Data Descriptor

Urinary Metabolite Panel Dataset for Bulgarian Children with Autism Spectrum Disorder (ASD)

Victor Slavov *, Lubomir Traikov, Stanislava Ciurinskiene, Maria Savcheva, Till Heine, Radka Tafradjiiska-Hadjiolova, Alexandra Zlatarova, Ivan Tourtourikov, Dilyana Madzharova, Anita Kavrakova, Tanya Kadiyska

Posted Date: 17 March 2026

doi: 10.20944/preprints202603.1210.v1

Keywords: autism spectrum disorder; urine metabolites; creatinine normalization; neurotransmitters; one-carbon metabolism; organic acids; microbiome-related amines; Bulgaria

 doi:10.20944/preprints202603.1210.v1

Data Descriptor

Urinary Metabolite Panel Dataset for Bulgarian Children with Autism Spectrum Disorder (ASD)

Victor Slavov [1,*], Lubomir Traikov [1], Stanislava Ciurinskiene [2], Maria Savcheva [3], Till Heine [4], Radka Tafradjiiska-Hadjiolova [5], Alexandra Zlatarova [5], Ivan Tourtourikov [6], Dilyana Madzharova [6], Anita Kavrakova [6] and Tanya Kadiyska [5,6]

[1] Department of Medical Physics and Biophysics, Medical University of Sofia, 1431 Sofia, Bulgaria
[2] Vsiaka Duma Society, 1000 Sofia, Bulgaria
[3] Laboratory of Transfusion Hematology, Saint Sophia Hospital, 1618 Sofia, Bulgaria
[4] Biovis Diagnostik MVZ GmbH, Limburg, Germany
[5] Department of Physiology and Pathophysiology, Medical University of Sofia, 1431 Sofia, Bulgaria
[6] Genetic Medico-Diagnostic Laboratory Genica and Genome Center Bulgaria, 1612 Sofia, Bulgaria
* Correspondence: vslavov@medfac.mu-sofia.bg

Abstract

This Data Descriptor presents an anonymized, shuffled dataset of creatinine-normalized urinary metabolite measurements from 73 Bulgarian children with autism spectrum disorder (ASD), released to support reuse in secondary analyses and cross-cohort comparisons. Spot urine results are provided as individual-level values after creatinine normalization; for trimethylamine, values below the limit of quantification (LOQ) were replaced with LOQ/2. The deposit contains measurements for 24 urinary markers grouped into three functional classes (neurotransmitters and aromatic amino acid precursors; one-carbon/methylation and vitamin-related metabolites; and energy metabolism/organic acids with microbiome-related amines). The release includes the results table in both XLSX and CSV formats, a reference limits and units file for contextual interpretation, a data dictionary, a README, a changelog, and SHA-256 checksums for integrity verification. Cohort-level demographics and additional sampling details are described in the companion publication cited in the main text. Dataset: https://doi.org/10.5281/zenodo.18614881 Dataset License: Creative Commons Attribution 4.0 International.

Keywords: autism spectrum disorder; urine metabolites; creatinine normalization; neurotransmitters; one-carbon metabolism; organic acids; microbiome-related amines; Bulgaria

1. Summary

This release provides anonymized, shuffled participant-level urinary measurements for 24 creatinine-normalized markers in a pediatric autism spectrum disorder (ASD) case cohort from Bulgaria (n = 73). The cohort comprised children aged 3–13 years; median age (interquartile range, IQR) was 5 (4–7) years, including 57 males and 16 females, and no contemporaneous neurotypical control group was enrolled (Table 1). Spot urine collection followed a second-morning, midstream, acid-stabilized protocol using a commercial kit (T928; biovis Diagnostik MVZ GmbH, Limburg, Germany), with collection 2–4 h after the first morning void and acidification verified using the provided pH strip. The dataset was deposited to enable reuse of a case-only urinary marker panel in descriptive distribution analyses, exploratory feature screening, and cross-cohort comparisons that incorporate external control data or published reference limits. Additional cohort-level and pre-analytical methodological details are described in the companion publication [1].

2. Data Description

 doi:10.20944/preprints202603.1210.v1

2.1. Files in This Release

The Zenodo record [2] includes the following files:

- ASD_urine_metabolites_panel_v1_results_anonymized_shuffled.xlsx (primary results table in Excel format);
- ASD_urine_metabolites_panel_v1_results_anonymized_shuffled.csv (primary results table in CSV format);
- ASD_v1_reference_ranges_units.csv (reporting units and laboratory reference limits);
- ASD_v1_data_dictionary.csv (variable definitions, units, and reference metadata);
- ASD_v1_README.md (dataset overview and usage notes);
- ASD_v1_CHANGELOG.md (release version history);
- ASD_v1_SHA256SUMS.txt (SHA-256 checksums for integrity verification).

2.2. Results Table

The primary results table is provided in both .xlsx and .csv formats. It contains one row per participant and one column for each of the 24 creatinine-normalized urinary metabolite variables included in the panel. The public results table contains de-identified analytical variables only. Individual-level age and sex are reported elsewhere [1]. Variable-level metadata are provided in ASD_v1_data_dictionary.csv.

2.3. Reference Limits File

ASD_v1_reference_ranges_units.csv provides the reporting units and laboratory reference limits for the panel variables. The file contains the metabolite name, reporting units, reference type, and lower and upper bounds where applicable. One-sided decision limits are encoded as single-threshold entries (< or >), whereas two-sided limits are encoded as ranges.

2.4. De-identification and Scope Boundaries

Participant rows are anonymized and shuffled. The public dataset contains no direct identifiers, dates, locations, or individual-level demographic variables. Cohort-level demographic summaries and clinical context are reported in the companion publication [1].

2.5. Summary Tables and Visualization

Table 1 provides an overview of the dataset and analytical characteristics. Table 2 summarizes each marker relative to laboratory-provided reference limits. Figure 1 visualizes marker distributions grouped into three functional classes.

Table 1. Dataset characteristics and analytical notes for the urinary metabolite panel dataset.

Characteristic	Value
Participants, *n*	73
Age range, years	3–13
Analytes, *n*	24
Data matrix dimensions (rows × columns)	73 × 24
Reporting basis	Creatinine-normalized spot urine concentrations
Units (as provided by laboratory)	µg/g creatinine; µmol/g creatinine; mg/g creatinine; ratio

 doi:10.20944/preprints202603.1210.v1

Reference limits	Laboratory-provided reference intervals/decision limits (two-sided ranges or one-sided thresholds)
Qualified results handling	Trimethylamine: values reported by the laboratory as <0.1 μmol/g creatinine were converted to LOQ/2 = 0.05 (*n* = 64, 87.7%)
Missing values	None (0/73 × 24)
De-identification	Anonymized and shuffled; no demographics or dates included in the data file

Note: Values are creatinine-normalized spot urine results. The deposited tabular dataset contains anonymized, shuffled participant rows and results for 24 urinary metabolite markers/ratios. Individual-level demographic variables (age, sex) and urinary creatinine distribution are not included in the public dataset; only de-identified metabolite results are provided. Reference limits are laboratory-provided and shown for contextual interpretation. Abbreviations: SAM/SAH, S-adenosylmethionine/S-adenosylhomocysteine; LOQ, limit of quantification.

Table 2. Summary of urinary metabolite panel markers relative to laboratory-provided reference limits for contextual interpretation in the ASD cohort (*n* = 73).

Analyte	Unit	Reference limit	Median (IQR)	Below ref *n* (%)	Above ref *n* (%)
Dopamine	μg/g creatinine	130–240	578.35 (483.42–750.03)	0 (0.0)	72 (98.6)
Noradrenaline	μg/g creatinine	15–36	48.78 (34.20–62.63)	3 (4.1)	51 (69.9)
Adrenaline	μg/g creatinine	2–5.5	8.69 (5.34–12.80)	1 (1.4)	53 (72.6)
Serotonin	μg/g creatinine	80–190	248.20 (199.45–300.90)	0 (0.0)	58 (79.5)
GABA	μmol/g creatinine	1.5–5	3.69 (2.55–5.20)	2 (2.7)	19 (26.0)
Glutamate	μmol/g creatinine	8–25	21.27 (14.72–30.10)	0 (0.0)	22 (30.1)
Phenylalanine	μmol/g creatinine	>31	94.90 (73.10–113.70)	0 (0.0)	0 (0.0)
Tyrosine	μmol/g creatinine	>42	138.00 (105.30–171.40)	0 (0.0)	0 (0.0)
Cystathionine	μmol/g creatinine	<25	9.50 (5.10–13.90)	0 (0.0)	5 (6.8)
Methylmalonic acid	mg/g creatinine	<1.8	1.66 (1.33–2.14)	0 (0.0)	32 (43.8)
Nicotinic acid	μmol/g creatinine	>0.5	1.80 (1.40–2.40)	1 (1.4)	0 (0.0)
Nicotinamide	μmol/g creatinine	>1.2	3.40 (2.64–4.36)	7 (9.6)	0 (0.0)
S-Adenosyl Methionine	μmol/g creatinine	>7.5	20.10 (17.50–25.20)	1 (1.4)	0 (0.0)

 doi:10.20944/preprints202603.1210.v1

Betaine	µmol/g creatinine	29–85	100.50 (66.70–151.70)	2 (2.7)	42 (57.5)
Choline	µmol/g creatinine	13–30	36.00 (26.10–47.10)	0 (0.0)	47 (64.4)
SAM/SAH	ratio	>9	16.10 (12.90–18.40)	5 (6.8)	0 (0.0)
Citrulline	µmol/g creatinine	<4	5.08 (4.30–7.19)	0 (0.0)	57 (78.1)
Citrate	mg/g creatinine	160–786	683.42 (425.74–964.60)	4 (5.5)	29 (39.7)
Lactate	mg/g creatinine	1.7–20.5	5.89 (4.03–8.88)	3 (4.1)	3 (4.1)
Pyruvate	mg/g creatinine	<5.4	5.39 (3.51–6.73)	0 (0.0)	35 (47.9)
Suberic acid	mg/g creatinine	<1.9	2.81 (1.89–4.44)	0 (0.0)	54 (74.0)
Carnitine	µmol/g creatinine	11–90	67.10 (21.60–166.00)	7 (9.6)	31 (42.5)
Trimethylamine N-Oxide	µmol/g creatinine	<600	413.10 (248.40–627.40)	0 (0.0)	20 (27.4)
Trimethylamine	µmol/g creatinine	<0.1	0.05 (0.05–0.05)	0 (0.0)	9 (12.3)

Note: Reference limits are laboratory-provided and are shown for contextual interpretation only, as no neurotypical control group was enrolled. For one-sided limits (< or >), only the indicated direction is considered outside reference. Values reported by the laboratory as below the quantification limit for trimethylamine (<0.1 µmol/g creatinine) were set to LOQ/2 = 0.05 µmol/g creatinine for summary statistics. Abbreviations: ASD, autism spectrum disorder; IQR, interquartile range; GABA, γ-aminobutyric acid; SAM/SAH, S-adenosylmethionine/S-adenosylhomocysteine; LOQ, limit of quantification.

 doi:10.20944/preprints202603.1210.v1

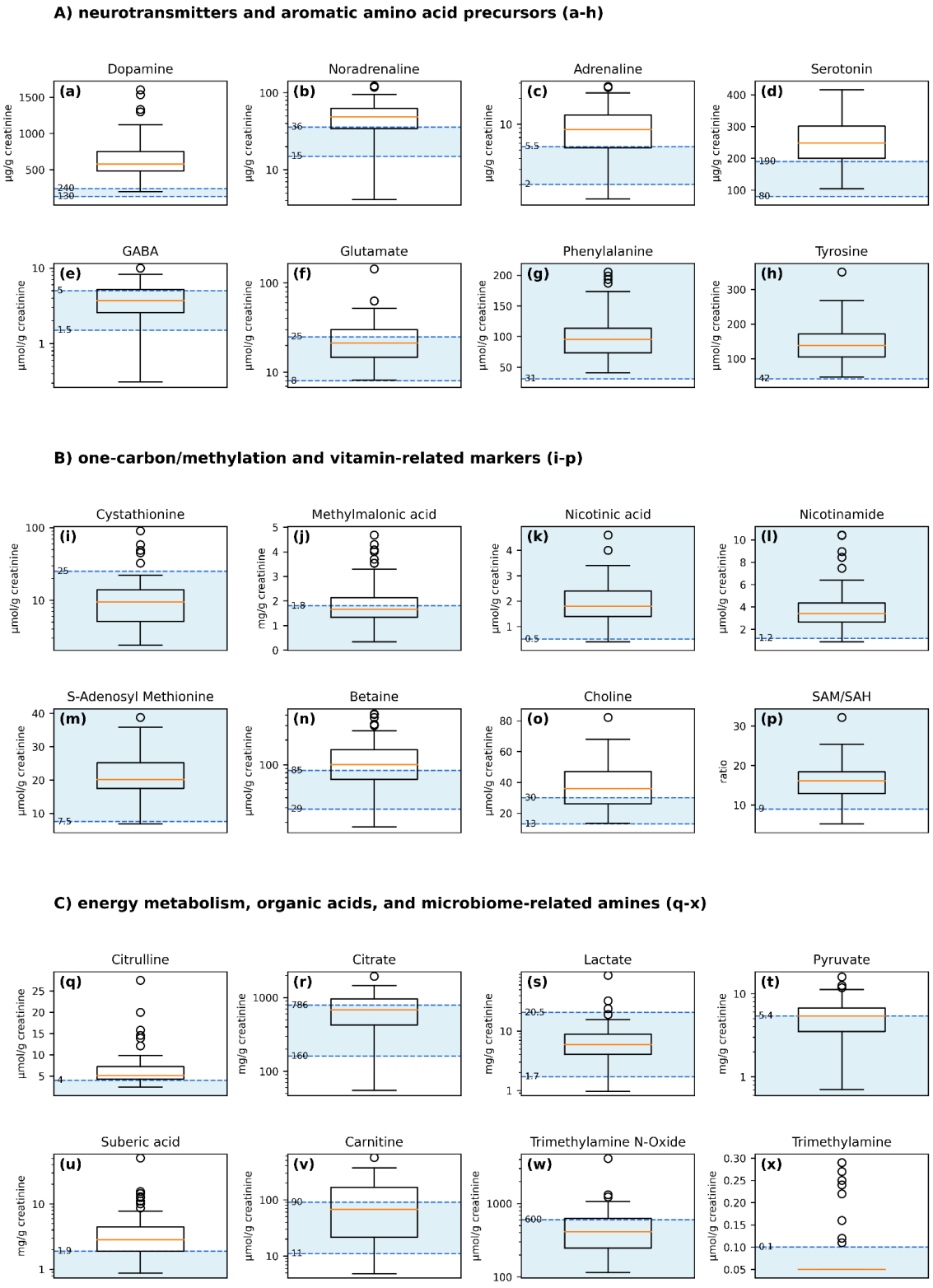

Figure 1. Creatinine-normalized distributions of urinary metabolite panel markers in the ASD cohort (n = 73), grouped by functional class: A) neurotransmitters and aromatic amino acid precursors (a–h); B) one-carbon/methylation and vitamin-related markers (i–p); C) energy metabolism, organic acids, and microbiome-related amines (q–x). Dashed lines indicate laboratory-provided reference limits, and shaded regions represent the corresponding reference ranges where applicable (two-sided ranges or one-sided thresholds). Boxplots show

 doi:10.20944/preprints202603.1210.v1

the median (center line), IQR (box), whiskers extend to the most extreme values within 1.5 × IQR, and points denote outliers. Panels (b), (c), (e), (f), (i), (r), (s), (t), (u), (v), and (w) are displayed on a logarithmic y-axis. Abbreviations: ASD, autism spectrum disorder; IQR, interquartile range.

3. Methods

3.1. Urine Collection

Urine collection for this cohort is described in the published Metabolites article [1]. Briefly, second-morning, midstream, acid-stabilized spot urine was collected at home using an acid-stabilized urine collection kit (T928; biovis Diagnostik MVZ GmbH) within the provider's second-morning workflow. Caregivers were instructed to collect the second-morning void approximately 2–4 h after the first-morning urine (minimum 2 h), avoid excessive fluid intake before sampling, refrain from strenuous physical activity on the morning of collection, and avoid caffeine-containing beverages and energy drinks for at least 16 h prior to collection. Recruitment and pre-analytical procedures (collection, stabilization, storage, and shipment logistics) were coordinated via Genome Center Bulgaria (Sofia, Bulgaria), and samples were shipped to the German provider laboratory for analytical measurement and report release. The present 24-marker Zenodo release derives from the same cohort and pre-analytical workflow.

3.2. Laboratory Measurement, Units, and Normalization

Analytical measurement for the cohort is described in [1]. Urinary metabolite measurements were performed by LC–MS/MS, whereas spot urinary creatinine was measured by a routine enzymatic method (UV/VIS photometric assay). Metabolite results were reported by the provider as creatinine-normalized concentrations in marker-specific units. Laboratory reference limits were taken from the commercial documentation accompanying the analytical service and are defined for the same second-morning, acid-stabilized urine collection and reporting units used by the provider. The present Zenodo release provides creatinine-normalized values for a 24-marker subset; the same LC–MS/MS-based analytical workflow and creatinine normalization framework were used for this subset release.

3.3. Handling of Qualified Results

Trimethylamine values reported as below the quantification limit (limit of quantification, LOQ) (<0.1 µmol/g creatinine) were set to LOQ/2 = 0.05 µmol/g creatinine for summary statistics and visualization (see Table 2 note).

3.4. Software for Figures

As reported in [1], statistical analyses were performed in Python (v3.11) using SciPy and statsmodels. Figure 1 for the present 24-marker release was generated from the curated subset in the same analysis environment.

3.5. Ethics Approval and Informed Consent

The study was conducted in accordance with the Declaration of Helsinki and approved by the Research Ethics Committee of University Hospital "St. Ivan Rilski", Sofia, Bulgaria (protocol number 3/26.04.2023, date of approval: 26 April 2023). Informed consent was obtained from the parents/guardians of all subjects involved in the study.

4. Usage Notes

This case-only, cross-sectional dataset is intended primarily for descriptive analyses, exploratory feature screening, and comparisons with external cohorts or laboratory-provided reference

 doi:10.20944/preprints202603.1210.v1

benchmarks. Because no contemporaneous neurotypical control group was enrolled, the deposited values and the proportions outside laboratory-provided reference limits should be interpreted as contextual descriptors rather than evidence of ASD-specific abnormalities.

The public release contains de-identified analytical variables only. Individual-level demographic variables, dates, urinary creatinine, and richer clinical covariates are not included in the deposited files; therefore, age- or sex-stratified analyses, direct assessment of urine dilution effects, and covariate-adjusted analyses are not possible from the public dataset alone. Users should consult the data dictionary for variable definitions and units and the reference-limits file for reference type (two-sided ranges or one-sided thresholds). As with other creatinine-normalized spot-urine datasets, some residual variability related to urine concentration and pre-analytical conditions may remain. Trimethylamine should be interpreted with particular caution, because values reported by the laboratory as below the quantification limit were represented as LOQ/2 in the public release.

Author Contributions: Conceptualization, V.S.; methodology, V.S. and T.H.; formal analysis, V.S., D.M., and A.K.; investigation/resources, S.C., M.S. and I.T.; visualization, V.S. and R.T.-H.; writing—original draft preparation, V.S. and A.Z.; writing—review and editing, L.T. and T.K.; supervision, L.T. and T.K.; project administration, R.T.-H. and T.K.; funding acquisition, R.T.-H. and T.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Research Ethics Committee of University Hospital "St. Ivan Rilski", Sofia, Bulgaria (protocol number 3/26.04.2023, date of approval: 26 April 2023).

Informed Consent Statement: Informed consent was obtained from the parents/guardians of all subjects involved in the study.

Data Availability Statement: The dataset presented in this study is publicly available in Zenodo at https://doi.org/10.5281/zenodo.18614881. The deposited record includes the primary results tables (.csv and .xlsx), the data dictionary, the reference-limits file, and accompanying documentation. Individual-level demographic variables and dates are not included in the public dataset.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ASD	Autism spectrum disorder
GABA	γ-Aminobutyric acid
IQR	Interquartile range
LC–MS/MS	Liquid chromatography–tandem mass spectrometry
LOQ	Limit of quantification
SAM	S-adenosylmethionine
SAH	S-adenosylhomocysteine
UV/VIS	Ultraviolet–visible spectrophotometry

References

1. Slavov, V.; Traikov, L.; Ciurinskiene, S.; Tafradjiiska-Hadjiolova, R.; Kadiyska, T. Urinary Tryptophan–Kynurenine Pathway Profiling in Bulgarian Children with Autism Spectrum Disorder (ASD): Neopterin Co-Varies with Kynurenine and Quinolinic Acid. *Metabolites* **2026**, *16*, 169. https://doi.org/10.3390/metabo16030169
2. Slavov, V. *Urinary metabolite panel dataset for Bulgarian children with ASD (N=73; ages 3–13), v1*. Zenodo. Available online: https://doi.org/10.5281/zenodo.18614881

 doi:10.20944/preprints202603.1210.v1