# ARTICLE INFORMATION

**Article title**

*Datasets for high-throughput metabolomic CE-MS drugs of abuse screening in clinically depressed patients*

**Authors**

*Liam Surry, \*Philip Britz-Mckibbin.*

**Affiliations**

*McMaster University*

**Corresponding author’s email address and Twitter handle**

*britz@mcmaster.ca*

**Keywords**

*Capillary electrophoresis (CE), mass spectrometry, metabolomics, drug screening, drugs of abuse, urine, automated processing*

**Abstract**

* *Canada has seen an alarming increase in the number of drug-related deaths and poisonings in the past 10 years, representing a public health crisis of staggering proportions. Multi-drug toxicity and unintentional drug overdoses have contributed to this public health crisis. As a result, there is an urgent need for high-throughput analytical screening platforms. Traditional urine drug screening methods such as immunoassays and confirmatory GC-MS/LC-MS testing suffer from false positives and target a limited number of known drug panels. Furthermore, this process requires complex sample preparation, hampering throughput. Here, we used multisegmented injection capillary electrophoresis mass spectrometry (MSI-CE-MS) was used to rapidly and accurately detect a broad range of drugs and their metabolites, including drugs not traditionally analyzed in drug screens directly in urine samples of clinically depressed patients. Urine samples were collected from clinically depressed patients at St. Josephs Healthcare Hamilton (Hamilton, ON, Canada), which is composed of the DISCOVER 9 cohort. Urine was analyzed for illicit, prescription, and over-the-counter drugs in addition to analyzing general metabolites. The dataset consists of a large set of mass-spectrometry metabolomic information corresponding to individual anonymized patients. CE-MS data can be reused by researchers as specific references or snapshots of time points. Furthermore, due to the method of data collection, retrospective analysis can be performed on the data at a later time point for further research. This data can be used by researchers and physicians to monitor psychiatric patients, allowing for fast decision-making in clinical settings (i.e. overdose). Furthermore, this allows for monitoring of the effect of a drug on a patient through metabolism (i.e., whether they are processing the medication fast or slow), allowing for a personalized approach to treatment. This data can better guide the management of drug addiction and mental health outcomes via individualized treatment strategies*

# SPECIFICATIONS TABLE

|  |  |
| --- | --- |
| **Subject** | *Omics: Metabolomics* . |
| **Specific subject area** | *Metabolomics is the scientific study of chemical processes involving metabolites, and provides a molecular link to the phenotype .* |
| **Type of data** | Table, Chart, Graph, Figure, Raw, Processed, Filtered, Analyzed |
| **Data collection** | *Data for drugs of abuse (DoA) screening will be collected after separation using an Agilent 6550 quadrupole time-of-flight (QTOF) mass spectrometer with a dual Agilent JetStream electrospray (dual AJS ESI) ionization source equipped with an Agilent 7100 CE unit. Raw data will be processed in two formats: .d format using the MassHunter Workstation Software (Agilent Technologies, V B.08.00) and .mzML format using PeakMeister, a custom open-source automated data processing algorithm. Data will be normalized using a log-transform and autoscaling approach.* |
| **Data source location** | *McMaster University, Hamilton, Ontario, Canada*  *1280 Main Street West, L8S 4L8*  *43.2609° N, 79.9192° W* |
| **Data accessibility** | ***Data will be made available in a publicly available GitHub repository****:*  Repository name: CE-MS\_DoA\_screening  Data identification number: *TBD*  Direct URL to data: TBD |
| **Related research article** | *DiBattista, A.; Rampersaud, D.; Lee, H.; Kim, M.; Britz-McKibbin, P. High Throughput Screening Method for Systematic Surveillance of Drugs of Abuse by Multisegment Injection–Capillary Electrophoresis–Mass Spectrometry. Anal. Chem.* ***2017****, 89 (21), 11853–11861. https://doi.org/10.1021/acs.analchem.7b03590.* |

# VALUE OF THE DATA

* This data is valuable because currently, there is a worldwide opioid and toxic drug crisis which requires new technologies for reliable high-throughput drug screening. Current gold standard methods (LC-MS) are time-consuming and require complex sample preparation. Physicians require an efficient platform to monitor both illicit medications and prescription medications in patients. Traditional drug screens fail to look at prescription medications such as antipsychotics or antidepressants.
* CE-MS data can be reused by researchers as specific references or snapshots of time points. Furthermore, due to the method of data collection, retrospective analysis can be performed on the data at a later time point for further research.
* Researchers and physicians can use this data to monitor psychiatric patients, allowing for fast decision-making in clinical settings (e.g., overdose). Furthermore, this allows for monitoring a drug's effect on a patient through metabolism (e.g., are they processing the medication fast or slow?), allowing for a personalized approach to treatment.
* |This data can better guide the management of drug addiction and mental health outcomes via individualized treatment strategies

# BACKGROUND

*Canada has seen an alarming increase in the number of drug-related deaths and poisonings, representing a public health crisis of staggering proportions. Since 2016, there have been over 38,000 opioid-related deaths across Canada, primarily impacting young adult males and flattening overall life expectancy. Multi-drug toxicity and unintentional drug overdoses have contributed to an opioid crisis, which stems from adulteration of the drug supply by extremely toxic synthetic opioids and other classes of psychoactive stimulants that evade conventional testing. Furthermore, there has been a significant increase in the use and prescription of psychiatric medications such as antipsychotics and antidepressants for a variety of psychiatric illnesses such as clinical depression, personality disorders, or anxiety disorders. Many of these prescribed medications are not typically analyzed in traditional drug screens, which primarily focus on opioids and related illicit medications. As a result, there is a critical need for high-throughput yet comprehensive analytical platforms to detect and identify the emergence of novel drugs and adulterants. Currently, a two-tiered urine drug testing strategy is widely used based on an immunoassay screen and confirmatory analysis by a test using GC-MS or LC-MS. This approach is prone to false positives, requires complex sample workup and preparation, and targets only a limited number of known drug panels. An alternative approach involves using multisegmented injection capillary electrophoresis mass spectrometry (MSI-CE-MS), which offers a high-throughput approach for drug screening directly in urine, allowing for the analysis of 13 samples in a single analytical run.*

# DATA DESCRIPTION

*The dataset consists of a large set of mass-spectrometry metabolomic information from urine samples from the DISCOVER cohort corresponding to individual anonymized patients*

*Data folders:*

*2024DoA*

*Metadata*

*Raw .mzML data*

*Code*

*Results*

*Results (for individual samples)*

*Plots*

*Metabolites (one plot for each metabolite)*

*Parameters*

*Output\_datafiles*

# EXPERIMENTAL DESIGN, MATERIALS AND METHODS

*Urine samples were collected from clinically depressed psychiatric patients admitted to St. Josephs Healthcare Hamilton (Hamilton, ON, Canada), which is composed of the DISCOVER 9 cohort. Information includes weight, height, biological sex, medications, ethnicity, diet, smoking status, drug use, alcohol use, prescribed medications, basic medical history, and physical health. Patients were given a numerical identifier to identify the sample. Samples were collected adhering to the McMaster Research Ethics Board. The type of sample collected is urine, collected in a plastic cup. All samples are stored at -80oC prior to analysis. Upon analysis, samples are thawed, an aliquot is taken out of the solution, and the sample is diluted four-fold. A series of internal standards will be added, including deuterated drug standards and a series of 5 other internal standards ( to be determined). Samples are then placed in CE-specific 50uL plastic sample vials. Samples are then analyzed using an Agilent 6550 quadrupole time-of-flight (QTOF) mass spectrometer with a dual Agilent JetStream electrospray (dual AJS ESI) ionization source equipped with an Agilent 7100 CE unit. An Agilent 1260 Infinity Isocratic pump and a 1260 Infinity degasser were used to create the coaxial sheath liquid interface. The Agilent mass spectrometer produces vendor-specific .d files. These .d files are processed using the Agilent MassHunter Qualitative Analysis (B.06.00) software. Qualitative analysis extracts the different masses of the metabolites of interest, which are provided by the user, resulting in a specific electropherogram (similar to a chromatogram). Electropherograms are processed in profile mode with a 10 ppm mass window. Peaks are then smoothed using a quadratic/cubic Savitzky-Golay function with a strength of 7 points. Peaks are then integrated using the agile2 integrator provided in the software. However, manual adjustment or quality checks are done on every single peak to ensure accuracy. Raw data (peak area and migration time) are then copied into excel and exported as .xlsx and .csv formats. Data will additionally be processed in tandem using a lab-developed open-source software called PeakMeister, written in R. Since vendor-specific instruments typically output vendor-specific files that are locked and require the vendor software to open; more focus has been placed on using open-source platforms. Vendor-specific .d files are converted to the open-source .mzML format using ProteoWizard’s msConvert. A user-supplied parameters .csv file is used to control what PeakMeister does. Raw .mzML data is then loaded into PeakMeister, which then calculates the migration time indices, performs mass calibration, extracts the electropherograms from the raw data, smooths them using a Gaussian kernel smoother, detects the internal standard and metabolite peaks, integrates peaks, filters peaks using peak height and full-width-at-half-maximum, and then plots the electropherograms. Processed data is exported in .csv format, which includes the migration index or relative migration times used, a copy of the parameters used, the migration times of peaks, and the peak areas. The results from each sample are stored in a folder called Results\_samplename. All parameter files are stored in a folder structure called “Parameters” that is included in each results folder. All of the formats mentioned will allow for data re-use, sharing, and preservation if someone has Qualitative Analysis, excel, or R installed, with the latter software R being open source. The .d files will be named based on the date the samples were run, day, month, and year (i.e. 24102024) with \_initial to indicate who ran it (i.e. 24102024\_LS) and a numerical identifier to indicate the sample it belonged to (i.e. 001\_24102024\_LS). All relevant sample metadata (i.e. identification) is stored in a .csv file named 2024DoA\_sample\_metadata.csv.*

# PRESERVATION AND STORAGE

*This project is anticipated to generate terabytes of data and will require an external hard drive to store data throughout the project (5 terabytes in size). Project data will be stored here for the duration of the project, after which it will be stored indefinitely. Raw MS data will be uploaded to NIST for other researchers to use openly. Patient identification data will be stored in a secure location and password protected. Code will be stored in a GitHub repository (available to the public), and relevant metadata will be uploaded to FigShare. A detailed README file will accompany raw data and GitHub code*

# RESPONSIBILITIES

*The individual responsible for managing the project and accompanying data is Liam Surry. Dr. Britz-McKibbin will be responsible for managing the project when the project is finished or when Liam Surry leaves the laboratory and is additionally responsible for who has access to the data. In the event that the principal investigator changes, the data management protocols and subsequent access to the data will be passed on to the next principal investigator*

# LIMITATIONS

*None*

# ETHICS STATEMENT AND LEGAL COMPLIANCE

*Any sensitive data collected on human subjects will be upheld to the highest ethical standards and guidelines outlined by the McMaster Research Ethics Board. Prior to this approval, no data from humans will be collected.*

*Sensitive data will be treated using the highest ethical standards and the McMaster Research Ethics Board guidelines. No data from human samples will be collected before the approval of  
 the relevant ethics committee has been obtained.  
 Person-identifiable data will not leave the unit from which they originated, and keys to identification numbers will be held confidentially within the respective clinical units.*

*All drug standards and associated deuterated internal standards were sourced from Cerriliant (Round Rock, TX, USA) and imported into Canada following all legal requirements. Drug standards are stored in a separate, locked, and secured fridge in the lab, separated from other chemicals.*

# CRediT AUTHOR STATEMENT

*Liam Surry – Conceptualization, methodology, validation, formal analysis, investigation, data curation, writing – original draft, writing – review and editing, visualization*

*Philip Britz-McKibbin - funding acquisition, resources, supervision, project administration, writing – reviewing and editing*

# ACKNOWLEDGEMENTS

*The authors acknowledge funding support from the Natural Sciences and Engineering Research Council of Canada, the Canada Foundation for Innovation, Genome Canada, and McMaster University.*

# DECLARATION OF COMPETING INTERESTS

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

# REFERENCES

*TBD*