of our knowledge, no studies are available regarding the impact of stereoselectivity on metabolome changes. This may explain higher peak areas in routine samples, but not vice versa or between B and C. However, given the number of (other) confounding factors, it is not possible to conclusively prove that these changes are related to amphetamine.

Sampling time point (circadian variation): Certain metabolites fluctuate between daytime (circadian variation). Samples for study A were all collected at the same time (CT2). Thus, no major endogenous differences regarding circadian metabolites are to be expected. For studies B/C, blood collection varied among the samples and was different from study A. Thus, circadian changes could partly explain the varying peak areas and higher variation among routine samples.

Time difference between intake and blood collection (metabolism): The time between intake and blood collection was constant for study A (TI2, 3.5 h, in the range of amphetamine's C_{max}), but uncontrolled in routine samples. Even if the time difference between the event (e.g., driving under the influence) and blood sampling is taken into account, this does not rule out significantly earlier drug use or even use after the event. Still, the shorter the time between drug intake and blood sampling, the less likely it is that significant changes in the metabolome will be detectable, as systemic responses may not yet have developed.". This could in part explain findings in routine samples which are not present in study A.

Matrix: Study A uses plasma, while studies B/C rely on whole blood. While it is known that metabolite concentrations are comparable between human serum and plasma with a slight advantage for plasma[40,41], no comparison between human plasma and whole blood in the context of metabolomics exists. However, as plasma is derived from whole blood, whole blood contains everything plasma does, but not necessarily in the same concentration and vice versa. Thus, since whole blood from studies B/C is compared to plasma from study A, every feature found in plasma was also present in whole blood. Differences in concentrations might explain part of the observed higher peak areas in routine samples compared to study A.

Freeze-thaw cycles and stability: Samples from studies B/C may have undergone one more freeze-thaw cycle compared to those in study A, which could impact sample integrity and is known to lead to false-positive results[42,43]. Some substances may have degraded between measurements of study B (within a few days after arrival at ZIFM) and study C (within one year), which would explain the higher peak areas in study B compared to study C.

Sample size: n=36 samples (18 per condition) were analyzed from a paired clinical placebo-controlled administration study, and n=63 (28 amphetamine-positive, 35 amphetamine-negative) routine FT samples. Meaningful calculation of statistical power in metabolomics is complicated, as actual effect sizes (corresponding to the difference between two sample groups relative to their within-group variance) cannot be reliably estimated for all metabolites. While the sample size of our routine samples is sufficient to reach an acceptable power of 0.8 for large effect sizes (d=0.8; calculated power 0.9, G*Power[44] version 3.1.9.7), it is not high enough to detect changes with medium effect sizes (e.g., d=0.5,

power 0.5). Less variation, or matching samples, as available in study A, can increase the effect size and increase statistical power. Detection of DoA-induced metabolomic changes in routine samples will likely benefit from large sample cohorts (e.g., a sample size of 60 per group would be necessary for a power of 0.8 to detect medium effect sizes)[45,46]. However, depending on the analytical question, the inclusion of additional samples is time-consuming or not possible.

Coincidence: Finally, the significant changes observed in individual study types could simply be coincidence. However, the three-fold data evaluation of samples randomly assigned to two groups proved this assumption to be highly unlikely.

5. Conclusion

This study aimed to systematically evaluate the usability of (retrospective) routine data for DoA metabolomics in humans. The comparison between a controlled-administration study (A), re-processed, authentic routine samples measured as multi-batch (B), and re-extracted authentic routine samples measured as a one-batch (C) revealed several key findings. Consistent results across all three study designs were mainly found for amphetamine itself, i.e., for substances for which there were logically large changes (exogenous substance administration) between the tested conditions. Placebo-controlled studies are still the gold standard, as confounding factors, which can have a massive influence on the metabolome, are controlled in the best possible way. Despite all possible confounders, the analysis of routine samples also led to the identification of distinguishing features. However, definitive proof that these are attributable to amphetamine remains open. At least the statistical evaluation of random sample groups speaks against random findings in the routine samples. On the contrary, routine samples may also offer advantages over controlled studies, as, for example, higher and/or multiple doses are consumed and thus may be much closer to routine FT. Both study types, B (re-evaluated) and C (re-extracted), have advantages and disadvantages. While B is mainly limited by the fact that inter-batch differences have to be normalized and higher variation is to be expected, direct measurement of the samples is better than repeated processing in terms of stability and is more resource-efficient.

In summary, our study shows that routine samples are generally suitable for detecting differences in the metabolome that do not appear to be random, even if they do not correspond to those of a controlled study. In general, larger differences between the groups are required to be detected with routine samples. In addition, the largest possible cohorts should be used (if available). This also allows specific inclusion and exclusion criteria to be applied in order to form comparison groups that are as homogeneous as possible and to ensure sufficient statistical power.

Data Availability

The datasets generated and/or analyzed during the current study are not publicly available due to ethical constriction regarding the private information present in routine data. Data can only be made available via the corresponding author upon reasonable request.

References

1 Xiao, Y. *et al.* Comprehensive metabolomics expands precision medicine for triple-negative breast cancer. *Cell Res* **32**, 477-490 (2022). https://doi.org:10.1038/s41422-022-00614-0

2 He, X., Liu, X., Zuo, F., Shi, H. & Jing, J. Artificial intelligence-based multi-omics analysis fuels cancer precision medicine. *Semin Cancer Biol* **88**, 187-200 (2023). https://doi.org:10.1016/j.semcancer.2022.12.009

3 Barberis, E. *et al.* Precision Medicine Approaches with Metabolomics and Artificial Intelligence. *Int J Mol Sci* **23** (2022). https://doi.org:10.3390/ijms231911269

4 Zhang, A., Sun, H., Yan, G., Wang, P. & Wang, X. Mass spectrometry-based metabolomics: applications to biomarker and metabolic pathway research. *Biomed Chromatogr* **30**, 7-12 (2016). https://doi.org:10.1002/bmc.3453

5 Klein, M. S. & Shearer, J. Metabolomics and Type 2 Diabetes: Translating Basic Research into Clinical Application. *Journal of Diabetes Research* **2016**, 3898502 (2016). https://doi.org:https://doi.org/10.1155/2016/3898502

6 Wang, X., Chen, S. & Jia, W. Metabolomics in Cancer Biomarker Research. *Current Pharmacology Reports* **2**, 293-298 (2016). https://doi.org:10.1007/s40495-016-0074-x

7 Ambati, C. S., Yuan, F., Abu-Elheiga, L. A., Zhang, Y. & Shetty, V. Identification and Quantitation of Malonic Acid Biomarkers of In-Born Error Metabolism by Targeted Metabolomics. *J Am Soc Mass Spectrom* **28**, 929-938 (2017). https://doi.org:10.1007/s13361-017-1631-1

8 Ren, S. *et al.* Integration of Metabolomics and Transcriptomics Reveals Major Metabolic Pathways and Potential Biomarker Involved in Prostate Cancer. *Mol Cell Proteomics* **15**, 154-163 (2016). https://doi.org:10.1074/mcp.M115.052381

9 Wurtz, P. *et al.* Metabolic profiling of alcohol consumption in 9778 young adults. *Int J Epidemiol* **45**, 1493-1506 (2016). https://doi.org:10.1093/ije/dyw175

10 Lu, Y. & Chen, C. Metabolomics: Bridging Chemistry and Biology in Drug Discovery and Development. *Current Pharmacology Reports* **3**, 16-25 (2017). https://doi.org:10.1007/s40495-017-0083-4

11 Mercier, K. A., Al-Jazrawe, M., Poon, R., Acuff, Z. & Alman, B. A Metabolomics Pilot Study on Desmoid Tumors and Novel Drug Candidates. *Sci Rep* **8**, 584 (2018). https://doi.org:10.1038/s41598-017-18921-7

12 Castillo-Peinado, L. S. & Luque de Castro, M. D. Present and foreseeable future of metabolomics in forensic analysis. *Anal Chim Acta* **925**, 1-15 (2016). https://doi.org:10.1016/j.aca.2016.04.040

13 Steuer, A. E., Brockbals, L. & Kraemer, T. Metabolomic Strategies in Biomarker Research-New Approach for Indirect Identification of Drug Consumption and Sample Manipulation in Clinical and Forensic Toxicology? *Front Chem* **7**, 319 (2019). https://doi.org:10.3389/fchem.2019.00319

14 Steuer, A. E., Brockbals, L. & Kraemer, T. Untargeted metabolomics approaches to improve casework in clinical and forensic toxicology-"Where are we standing and where are we heading?". *Wires Forensic Sci* **4** (2021). https://doi.org:ARTN e1449
10.1002/wfs2.1449

15 Manier, S. K. & Meyer, M. R. Current Situation of the Metabolomics Techniques Used for the Metabolism Studies of New Psychoactive Substances. *Ther Drug Monit* **42**, 93-97 (2020). https://doi.org:10.1097/FTD.0000000000000694

16 Szeremeta, M., Pietrowska, K., Niemcunowicz-Janica, A., Kretowski, A. & Ciborowski, M. Applications of Metabolomics in Forensic Toxicology and Forensic Medicine. *Int J Mol Sci* **22** (2021). https://doi.org:10.3390/ijms22063010

17 Dinis-Oliveira, R. J. Metabolomics of drugs of abuse: a more realistic view of the toxicological complexity. *Bioanalysis* **6**, 3155-3159 (2014). https://doi.org:10.4155/bio.14.260

18 Zaitsu, K., Hayashi, Y., Kusano, M., Tsuchihashi, H. & Ishii, A. Application of metabolomics to toxicology of drugs of abuse: A mini review of metabolomics approach to acute and chronic toxicity studies. *Drug Metab Pharmacokinet* **31**, 21-26 (2016). https://doi.org:10.1016/j.dmpk.2015.10.002

19 Nielsen, K. L., Telving, R., Andreasen, M. F., Hasselstrom, J. B. & Johannsen, M. A Metabolomics Study of Retrospective Forensic Data from Whole Blood Samples of Humans Exposed to 3,4-Methylenedioxymethamphetamine: A New Approach for Identifying Drug Metabolites and Changes in Metabolism Related to Drug Consumption. *J Proteome Res* **15**, 619-627 (2016). https://doi.org:10.1021/acs.jproteome.5b01023

20 Wang, T. *et al.* A Retrospective Metabolomics Analysis of Gamma-Hydroxybutyrate in Humans: New Potential Markers and Changes in Metabolism Related to GHB Consumption. *Front Pharmacol* **13**, 816376 (2022). https://doi.org:10.3389/fphar.2022.816376

21 Pasin, D. *et al.* Metabolomics-driven determination of targets for salicylic acid and ibuprofen in positive electrospray ionization using LC-HRMS. *Drug Test Anal* **14**, 747-756 (2022). https://doi.org:10.1002/dta.3215

22 Mollerup, C. B. *et al.* Retrospective analysis for valproate screening targets with liquid chromatography-high resolution mass spectrometry with positive electrospray ionization: An omics-based approach. *Drug Test Anal* **11**, 730-738 (2019). https://doi.org:10.1002/dta.2543

23 Hoj, L. J. *et al.* Identification of phenobarbital and other barbiturates in forensic drug screening using positive electrospray ionization liquid chromatography-high resolution mass spectrometry. *Drug Test Anal* **11**, 1258-1263 (2019). https://doi.org:10.1002/dta.2603

24 Ward, L. J. *et al.* Postmortem metabolomics as a high-throughput cause-of-death screening tool for human death investigations. *iScience* **27**, 109794 (2024). https://doi.org:10.1016/j.isci.2024.109794

25 Kronstrand, R. *et al.* The metabolism of the synthetic cannabinoids ADB-BUTINACA and ADB-4en-PINACA and their detection in forensic toxicology casework and infused papers seized in prisons. *Drug Test Anal* **14**, 634-652 (2022). https://doi.org:10.1002/dta.3203

26 Roman, M., Strom, L., Tell, H. & Josefsson, M. Liquid chromatography/time-of-flight mass spectrometry analysis of postmortem blood samples for targeted toxicological screening. *Anal Bioanal Chem* **405**, 4107-4125 (2013). https://doi.org:10.1007/s00216-013-6798-0

27 Brockbals, L. *et al.* Time- and Site-Dependent Postmortem Redistribution of Antidepressants and Neuroleptics in Blood and Alternative Matrices. *Journal of Analytical Toxicology* **45**, 356-367 (2020). https://doi.org:10.1093/jat/bkaa092

28 Brockbals, L. *et al.* Postmortem Metabolomics: Strategies to Assess Time-Dependent Postmortem Changes of Diazepam, Nordiazepam, Morphine, Codeine, Mirtazapine and Citalopram. *Metabolites* **11**, 643 (2021).

29 Steuer, A. E. *et al.* Identification of new urinary gamma-hydroxybutyric acid markers applying untargeted metabolomics analysis following placebo-controlled administration to humans. *Drug Testing and Analysis* **11**, 813-823 (2019). https://doi.org:https://doi.org/10.1002/dta.2558

30 Steuer, A. E., Keller, M., Kraemer, T. & Poetzsch, S. N. Multianalyte Approach-Including Automated Preparation of Calibrators-for Validated Quantification of 82 Drugs in Whole Blood by Liquid Chromatography-Tandem Mass Spectrometry. *Drug Test Anal* (2024). https://doi.org:10.1002/dta.3794

31 Holze, F. *et al.* Distinct acute effects of LSD, MDMA, and D-amphetamine in healthy subjects. *Neuropsychopharmacology* **45**, 462-471 (2020). https://doi.org:10.1038/s41386-019-0569-3

32 Guo, J., Shen, S. & Huan, T. Paramounter: Direct Measurement of Universal Parameters To Process Metabolomics Data in a "White Box". *Anal Chem* **94**, 4260-4268 (2022). https://doi.org:10.1021/acs.analchem.1c04758

33 Tsugawa, H. *et al.* MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat Methods* **12**, 523-526 (2015). https://doi.org:10.1038/nmeth.3393

34 R: A language and environment for statistical computing v. 4.4.1 (R Foundation for Statistical Computing, Vienna, Austria, 2024).

35 Dieterle, F., Ross, A., Schlotterbeck, G. & Senn, H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. *Anal Chem* **78**, 4281-4290 (2006). https://doi.org:10.1021/ac051632c

36 Pang, Z. *et al.* Using MetaboAnalyst 5.0 for LC-HRMS spectra processing, multi-omics integration and covariate adjustment of global metabolomics data. *Nat Protoc* **17**, 1735-1761 (2022). https://doi.org:10.1038/s41596-022-00710-w

37 S.S. Shapiro, M. B. W. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika* **52**, 591-611 (1965).

38 Steuer, A. E. *et al.* Comparative Untargeted Metabolomics Analysis of the Psychostimulants 3,4-Methylenedioxy-Methamphetamine (MDMA), Amphetamine, and the Novel Psychoactive Substance Mephedrone after Controlled Drug Administration to Humans. *Metabolites* **10** (2020). https://doi.org:10.3390/metabo10080306

39 Fernandez-Albert, F. *et al.* Intensity drift removal in LC/MS metabolomics by common variance compensation. *Bioinformatics* **30**, 2899-2905 (2014). https://doi.org:10.1093/bioinformatics/btu423

40 Yu, Z. *et al.* Differences between human plasma and serum metabolite profiles. *PLoS One* **6**, e21230 (2011). https://doi.org:10.1371/journal.pone.0021230

41 Handley, S. A., Silk, S. W., Fisher, D. S., Subramaniam, K. & Flanagan, R. J. Clozapine and Norclozapine Concentrations in Paired Human Plasma and Serum Samples. *Therapeutic Drug Monitoring* **40**, 148-150 (2018). https://doi.org:10.1097/ftd.0000000000000478

42 Hirayama, A. *et al.* Effects of processing and storage conditions on charged metabolomic profiles in blood. *Electrophoresis* **36**, 2148-2155 (2015). https://doi.org:10.1002/elps.201400600
43 Chen, D., Han, W., Huan, T., Li, L. & Li, L. Effects of Freeze-Thaw Cycles of Blood Samples on High-Coverage Quantitative Metabolomics. *Anal Chem* **92**, 9265-9272 (2020). https://doi.org:10.1021/acs.analchem.0c01610
44 Faul, F., Erdfelder, E., Lang, A. G. & Buchner, A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* **39**, 175-191 (2007). https://doi.org:10.3758/bf03193146
45 Stanislava Rakusanova, T. C. Tips and tricks for LC–MS-based metabolomics and lipidomics analysis. *TrAC Trends in Analytical Chemistry* **180** (2024). https://doi.org:https://doi.org/10.1016/j.trac.2024.117940.
46 Blaise, B. J. *et al.* Power Analysis and Sample Size Determination in Metabolic Phenotyping. *Anal Chem* **88**, 5179-5188 (2016). https://doi.org:10.1021/acs.analchem.6b00188

Acknowledgement

The authors would like to thank Maja Keller for her support and express their gratitude to Emma Louise Kessler, MD for her generous legacy she donated to the Institute of Forensic Medicine at the University of Zurich, Switzerland for research purposes.

Author Contributions:

Annina Bovens conducted the experiments, the statistical analysis and wrote the manuscript. Claudio Leu conducted the experiments. Lana Brockbals wrote the manuscript. Friederike Holze and Matthias E. Liechti provided the samples from the controlled administration study. Thomas Kraemer wrote the manuscript. Andrea E. Steuer had the organizational lead and wrote the manuscript.

Additional Information / Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Consent Statement

Informed consent was obtained for all data used from the placebo-controlled study (study A) performed by Holze et al. The study was registered at ClinicalTrials.gov (NCT03019822) and was in full accordance with the Declaration of Helsinki, as well as approved by the Ethics Committee Northwest Switzerland (EKNZ).

Due to the retrospective nature of the routine data used for studies B and C a waiver and a declaration of no objection for ethical approval of the Cantonal Ethics Board of the Canton of Zurich were obtained (KEK waiver no. 42.2005 and BASEC-Nr. Req2017-00946).

Figure Legends

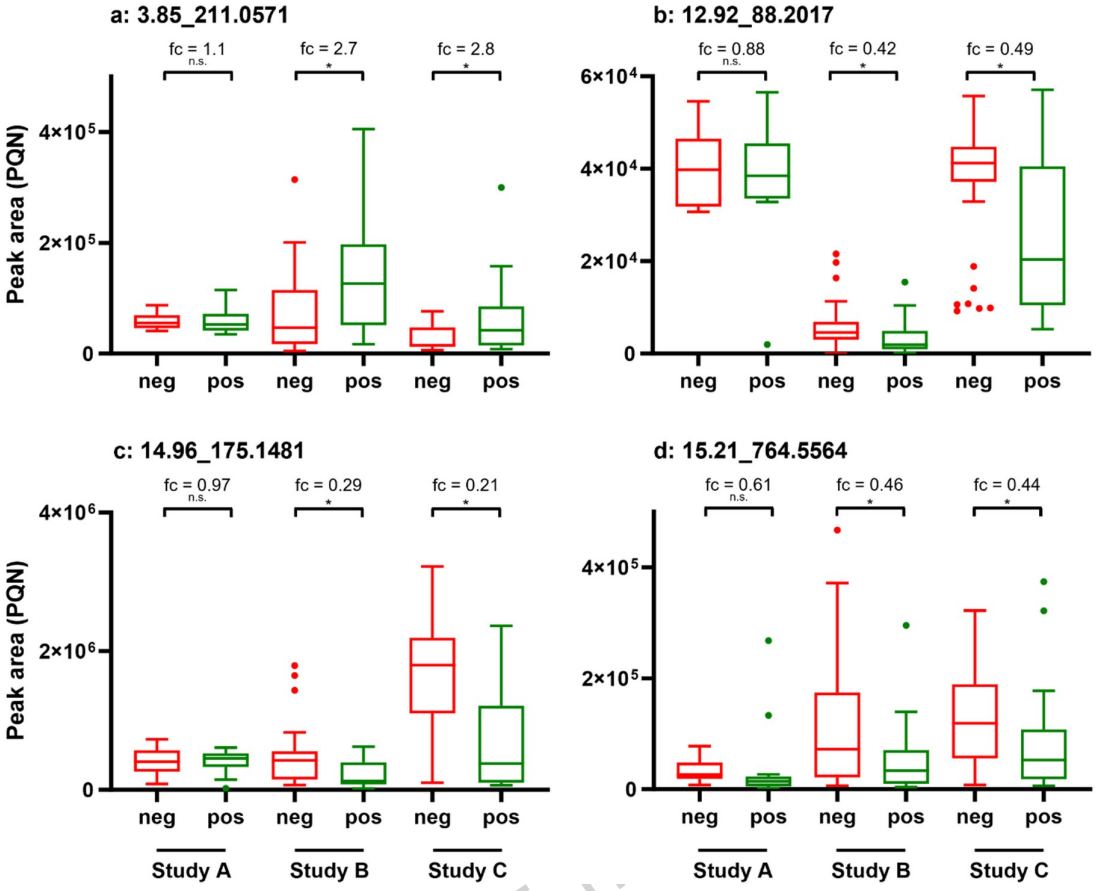
Figure 1: Venn diagram of features which are significantly different between the amphetamine-positive and amphetamine-negative groups in studies A (controlled-administration, green), B (re-processed, lilac), and C (re-extracted, orange).

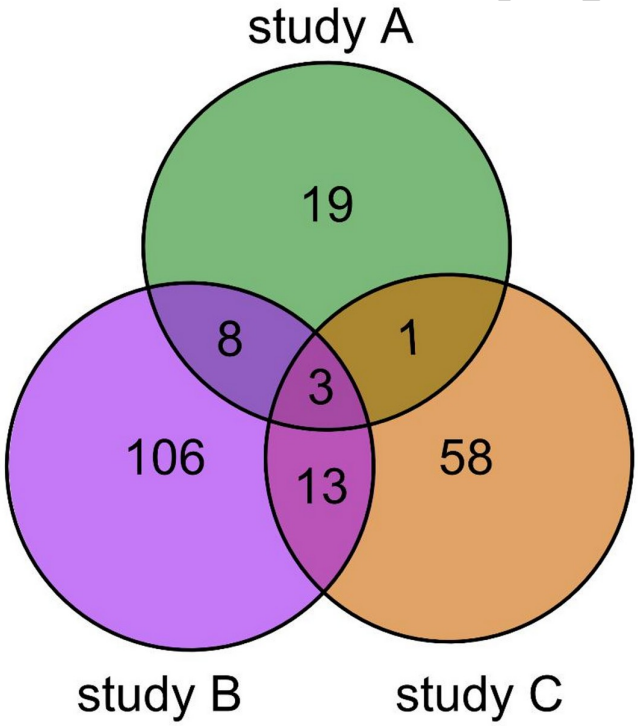
Figure 2: Boxplots of the 3 features (a, b, and c) significantly changed in studies A, B, and C. Boxplot d shows the feature that is only significant in A and C (but not in B). The box represents the median and 25%/75% percentiles, the whiskers indicate the 5-95% percentiles. The y-axis shows the PQN peak area of each feature. Amphetamine-positive groups are given in green and amphetamine-negative ones in red for each study type. Features with a p-value ≤ 0.05 (Wilcoxon rank-sum test) are indicated with the *-symbol. Feature increases or decreases between amphetamine-positive and amphetamine-negative groups are shown as fc-values. Features were considered significant with a p-value ≤ 0.05 and a fc-value ≥ 1.9 or ≤ 0.5. Features 4.11_91.0242 and 4.1_136.1114 were identified as amphetamine (*m/z* 136.1114) and its main fragment ion (*m/z* 91.0242).

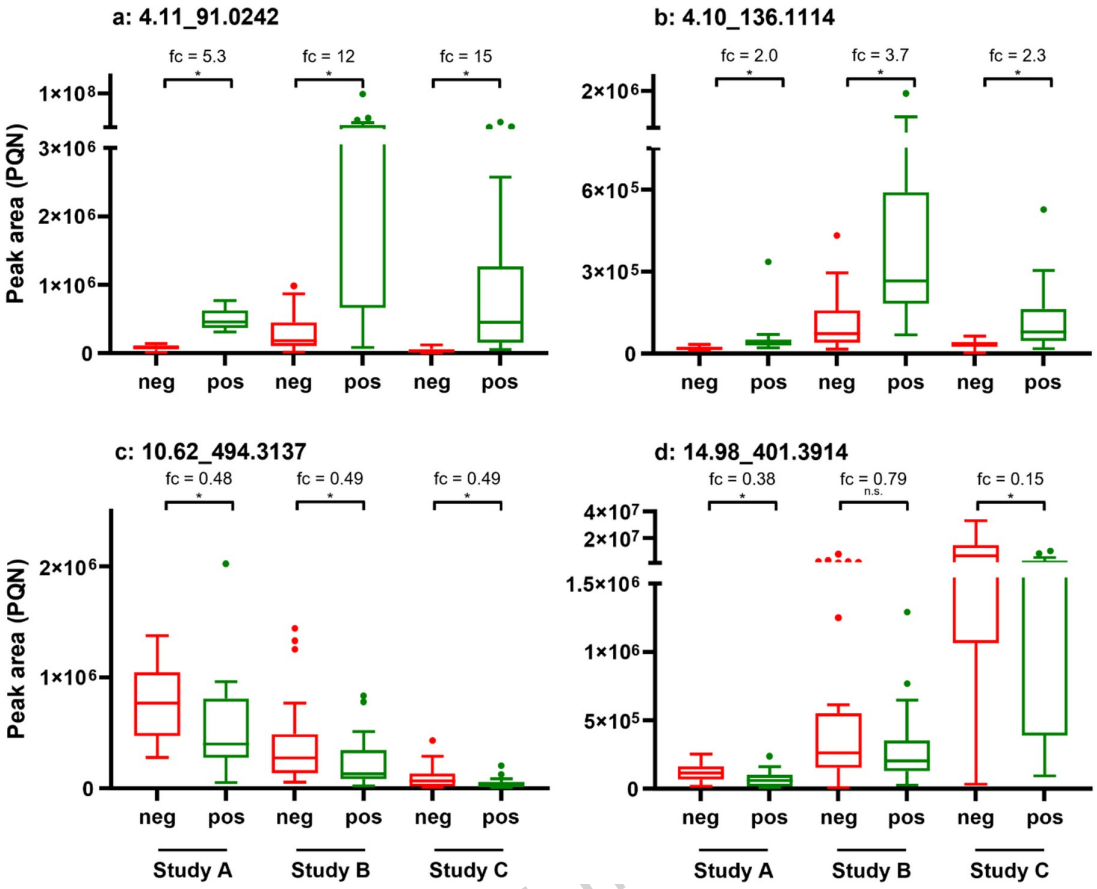
Figure 3: Boxplots of 2 representative features per study A (a, b), B (e, f) and C (c, d) which are only significantly changed in the respective study type. The box represents the median and 25%/75% percentiles, the whiskers indicate the 5-95% percentiles. The y-axis shows the PQN peak area of each feature. Amphetamine-positive groups are given in green and amphetamine-negative ones in red for each study type. Features with a p-value ≤ 0.05 (Wilcoxon rank-sum test) are indicated with the *-symbol. Feature increases or decreases between amphetamine-positive and amphetamine-negative groups are shown as fc-values. Features were considered significant with a p-value ≤ 0.05 and a fc-value ≥ 1.9 or ≤ 0.5.

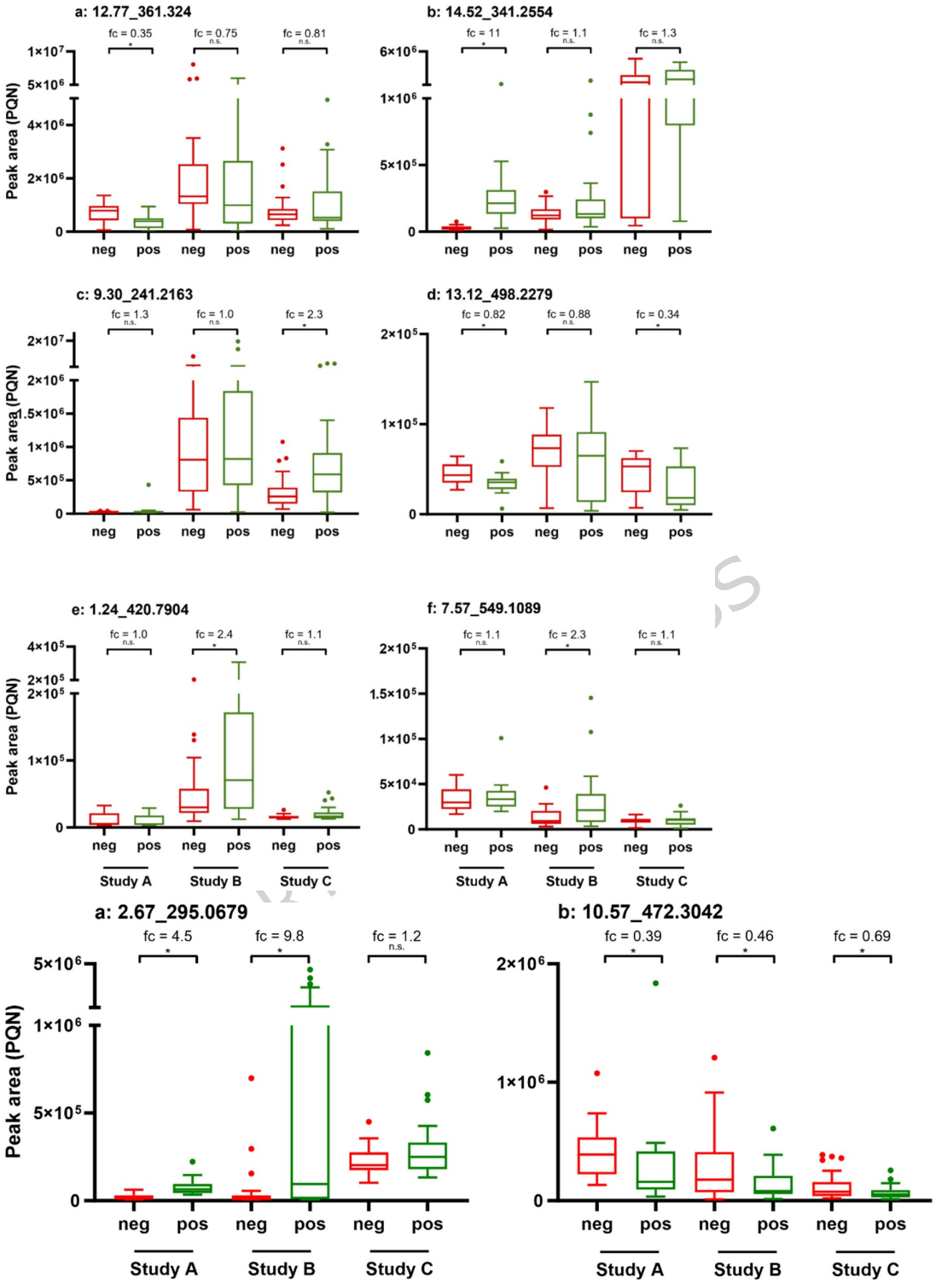
Figure 4: Boxplots of 2 representative features which are significant in studies A and B (but not in C). The box represents the median and 25%/75% percentiles, the whiskers indicate the 5-95% percentiles. The y-axis shows the PQN peak area of each feature. Amphetamine-positive groups are given in green and amphetamine-negative ones in red for each study type. Features with a p-value ≤ 0.05 (Wilcoxon rank-sum test) are indicated with the *-symbol. Feature increases or decreases between amphetamine-positive and amphetamine-negative groups are shown as fc-values. Features were considered significant with a p-value ≤ 0.05 and a fc-value ≥ 1.9 or ≤ 0.5.

Figure 5: Boxplots of 4 representative features which are significant in studies B and C (but not in A). The box represents the median and 25%/75% percentiles, the whiskers indicate the 5-95% percentiles. The y-axis shows the PQN peak area of each feature. Amphetamine-positive groups are given in green and amphetamine-negative ones in red for each study type. Features with a p-value ≤ 0.05 (Wilcoxon rank-sum test) are indicated with the *-symbol. Feature increases or decreases between amphetamine-positive and amphetamine-negative groups are shown as fc-values. Features were considered significant with a p-value ≤ 0.05 and a fc-value ≥ 1.9 or ≤ 0.5.


a: 3.85_211.0571
fc = 1.1
n.s.
fc = 2.7
*
fc = 2.8
*
Peak area (PQN)
4×10⁵
2×10⁵
0
neg
pos
b: 12.92_88.2017
fc = 0.88
n.s.
fc = 0.42
fc = 0.49
6×10⁴
4×10⁴
2×10⁴
c: 14.96_175.1481
fc = 0.97
fc = 0.29
fc = 0.21
4×10⁶
2×10⁶
d: 15.21_764.5564
fc = 0.61
fc = 0.46
fc = 0.44
Study A
Study B
Study C


study A
19
8
1
3
106
13
58
study B
study C




a: 12.77_361.324
b: 14.52_341.2554
c: 9.30_241.2163
d: 13.12_498.2279
e: 1.24_420.7904
f: 7.57_549.1089
a: 2.67_295.0679
b: 10.57_472.3042
Peak area (PQN)
fc = 0.35
fc = 0.75
fc = 0.81
fc = 11
fc = 1.1
fc = 1.3
fc = 1.3
fc = 1.0
fc = 2.3
fc = 0.82
fc = 0.88
fc = 0.34
fc = 1.0
fc = 2.4
fc = 1.1
fc = 1.1
fc = 2.3
fc = 1.1
fc = 4.5
fc = 9.8
fc = 1.2
fc = 0.39
fc = 0.46
fc = 0.69
n.s.
*
neg
pos
Study A
Study B
Study C