Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Mapping Neurotoxic Mixture Burdens Across European Cities

Bin Wang

binwang@pku.edu.cn

Peking University https://orcid.org/0000-0002-1164-8430

Nan Fu

Peking University

Shuxin An

Peking University

Shuo Yang

Peking University

Qianyi Wang

Peking University

Tianxiang Wu

Peking University

Thanh Wang

Linköping University

Zhijiang Chen

Zhejiang Provincial Center for Disease Control and Prevention

Xiaoming Lou

Zhejiang Provincial Center for Disease Control and Prevention

Hanbin Zhang

Imperial College London

Mengyuan Ren

Emory University https://orcid.org/0000-0003-3819-4181

Jian Xu

Chinese Research Academy of Environmental Sciences

Jingchuan Xue

Guangdong Basic Research Center of Excellence for Ecological Security and Green Development, Key Laboratory for City Cluster Environmental Safety and Green Development of the Ministry of Education,

Xiaoxia Lu

Peking University

Yichao Huang

Anhui Medical University https://orcid.org/0000-0003-1186-1300

Mingliang Fang
Department of Environmental Science and Engineering, Fudan University https://orcid.org/0000-0002-2204-9783
Pablo Gago-Ferrero
Institute of Environmental Assessment and Water Research
Nikiforos Alygizakis
Swedish University of Agricultural Sciences
Fangang Meng
Department of Neurosurgery, Beijing Tiantan Hospital, Capital Medical University, Beijing 100070, China
Xiqing Li
Peking University

Article

Keywords: Wastewater, neurotoxic chemicals, urban population surveillance, risk assessment

Posted Date: November 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11693239/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Title: Mapping Neurotoxic Mixture Burdens Across European Cities

Author: Nan Fu[1], Shuxin An[1], Shuo Yang[2,3], Qianyi Wang[2,3], Tianxiang Wu[2,3], Thanh Wang[4,5], Zhijian Chen[6], Xiaoming Lou[6], Hanbin Zhang[7,8], Mengyuan Ren[9], Jian Xu[10], Jingchuan Xue[11], Xiaoxia Lu[1], Yichao Huang[12], Mingliang Fang[13], Pablo Gago-Ferrero[14], Nikiforos Alygizakis[15], Fangang Meng[16], Xiqing Li[1*], Bin Wang[1,2,3,17*]

[1] Laboratory for Earth Surface Processes, College of Urban and Environmental Science, Peking University, Beijing 100871, China
[2] Institute of Reproductive and Child Health, Peking University/Key Laboratory of Reproductive Health, National Health Commission of the People's Republic of China, Beijing 100191, China
[3] Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing 100191, China
[4] State Key Laboratory of Environmental Chemistry and Toxicology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China
[5] Department of Physics, Chemistry and Biology (IFM), Linköping University, Linköping SE-58183, Sweden.
[6] Department of Environmental Health, Zhejiang Provincial Center for Disease Control and Prevention, 3399 Binsheng Road, Hangzhou 310051, China.
[7] European Centre for Environment and Human Health, University of Exeter Medical School, Penryn, United Kingdom, TR10 8RD;
[8] Environmental Research Group, Faculty of Medicine, Imperial College London, London, SW7 2AZ.
[9] Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University, 1518 Clifton Road, Atlanta, GA, 30322, USA
[10] State Key Laboratory of Environmental Criteria and Risk Assessment, Chinese Research Academy of Environmental Sciences, Beijing 100012, China.
[11] Guangdong Basic Research Center of Excellence for Ecological Security and Green Development, Key Laboratory for City Cluster Environmental Safety and Green Development of the Ministry of Education, School of Ecology, Environment and Ocean, Guangdong University of Technology, Guangzhou 510006, China
[12] School of Public Health, Key Laboratory of Environmental Toxicology of Anhui Higher Education Institutes, Center for Big Data and Population Health of IHM and MOE Key Laboratory of Population Health Across Life Cycle, Anhui Medical University, Hefei, 230032, China
[13] Department of Environmental Science and Engineering, Fudan University, Shanghai 200433, China
[14] Institute of Environmental Assessment and Water Research, IDAEA-CSIC, 08034 Barcelona, Spain.
[15] Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, Uppsala SE-75007, Sweden
[16] Beijing Neurosurgical Institute, Beijing Tiantan Hospital, Capital Medical University, Beijing, 100070, China
[17] Key Laboratory of Epidemiology of Major Diseases (Peking University), Ministry of Education, Beijing 100191, China

[*]Correspondence should be addressed to Dr. Bin Wang (Email: binwang@pku.edu.cn), and Dr. Xiqing Li (Email: xli@urban.pku.edu.cn)

SUMMARY

Introduction:

Neurological disorders are among the leading contributors to global disease burden. Exposure to neurotoxic chemicals can impair neurological health, particularly during development. Such exposures arise from consumer products, food, pharmaceuticals, pesticides and the broader environment, often involving mixtures rather than individual compounds. European biomonitoring has documented widespread exposure to multiple priority chemical classes. Experimental evidence further indicates that chemicals below their individual effect thresholds can contribute additively to neurite outgrowth inhibition. Yet current surveillance remains constrained by uneven chemical coverage and an emphasis on occurrence or concentration rather than biological relevance.

Rationale:

Wastewater provides a complementary population-scale view because municipal influent integrates chemical signals from large contributing populations and can be sampled repeatedly without individual recruitment. We developed a Wastewater-based framework for Prioritizing Neurotoxic Chemical Mixtures (WPN) to translate wastewater signals into biologically informed mixture priorities across European cities. Target chemicals for health risk assessment were identified by integrating wastewater occurrence, neurotoxicity evidence, and structured exposure-biology relationships. Georeferenced NORMAN records were harmonized across 117 GUB (Global Urban Boundaries) urban areas in 18 European countries, and sparse monitoring data were reconstructed while retaining all measured values. WPN sequentially translated wastewater concentration into estimated daily exposure, model-derived internal-equivalent concentration, and neurite-outgrowth-inhibition–weighted bioanalytical equivalent concentration (BEQ).

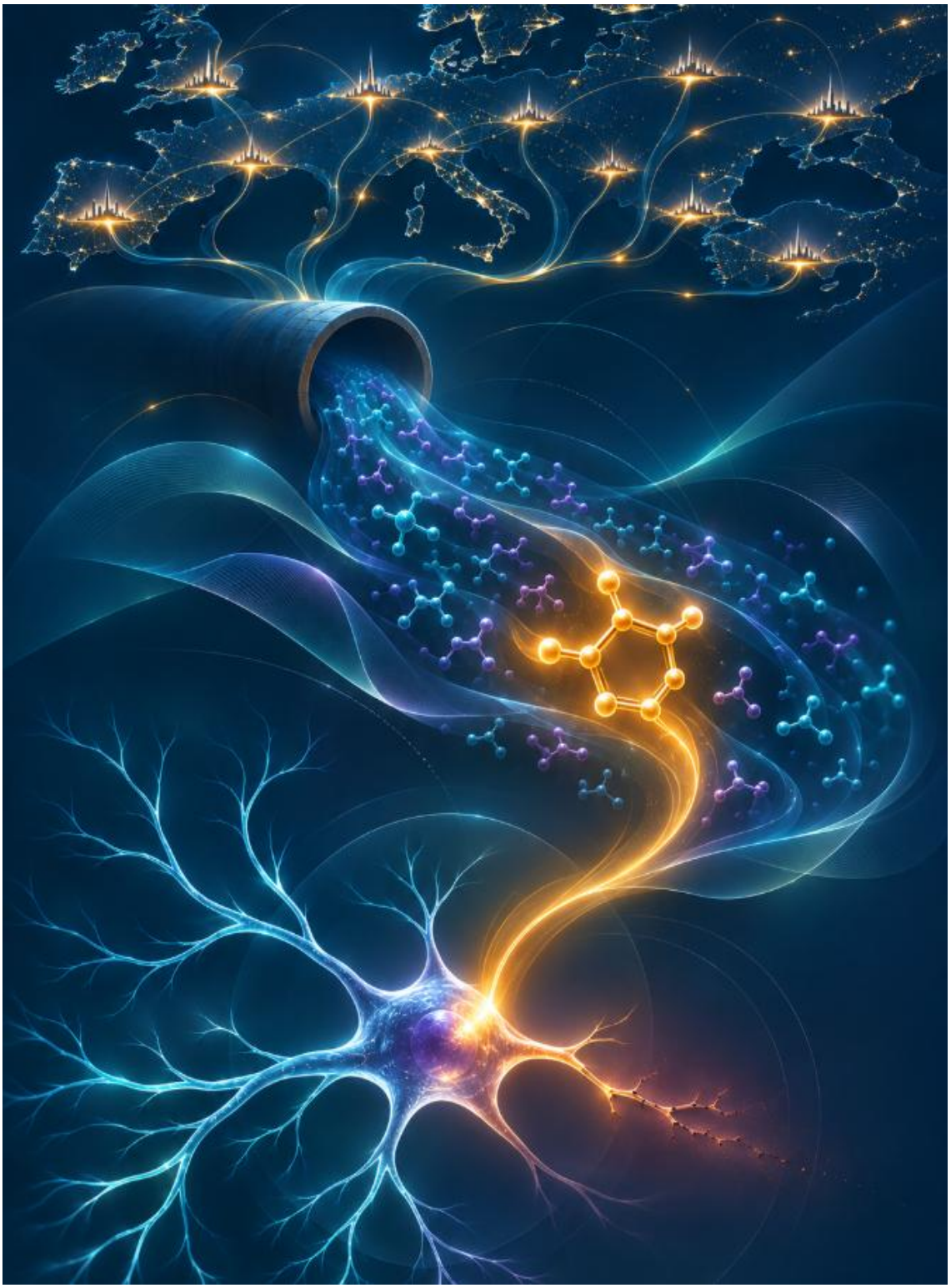
Results:

We identified 43 exogenous neurotoxic chemicals spanning eight major chemical classes. Biological translation markedly reorganized chemical priorities. Although wastewater concentration and estimated exposure rankings remained positively associated, nearly half of the chemicals changed priority stratum after the first translation. Subsequent toxicokinetic and potency weighting produced a stronger reordering: wastewater abundance and BEQ rankings shared only two chemicals among their top ten. The clearest example was 1,2-benzisothiazol-3(2H)-one (BIT), which ranked only 18th by wastewater concentration but dominated the effect-weighted mixture profile, accounting for approximately 67.7% of mean BEQ. Removing BIT revealed a broader set of secondary contributors, including Tris(1,3-dichloro-2-propyl) phosphate, Gabapentin and PFOA. Total modeled BEQ varied comparatively little among cities despite substantial differences in chemical composition, indicating that similar aggregate burdens can arise from distinct mixture profiles. National food-supply structure showed detectable matrix-level covariation with wastewater chemistry, but individual associations were weak after multiple-testing correction and explained only a limited fraction of effect-weighted chemical priority. The interactive WPN platform is available at https://www.exposomex.cn/wpn.

Conclusion:

Our findings show that chemical abundance is not equivalent to biological importance. By integrating wastewater occurrence with exposure, toxicokinetics, and endpoint-specific potency, WPN shifts urban chemical surveillance from concentration-based monitoring toward effect-informed mixture prioritization. This framework provides a scalable basis for identifying priority chemicals and cities for targeted biomonitoring and experimental validation.

Cover Art



Illustrations: Wastewater as a city-scale lens translates complex chemical mixtures across Europe into biologically meaningful neurotoxic signals. Flowing chemical signatures converge from urban populations, are transformed through exposure and effect-based prioritization, and ultimately connect to neuronal responses, highlighting the shift from chemical abundance to biological relevance.

Title: Mapping Neurotoxic Mixture Burdens Across European Cities

Author: Nan Fu[1], Shuxin An[1], Shuo Yang[2,3], Qianyi Wang[2,3], Tianxiang Wu[2,3], Thanh Wang[4,5], Zhijian Chen[6], Xiaoming Lou[6], Hanbin Zhang[7,8], Mengyuan Ren[9], Jian Xu[10], Jingchuan Xue[11], Xiaoxia Lu[1], Yichao Huang[12], Mingliang Fang[13], Pablo Gago-Ferrero[14], Nikiforos Alygizakis[15], Fangang Meng[16], Xiqing Li[1*], Bin Wang[1,2,3,17*]

[1] Laboratory for Earth Surface Processes, College of Urban and Environmental Science, Peking University, Beijing 100871, China
[2] Institute of Reproductive and Child Health, Peking University/Key Laboratory of Reproductive Health, National Health Commission of the People's Republic of China, Beijing 100191, China
[3] Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing 100191, China
[4] State Key Laboratory of Environmental Chemistry and Toxicology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China
[5] Department of Physics, Chemistry and Biology (IFM), Linköping University, Linköping SE-58183, Sweden.
[6] Department of Environmental Health, Zhejiang Provincial Center for Disease Control and Prevention, 3399 Binsheng Road, Hangzhou 310051, China.
[7] European Centre for Environment and Human Health, University of Exeter Medical School, Penryn, United Kingdom, TR10 8RD;
[8] Environmental Research Group, Faculty of Medicine, Imperial College London, London, SW7 2AZ.
[9] Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University, 1518 Clifton Road, Atlanta, GA, 30322, USA
[10] State Key Laboratory of Environmental Criteria and Risk Assessment, Chinese Research Academy of Environmental Sciences, Beijing 100012, China.
[11] Guangdong Basic Research Center of Excellence for Ecological Security and Green Development, Key Laboratory for City Cluster Environmental Safety and Green Development of the Ministry of Education, School of Ecology, Environment and Ocean, Guangdong University of Technology, Guangzhou 510006, China
[12] School of Public Health, Key Laboratory of Environmental Toxicology of Anhui Higher Education Institutes, Center for Big Data and Population Health of IHM and MOE Key Laboratory of Population Health Across Life Cycle, Anhui Medical University, Hefei, 230032, China
[13] Department of Environmental Science and Engineering, Fudan University, Shanghai 200433, China
[14] Institute of Environmental Assessment and Water Research, IDAEA-CSIC, 08034 Barcelona, Spain.
[15] Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, Uppsala SE-75007, Sweden
[16] Beijing Neurosurgical Institute, Beijing Tiantan Hospital, Capital Medical University, Beijing, 100070, China
[17] Key Laboratory of Epidemiology of Major Diseases (Peking University), Ministry of Education, Beijing 100191, China

[*]Correspondence should be addressed to Dr. Bin Wang (Email: binwang@pku.edu.cn), and Dr. Xiqing Li (Email: xli@urban.pku.edu.cn)

ABSTRACT:
Urban populations are exposed to complex chemical mixtures, yet chemical surveillance commonly prioritizes compounds by occurrence or abundance rather than biological relevance. Here we developed a Wastewater-based framework for Prioritizing Neurotoxic Chemical Mixtures (WPN) to integrate wastewater occurrence, estimated human exposure, toxicokinetics, and neurite-outgrowth-inhibition potency across 43 chemicals in 117 urban areas from 18 European countries. Biological translation substantially reordered chemical priorities: wastewater abundance and effect-weighted bioanalytical equivalent concentration (BEQ) rankings shared only two chemicals among their top ten. The biocide 1,2-benzisothiazol-3(2H)-one (BIT) ranked only 18th by wastewater abundance but dominated the predicted neurotoxic mixture burden, accounting for 67.71% of mean BEQ. Removing BIT revealed a broader set of secondary contributors, whereas total modeled BEQ varied comparatively little among cities. These results show that chemical abundance is a poor proxy for biological importance and establish wastewater as a scalable population-level platform (www.exposomex.cn/wpn) for prioritizing chemicals and cities for targeted biomonitoring and effect-based validation.

KEYWORDS: Wastewater, neurotoxic chemicals, urban population surveillance, risk assessment

One-sentence Summary:
Translating urban wastewater signals into exposure- and potency-weighted effects reveals neurotoxic chemical priorities that are largely invisible from chemical abundance alone, and identified 1,2-benzisothiazol-3(2H)-one as a critical toxicant for most European cities.

INTRODUCTION

Neurological conditions are among the largest contributors to global disease burden. In 2021, disorders affecting the nervous system were the leading cause of disability-adjusted life years worldwide and affected more than 3.4 billion people [1]. Although this burden arises from multiple risk factors, exposure to neurotoxic chemicals is a preventable contributor, particularly during prenatal and early-life development, when disruption of neuronal differentiation, neurite formation and circuit maturation can produce lasting effects on brain function [2]. Importantly, people are not exposed to chemicals one by one at a time. Instead, they are continuously exposed to complex mixtures originating from consumer products, food, pharmaceuticals, pesticides, industrial activities and the broader environment [3,4]. Assessing chemicals individually can overlook the combined effects of these exposures. Backhaus and colleagues propose a mixture allocation factor (MAF) for the revision of REACH, the European Union's chemicals regulation, to account for combined exposures [5].

Human biomonitoring (HBM) assesses exposure by measuring chemicals or their metabolites in biospecimens such as urine, blood and breast milk. HBM has become a cornerstone of population exposure surveillance and is implemented through national and international programmes. In Europe, the HBM4EU initiative brings together population studies to harmonize measurements of internal chemical exposure [6]. For example, broad chemical screening of maternal plasma in Germany, combined with experiments on reconstructed mixtures, showed that numerous detected chemicals can contribute additively to neurite outgrowth inhibition even when individual compounds are present below their respective effect thresholds [7]. However, HBM studies generally target only a selected fraction of the chemical universe. Furthermore, participant recruitment, specimen collection, analytical cost, and ethical requirements limit the number of chemicals,

locations and time points that can be covered [8]. These constraints are especially consequential for urban exposome surveillance, which encompasses chemical exposures alongside the environmental and social conditions of city life [9].

Wastewater-based epidemiology (WBE) offers a complementary population-scale lens. Municipal wastewater integrates chemical and metabolic signals from large contributing populations and can be sampled anonymously and repeatedly without recruiting individuals [10,11]. WBE has progressed from estimating illicit-drug and pharmaceutical consumption to monitoring pesticides, bisphenols, plasticizers, flame retardants, household and personal-care chemicals, and lifestyle-related biomarkers [12-19]. It is worth noting, however, that wastewater signals include both human excretion and chemicals entering sewers through product use, disposal or other releases. Thus, wastewater occurrence is informative about a city's chemical environment but does not represent a direct measurement of what its residents absorb.

For urban wastewater observations to guide priorities for neurotoxicity-related investigation, two linked challenges must be addressed. The first one is to derive comparable exposure-related information for screening from heterogeneous wastewater observations that differ across chemicals and cities. Conventional wastewater back-calculation is most applicable for human-specific biomarkers with characterized excretion factors. For many other chemicals, sampling bias, in-sewer transformation, and non-human inputs complicate the connection between measured concentrations and population exposure [10,11,20]. This difficulty is compounded by differences in which substances are measured. Monitoring archives, such as the NORMAN network database (https://www.norman-network.com/), document together valuable observations from diverse campaigns, but the city-by-chemical matrix is sparse [21]. Incomplete records in these databases reduce chemical and spatial coverage. Recent advances in learning from small tabular datasets offer tools for reconstructing missing information in heterogeneous datasets [22], which can support data reconstruction in small, heterogeneous datasets. Noted that, their role here is not to treat predictions as measurements, but to support a common comparison in which observations remain unchanged and reconstructed values remain identifiable. However, whether such reconstruction supports comparisons among cities must be assessed using comparable spatial units, tests against withheld observations and sensitivity to the amount of available information.

The second challenge is to determine which chemicals warrant priority for follow-up by combining exposure-related information with their relative potency at a defined biological endpoint. Exposure magnitude alone cannot define priority because chemical-specific toxicokinetics and biological potency determine how strongly a compound contributes to a shared endpoint. Addressing this challenge requires sufficiently broad toxicokinetic and endpoint-specific effect data to compare chemically diverse substances on a common basis [23,24]. At mixture scale, the number and composition of co-occurring chemicals vary across locations and generate far more combinations than can be tested experimentally, making integrated occurrence and effect data essential for scalable prioritization [25,26]. Such prioritization must distinguish chemical abundance from biological importance. For example, a mass-dominant compound may have limited activity at the selected endpoint, whereas a less abundant but more potent or persistent compound may contribute more strongly [23,24]. This mismatch may be addressed by effect-based methods that scale chemical levels to a reference compound and combine components that act on a shared endpoint as a bioanalytical equivalent concentration (BEQ) [27-29]. For neurite outgrowth inhibition, concentration addition has been supported experimentally in human-relevant reconstructed mixtures [7], and related bioassays have been applied to surface water and wastewater extracts [30-32]. However, exposure, toxicokinetic and effect data remain uneven across the chemical universe. The

scaling framework therefore must integrate the available high-throughput evidence into an endpoint-specific, screening-level risk ranking, rather than interpret any intermediate model output as an absolute measure of human health risk.

Here, we developed a Wastewater-based Framework for Prioritizing Neurotoxic Chemical Mixtures (WPN) to address these two challenges together (Fig. 1). We selected 43 exogenous chemicals across eight predefined classes by combining the observed wastewater records meeting the urban coverage criterion with published relative effect potencies for neurite-outgrowth inhibition, and then characterized their recorded disease associations using E-BIO (Fig. 1a). Georeferenced observations from the NORMAN monitoring network were then harmonized using a consistent definition of urban areas, converting heterogeneous site-level records and repeated measurements into comparable profiles for 117 urban areas across 18 European countries (Fig. 1b) [33]. Within WPN, wastewater concentrations were sequentially translated into estimated daily exposure (DE) and model-derived internal-equivalent concentrations, and subsequently weighted by chemical-specific relative effect potency (REP) for neurite outgrowth inhibition to derive narciclasine-equivalent BEQ (Fig. 1c). This framework therefore separates three quantities that are frequently treated as interchangeable: environmental abundance, internal exposure and biological effect contribution. By comparing chemical rankings before and after biological translation, WPN identifies chemicals whose neurotoxic relevance is disproportionately high or low relative to their wastewater abundance and maps the resulting effect-weighted mixture burden across European urban populations (Fig. 1d). This framework provides a scalable basis for identifying both priority chemicals and urban areas for subsequent chemical measurements, human biomonitoring and paired chemical–bioassay assessment.

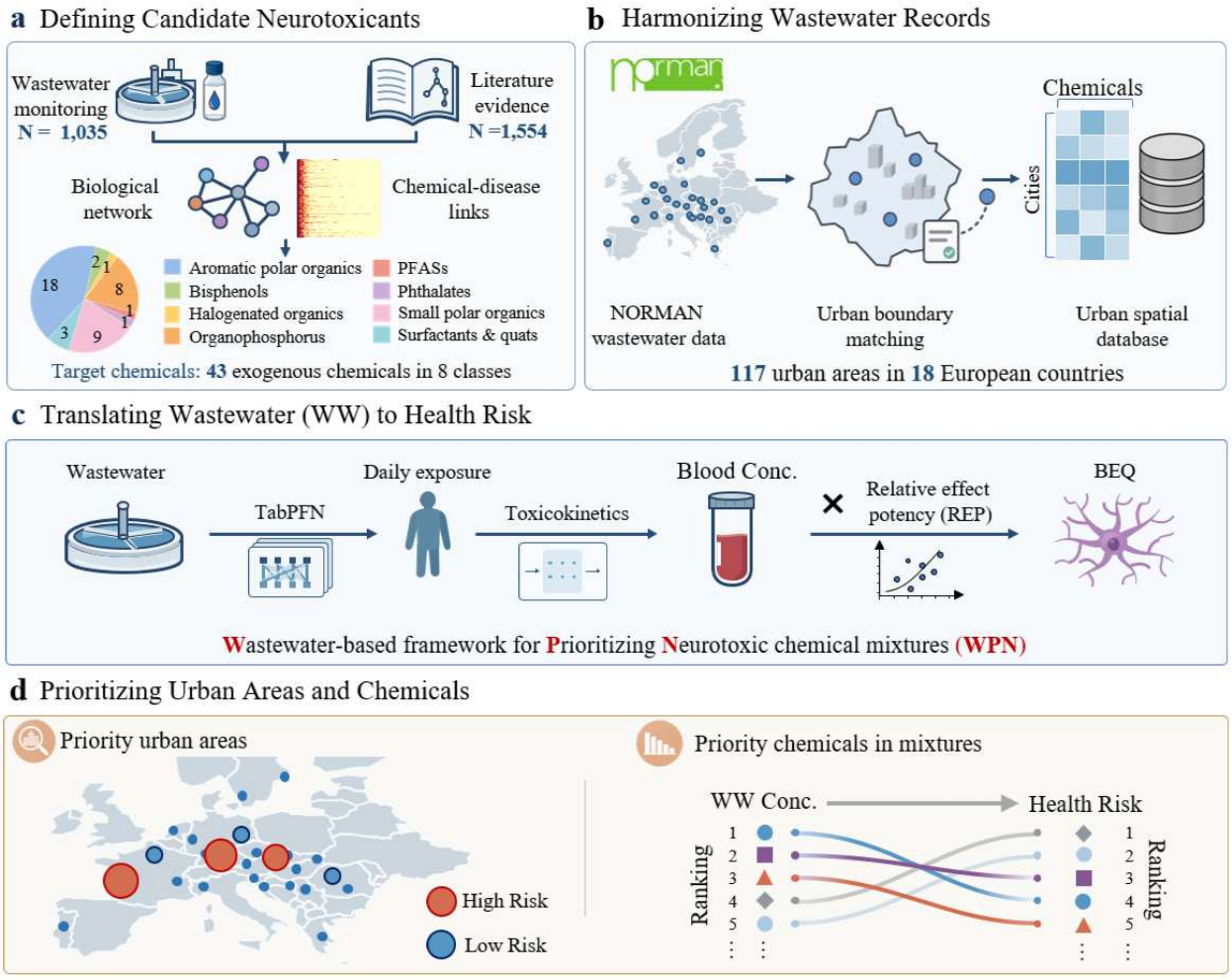


Fig. 1 | Wastewater-based framework for prioritizing neurotoxic chemical mixtures across European cities.

a, Candidate chemicals were identified by intersecting three complementary evidence streams: the wastewater chemical inventory, literature-derived candidates and positive numerical REP values for neurite-outgrowth inhibition. **b**, Georeferenced wastewater-monitoring records from the NORMAN network were spatially harmonized across European cities. **c**, The WPN translated wastewater signals (WW) into estimated daily exposure (DE), model-derived internal-equivalent exposure (Cb), and subsequently effect-weighted mixture burden using chemical-specific relative effect potency (REP). **d**, Biological translation can reorganize chemical priorities, distinguishing chemicals that dominate by environmental abundance from those that dominate by neurotoxic potency.

RESULTS

Targeting the chemical mixture for city-level neurotoxic burden assessment

We selected a focused screening panel by combining wastewater observability with published potency data for neurite-outgrowth inhibition. Chemical identities were harmonized across the NORMAN wastewater archive (1,035 compounds; Table S1) and the literature-derived inventory of 1,554 candidate chemicals (Table S2). Among chemicals meeting the urban coverage criterion, 53 matched the reference set of 97 chemicals with strong evidence on neurotoxicity and positive numerical relative effect potencies (REPs) for neurite-outgrowth inhibition. Excluding creatinine yielded an initial panel of 52 chemicals. Further screening retained 43 target chemicals (NT43) with the volume of distribution (Vd) and elimination half-life ($t_{0.5}$) required for downstream analysis, spanning eight classes based on structural and physicochemical characteristics (Extended Data Fig. 1a). These comprised aromatic polar organics (n = 22), bisphenols (n = 2), halogenated persistent organics (n = 1), organophosphorus compounds (n = 6), PFASs (n = 1), phthalates (n = 1), small polar organics (non-halogenated) (n = 7), surfactants & quats (n = 3) (Extended Data Fig. 1b).

We then used ExposomeX E-BIO module (www.exposomex.cn/ebio) and Exposure-Disease Graph Neural Network (ExDGNN) model to characterize the disease associations and biological coverage of the final panel. The ExposomeX E-BIO knowledge system links chemicals to diseases through proteins (exposure-protein-protein-disease, EPPD) or Gene Ontology terms (exposure-GO-disease, EGoD), placing the selected panel in a broader biological context (Fig. 2; Extended Data Fig. 2 and Table S3). NT43 reached 246 of the 261 diseases returned by the full EPPD network (94.25%) and 198 of the 278 returned by EGoD (71.22%; Table S4). Thus, the panel retained broad recorded disease associations despite comprising only a subset of the candidate chemicals. The corresponding NT43 association networks are shown in Fig. 2c-d. In a separate analysis of mechanism overlap, random sets of 43 chemicals drawn from each disease's own associated chemical pool covered, on average, 69.44% of its proteins and 72.77% of its GO terms across the 89 EPPD and 77 EGoD diseases with at least 43 associated chemicals, respectively (Extended Data Fig. 2c-d and Table S4).

ExDGNN further characterized NT43 across 344 graph-connected neurological disease endpoints (Supplementary Section 1; Figs. S1-S2), generating 14,792 chemical-disease association scores. The ExDGNN knowledge graph contained 341,304 nodes across five entity types (Table S5) and 3,642,481 associations across nine relation types (Table S6). Fig. 2e presents the 1,849 scores for NT43 and the 43 displayed disease endpoints with a score of at least 0.3 for one or more target chemicals (Table S7). Across five held-out folds, ExDGNN achieved a test macro-AUROC of 0.816 ± 0.004 and macro-AUPRC of 0.804 ± 0.004 (Fig. S3a-b). The predicted profiles showed shared high-scoring associations alongside differences among chemicals. In sum, the E-BIO links and predicted ExDGNN profiles support the biological relevance of NT43 for neurological disease-oriented screening.

Concentration levels of the target wastewater neurotoxicant chemicals in Europe cities

Monitoring coverage was highly uneven across both chemicals and urban areas. Missing values were reconstructed using a foundation AI model (TabPFN) coupled with chemical properties and correlations between wastewater concentrations (Fig. S4a). The uneven coverage reflected differences in sampling intensity and analytical scope among NORMAN campaigns. To retain chemical co-occurrence structure without discarding sparsely monitored substances or cities, we reconstructed the broader 117-urban-area × 663-chemical wastewater matrix containing 12,746 observed cells and 64,825 missing combinations (83.57% of the matrix) before extracting the

316 predefined 43-chemical neurotoxicant panel. All measured values were retained unchanged, and
317 only missing cells were estimated.
318 In ten repeated masking experiments, 12,750 held-out predictions covering 8,339 distinct
319 observed cells yielded R^2 = 0.785, RMSE = 0.454 and MAE = 0.319 on the $\log_{10}$ concentration
320 scale (Fig. S4b). Additional tests examined newly available observations and agreement between
321 common concentration profiles (Supplementary Section 2; Extended Data Fig. 3a-e and Table S8).
322 Although prediction accuracy declined as observations were removed, reconstruction retained a
323 mean R^2 of 0.595 even after removing 75% of the observed values (Supplementary Section 2;
324 Extended Data Fig. 3f). The completed target panel comprised 5,031 city-chemical values.
325 Reconstruction substantially increased spatial comparability of the 43 target chemicals across the
326 117 European urban areas (Fig. S4c), providing a harmonized concentration basis for subsequent
327 exposure translation and mixture-effect analyses.

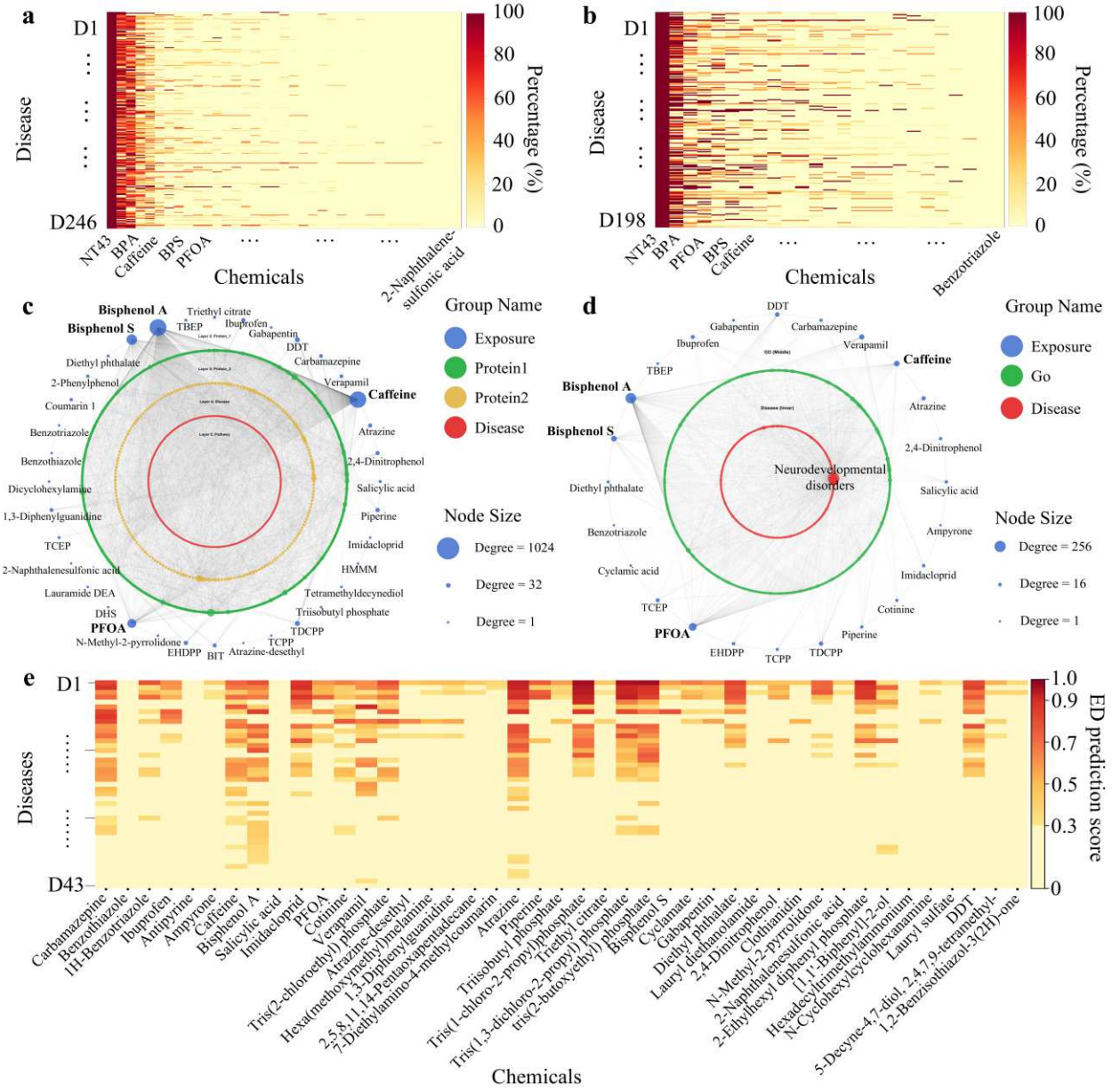


Fig. 2 | Forty-three target wastewater neurotoxic chemicals (NT43) show distinct disease-relevant protein and biological-process profiles in the exposure-biology-disease network and the predicted disease associations.

a, Proportional contribution of each target wastewater chemical to disease-associated proteins in the exposure-protein-protein-disease (EPPD) network, expressed as the percentage of proteins linked to each chemical among all disease-associated proteins. NT43 denotes the complete set of 43 target wastewater neurotoxic chemicals. **b**, Proportional contribution of each chemical to disease-associated biological processes in the exposure-Gene Ontology (GO)-disease (EGoD) network, expressed as the percentage of total disease-associated GO terms. **c**, Layered circular visualization of the EPPD network for NT43. **d**, Layered circular visualization of the EGoD network for NT43. **e**, ExDGNN scores for NT43 across 43 neurological diseases selected by a score of at least 0.3 for one or more target chemicals. Scores range from 0-1; values < 0.3 share the same pale color. The detailed disease IDs' information of various were provided in Table S4.

WPN translates wastewater signals into effect-weighted priorities

WPN translated wastewater concentrations first into estimated daily exposure (DE) and then into an effect-weighted neurotoxicity metric (Fig. 3a). The WW-to-DE model retained a clear wastewater concentration signal while incorporating chemical-specific properties that modify the relation between environmental occurrence and population exposure. In chemical-level tenfold cross-validation, pooled out-of-fold predictions yielded $R^2 = 0.543$, RMSE = 1.001 and MAE = 0.656 on the $\log_{10}$ scale. Wastewater concentration accounted for 22.5% of normalized permutation importance, following logKoa and logKoo (Fig. S5a-b). Sensitivity analysis further showed that increasing wastewater concentration produced an increasing but subproportional DE response, with distinguishable predictions under the 0.1×, 1× and 10× concentration scenarios (Fig. S5c).

The WW→DE step retained some abundance information but redistributed it across the mixture. Across these profiles, concentrations ranged from 0.00220 to 1,181 nM, and the sum of the 43 concentrations varied 36.42-fold among cities. No chemical approached complete dominance of the mean wastewater profile: caffeine, 1H-benzotriazole and hexa(methoxymethyl)melamine (HMMM) contributed 14.77%, 13.85% and 11.66%, respectively (Table S9). The total DE varied 1.67-fold among cities. The largest mean DE contributions were benzothiazole (32.91%), N-methyl-2-pyrrolidone (NMP, 11.18%) and lauryl sulfate (6.90%). Across chemicals, ranks based on the across-urban-area medians of WW and DE remained positively associated ($R = 0.568$, $P < 0.001$; Table S10).

The subsequent DE→BEQ step further reordered chemical priorities (Fig. 3b and Table S9). Chemical-specific DE was scaled using common body-mass and single-compartment toxicokinetic assumptions and then multiplied by the REP for neurite-outgrowth inhibition to obtain narciclasine-equivalent BEQ. Across-urban-area median DE and BEQ ranks were weakly associated ($R = 0.237$, $P = 0.126$), as were WW and BEQ ranks ($R = 0.211$, $P = 0.174$). 1,2-benzisothiazol-3(2H)-one (BIT) shifted from rank 18 by mean WW and rank 4 by mean DE to rank 1 by mean BEQ, accounting for 67.71% of the mean effect-weighted profile.

Rank-stratum transitions summarized the full WW-DE-BEQ sequence (Fig. 3c). Chemicals were ranked at each stage using their across-urban-area median values and divided into high (ranks 1-15), medium (16-29) and low (30-43) strata. 21 chemicals (48.8%) changed stratum from WW to DE, and 25 (58.1%) changed from DE to BEQ. 27 chemicals (62.8%) ended in a different BEQ stratum from their initial WW stratum, while only 10 (23.3%) remained in the same stratum throughout. Both translations altered which chemicals occupied the highest-priority group. Thus, WPN did not simply rescale wastewater abundance. Exposure translation and effect weighting supplied distinct information that changed the resulting screening priorities.

To place these chemical priorities in a broader population context, we compared national median wastewater concentrations of the 43 target chemicals with food-supply patterns across 18 European countries (Fig. S6a; Supplementary Section 3). We identified 191 strong positive associations ($R \geq 0.5$), 3.46 times the mean number under country-label permutation ($P = 0.0001$; Fig. S7). Six individual associations (five positive and one negative) survived global false-discovery-rate correction (Table S11). Food-module summaries showed different patterns of positive correlations among the ten chemicals with the highest across-city arithmetic mean BEQ (Fig. S6b). These national-level associations provide contextual leads for follow-up source investigations, although food availability does not measure individual dietary intake.

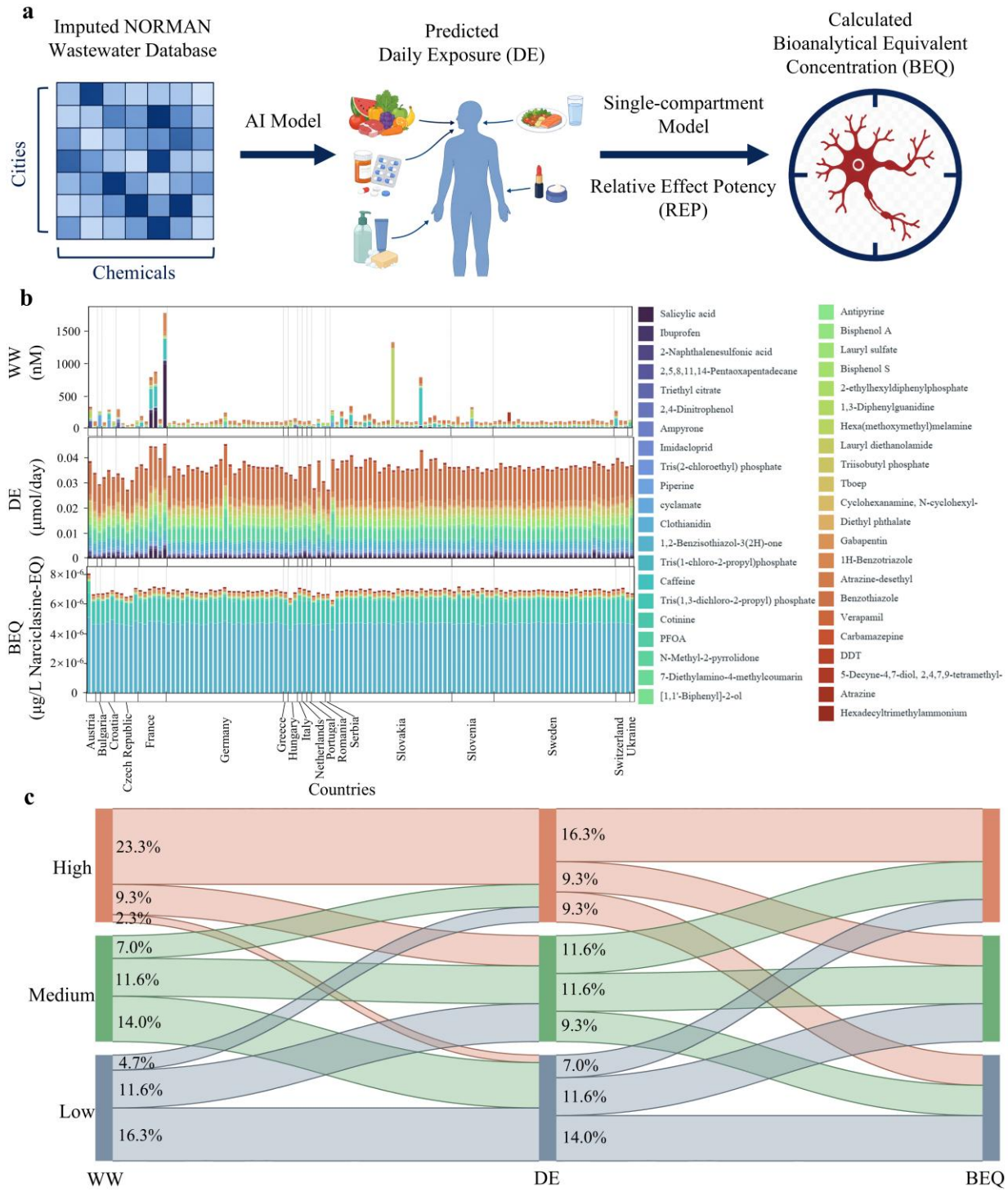


Fig. 3. Wastewater-to-effect translation reorders chemical priorities.
a, WPN translates wastewater concentration (WW) into estimated daily exposure (DE) and then neurite-outgrowth-inhibition bioanalytical equivalent concentration (BEQ). **b**, Chemical-resolved WW, DE and BEQ profiles across 117 urban areas. **c**, Alluvial diagram of stage-specific rank transitions. Chemicals were ranked separately by their across-urban-area median values and classified as high (ranking 1-15), medium (ranking 16-29) or low (ranking 30-43).

Effect Weighting Reveals City-level Neurotoxic Mixture Burdens

Effect weighting produced a highly concentrated chemical profile but comparatively modest geographic variation across the 117 European urban areas (Fig. 4a). Total BEQ ranged from 6.424 × 10^{-6} to 8.252 × 10^{-6} μg L^{-1} narciclasine-EQ, with a mean of 7.075 × 10^{-6} μg L^{-1}, a 1.285-fold maximum-to-minimum range and a coefficient of variation of 2.61%. The highest model-derived totals were found in Linz-Asten, Austria (8.252 × 10^{-6} μg L^{-1}), Ljubljana, Slovenia (7.359 × 10^{-6} μg L^{-1}), and Biganos, France (7.352 × 10^{-6} μg L^{-1}), whereas the lowest were found in Vrakuňa, Slovakia (6.424 × 10^{-6} μg L^{-1}), Győr, Hungary (6.546 × 10^{-6} μg L^{-1}), and Boskovice, Czech Republic (6.645 × 10^{-6} μg L^{-1}) (Fig. 4a).

The narrow BEQ ranges among cities all accompanied a strongly shared dominant contributor. BIT contributed 62.88-70.75% of each city's total and 67.71% of the across-city mean (Fig. 4b and c). The next contributors were tris(1,3-dichloro-2-propyl) phosphate (TDCPP) (22.52%), gabapentin (2.98%), perfluorooctanoic acid (PFOA, 1.94%), hexadecyltrimethylammonium (HDTMA, 1.38%), 1,3-diphenylguanidine (DPG13, 1.12%). BIT and the next nine chemicals together accounted for 98.99% of mean BEQ, leaving approximately 1% distributed among the other 33 chemicals.

Because BIT overwhelmingly dominated the total, we repeated all city- and composition-level summaries after excluding BIT, without refitting any preceding model (Extended Data Fig. 4a-c). The remaining total BEQ ranged from 1.997 × 10^{-6} to 3.063 × 10^{-6} μg L^{-1}, with a mean of 2.284 × 10^{-6} μg L^{-1}, a 1.534-fold range and a coefficient of variation of 5.26%. After BIT exclusion, Linz-Asten, Austria (3.063 × 10^{-6} μg L^{-1}), Krzko, Slovenia (2.640 × 10^{-6} μg L^{-1}), and Ljubljana, Slovenia (2.595 × 10^{-6} μg L^{-1}), had the highest residual totals, whereas Boskovice, Czech Republic (1.997 × 10^{-6} μg L^{-1}), Varaždin, Croatia (2.053 × 10^{-6} μg L^{-1}), and Šabac, Serbia (2.069 × 10^{-6} μg L^{-1}), had the lowest (Extended Data Fig. 4a). TDCPP accounted for 69.74% of residual mean BEQ, followed by gabapentin (9.22%), PFOA (6.00%), HDTMA (4.28%), DPG13 (3.47%) and atrazine-desethyl (DEA, 1.10%).

Chemical-class summaries distinguished aggregate contributions from the average contribution of an individual class member. With BIT included, the largest class-total contributions were aromatic polar organics (70.52%), organophosphorus compounds (22.88%), small polar organics (non-halogenated) (3.08%). After BIT exclusion, the largest class-total contributions were organophosphorus compounds (70.87%), small polar organics (non-halogenated) (9.47%), and aromatic polar organics (8.69%). With BIT included, organophosphorus compounds had the highest arithmetic mean BEQ per chemical (2.698 × 10^{-7} μg L^{-1} narciclasine-EQ). Aromatic polar organics and PFASs followed at 2.268 2.057 × 10^{-7} and 1.371 × 10^{-7} μg L^{-1} narciclasine-EQ, respectively. After BIT exclusion, organophosphorus compounds remained the highest-ranking class by mean BEQ per chemical, followed by PFASs and surfactants & quats (3.724 × 10^{-8} μg L^{-1} narciclasine-EQ; Fig. S8a-b).

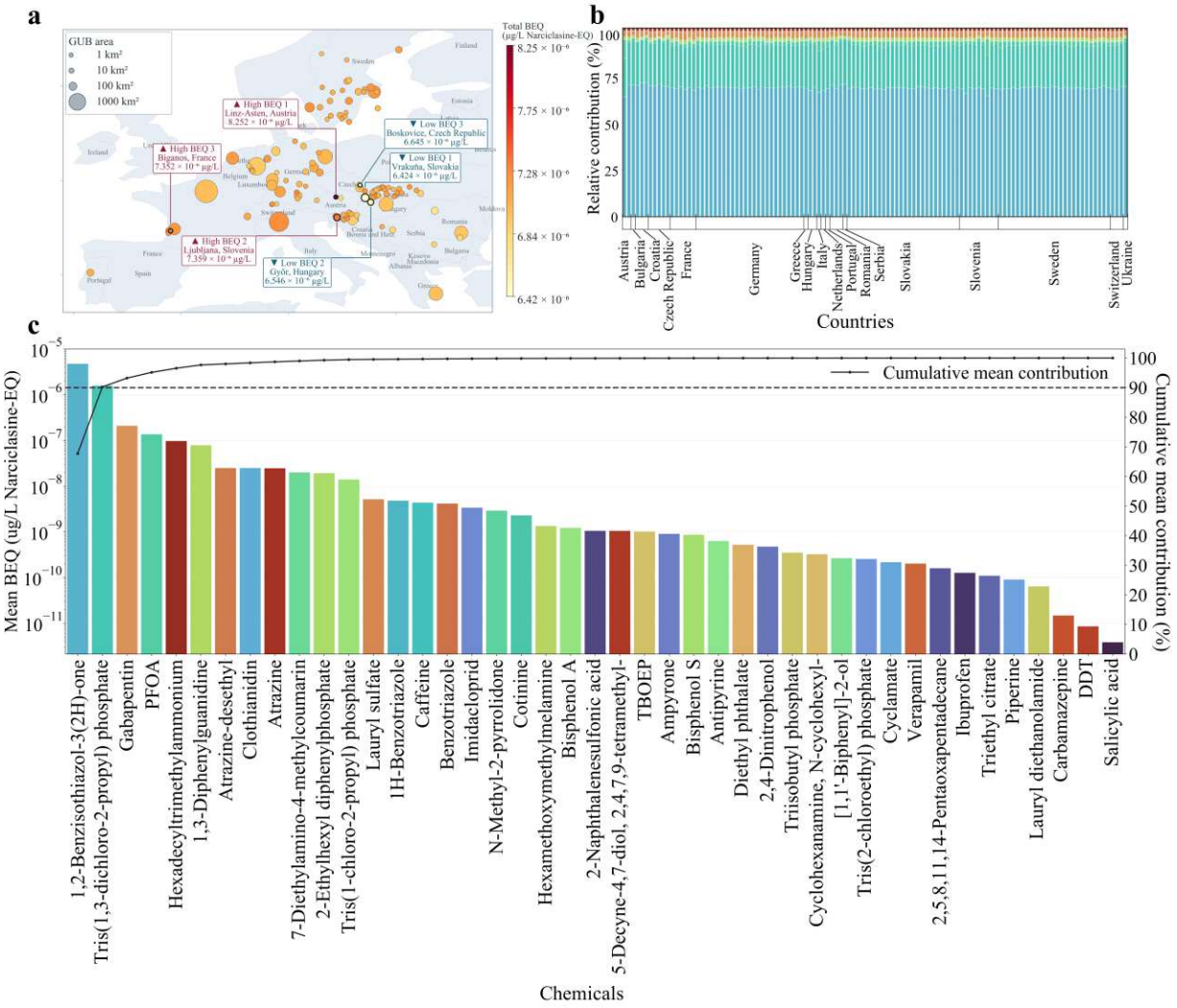


Fig. 4. Effect-weighted neurotoxic mixture burdens across European GUB urban areas. **a**, Total bioanalytical equivalent concentration (BEQ) across 117 Global Urban Boundary (GUB) urban areas. Symbol size represents GUB area and colour represents city-level total BEQ. **b**, Relative chemical contributions to total BEQ within each urban area. **c**, Across-city mean chemical-specific BEQ and cumulative contribution, ordered by mean BEQ.

DISCUSSION

This study establishes WPN as a framework for turning heterogeneous wastewater signals into an endpoint-specific screening priority. It connects broad community-level chemical monitoring with biologically interpretable mixture assessment, two aspects of population chemical surveillance that have usually remained separate. This connection is needed because an inventory of detected chemicals can show what reaches wastewater but cannot determine which components most warrant biological confirmation. Human biomonitoring and effect-based measurements provide stronger biological relevance, but repeated application across many chemicals and urban populations is more difficult [3,4,6,10,11]. WPN addresses this gap through an explicit WW→DE→BEQ sequence. In that sequence, wastewater concentration represents an available population-scale signal, estimated daily exposure provides a chemically informed translation of that signal, and bioanalytical equivalent concentration expresses the included chemicals on a common neurite-outgrowth-inhibition scale. In our analysis, BIT emerged as the leading candidate for neurite-outgrowth inhibition, making an otherwise broad prioritization exercise a focused biological question. The remaining mixture also matters because a dominant contributor can cover up other candidates that may become important in different settings. Therefore, the practical contribution of WPN is to help determine what to measure and test next, linking broad urban monitoring to a manageable set of chemical and mixture hypotheses. This is a shift in the use of wastewater information: from describing the chemicals present towards selecting those whose exposure and biological relevance most warrant confirmation.

A central contribution of WPN is to bring several incomplete evidence streams into one decision process while keeping the scientific meaning of each transition distinct. European cities monitoring complex wastewater mixtures need to decide which chemicals warrant closer investigation and how best to investigate them. The overlap between chemicals detected in influent wastewater and those with published potency data for neurite-outgrowth inhibition gives cities a set of chemical candidates that can be monitored and compared at a common biological endpoint. This focused panel retains links to a broad range of diseases in ExposomeX E-BIO, and ExDGNN predicts both shared and distinct disease associations across the selected chemicals. In the selected panel, WPN uses wastewater concentrations and chemical properties to estimate daily exposure, then combines those estimates with toxicokinetic and potency data to rank chemicals by their predicted contributions to the mixture BEQ. Keeping these quantities distinct allows investigators to follow a chemical from its occurrence in wastewater to its position in the effect ranking and to choose measurements that can test that priority. A high estimated exposure gives cities a reason to look for the chemical or its metabolites in residents using harmonized biomonitoring methods such as those developed through HBM4EU [6]. A large predicted contribution to mixture BEQ gives a reason to pair chemical analysis with bioassays, as illustrated by a German pregnancy cohort in which both maternal plasma extracts and mixtures prepared from the detected chemicals were tested for neurite-outgrowth inhibition [7]. Across cities, a common shortlist could guide repeated influent sampling, comparisons of sample extracts with mixtures prepared at the measured concentrations, and targeted biomonitoring in the populations served [11]. The principal advance is therefore an auditable route from scalable surveillance to a bounded set of chemicals, urban areas and mixture hypotheses for which confirmatory resources can be deployed deliberately.

The second contribution is a practical resolution of the first scientific challenge: obtaining comparable exposure-related information from monitoring records whose chemical and geographical coverage is intrinsically uneven. Large wastewater monitoring programmes differ in the treatment plants they sample, the chemicals they measure and the concentrations they can

detect [21]. A chemical may therefore be missing from a city's records simply because local campaigns did not measure it. Comparing only complete records would limit both the number of cities and the range of chemicals, while reconstructing missing concentrations makes it possible to compare the same panel of chemicals across cities with different levels of monitoring coverage. WPN first links sampling records to cities using common urban boundaries [33] and documented links between treatment plants and cities. For each chemical, it retains the measured city median where available and estimates missing concentrations from patterns among chemicals measured together in wastewater [22]. Once these concentration profiles are comparable across cities, the next question is how a given influent concentration relates to population exposure for each chemical. For some chemicals, intake can be estimated from wastewater when human excretion, stability in sewers, wastewater flow and the size of the contributing population are well characterized [10,11,16,20]. Across a diverse chemical panel, these factors are unevenly documented, and both human excretion and direct discharges can contribute to influent concentrations. The subsequent WW→DE model accounts for differences among chemicals by combining influent concentrations with chemical properties to estimate daily exposure. These exposure estimates provide a common basis for comparing the predicted contributions of different chemicals to neurite-outgrowth inhibition.

The third contribution is the resolution of the second scientific challenge: converting exposure-related information into a biologically informed mixture priority. Chemicals with similar estimated daily exposures can reach different internal concentrations because they differ in how widely they distribute in the body and how long they persist [23,24,28]. At the same internal concentration, a more potent chemical contributes more to the predicted inhibition of neurite outgrowth. Even at the same internal concentration, their contributions to neurite-outgrowth inhibition can differ due to their different inhibition potency [7]. WPN uses each chemical's potency relative to narciclasine to express its estimated internal concentration on a common effect scale, allowing contributions to be summed within each city and compared across chemicals. The resulting reordering can be read as a change in the prioritization question: from which chemical is most abundant to which included chemical contributes most strongly at the selected endpoint. BIT makes this distinction especially clear. In our analysis, BIT accounted for approximately 68% of the across-city mean BEQ despite its lower rank by wastewater concentration. And in a separate study of Elbe River water, BIT accounted, for approximately 62% of the total narciclasine equivalents on average calculated from the chemicals quantified for the same neurite-outgrowth endpoint [34]. BIT has also been detected in maternal plasma [7] and in a semi-quantitative analysis of human urine [35] (see Supplementary Section 4 for further evidence on its occurrence and human exposure). Together, these findings make BIT a priority for human biomonitoring in European cities and for experiments testing its contribution to neurite-outgrowth inhibition in local chemical mixtures.

Several limitations define both the interpretation of WPN and the sequence in which it should be validated. First, the concentrations of the concerned chemicals in WW were affected various factors. Their selection should be further supported in wastewater-based epidemiology with well-characterized biomarkers for real-life health risk assessment[8,20,36]. Second, both reconstruction and WW→DE translation are evaluated against available data, and may perform less reliably for structured gap. Poorly represented chemical domains or city-specific source events than for observations resembling the training distribution. Third, the BEQ is restricted to chemicals with an available relative effect potency for neurite-outgrowth inhibition. It expresses their predicted combined activity at this endpoint in narciclasine equivalents. Other aspects of neurotoxicity,

including effects on synapse formation and neuronal network function, require complementary assays.

To improve reproducibility and transparent interpretation, we developed an interactive WPN web platform that provides downloadable data, visualization tools and WPN model analysis code (see: https://www.exposomex.cn/wpn). The brief interfaces were presented in Extended Data Fig. 5. The interactive WPN platform distinguishes measured from reconstructed wastewater concentrations and allows users to compare abundance and BEQ rankings and download the corresponding data and figures. Users can explore measured and reconstructed influent concentrations by city and chemical in WASTEWATER module, examine transformed BEQs and download model code in WPN module, and compare rankings by wastewater concentrations and BEQs in RANK module. For individual cities and for the mean profile across cities, RANK module also shows each chemical's contribution to total BEQ. The downloadable code extends the analysis to users' own samples, allowing them to enter influent concentrations for one or more of the NT43, estimate their individual BEQs and rank the selected chemicals by BEQ.

Overall, WPN advances wastewater surveillance from occurrence mapping towards effect-organized, screening-level risk prioritization. Linking wastewater occurrence to exposure and biological potency helps identify chemicals that abundance-based monitoring would give lower priority. A shared shortlist gives European cities a common starting point for coordinated monitoring while allowing each city to include chemicals of local concern. Follow-up studies should pair repeated influent sampling with human biomonitoring to identify wastewater markers that reliably track population exposure. Measurements from additional cities can test concentration reconstruction, while independent paired wastewater and exposure data can test WW-to-DE translation across a wider range of chemicals. Testing the same mixtures and their individual constituents for neurite outgrowth, synapse formation and neuronal network development would show which chemicals remain priorities across endpoints. Its intended advance is to make broad chemical surveillance more useful for choosing what to verify next.

MATERIALS AND METHOD

1. Identifying Target Neurotoxicant Chemicals

Target chemicals were selected using a stepwise evidence-integration procedure designed to identify exogenous compounds that were both observable in European wastewater and relevant to neurotoxicity. We first assembled the chemicals monitored in the NORMAN wastewater archive, comprising 1,035 compounds with georeferenced concentration records that could be assigned to the European urban areas included in this study. The NORMAN network brings together reference laboratories and research organizations working on emerging environmental substances. Its EMPODAT Chemical Occurrence Database was established in 2005 and compiles georeferenced monitoring records from research projects and national programmes. Records are harmonized in a common format across Europe and beyond [21]. Among them, 663 chemicals were measured in at least three cities. To establish broader evidence of wastewater occurrence and human relevance, we harmonized these compounds against an independently compiled wastewater chemical inventory and against the 1199 chemicals evaluated in a recent human-relevant neurotoxicity study [7]. Among them, 97 chemicals have a positive numerical REP for this endpoint. Chemical identities across sources were standardized using canonical names, synonyms, CAS Registry Numbers where available, and structural identifiers. InChIKeys and SMILES were used as the primary structure-based identifiers to resolve synonyms and avoid duplicate counting of the same compound. Overlap analysis was conducted after chemical-name normalization and structure matching (Extended Data Fig. 1a). Matching the 663 chemicals meeting the urban coverage

criterion to the 97 chemicals with positive numerical REPs yielded 53 compounds. Exclusion of the endogenous marker creatinine left 52 target chemicals. The final analysis retained 43 chemicals with available volume of distribution (*Vd*) and elimination half-life ($t_{0.5}$) values from HExpPredict (Extended Data Fig. 1a; Tables S9-S10) [37]. The final panel comprised 22 aromatic polar organics, 2 bisphenols, 1 halogenated persistent organic, 6 organophosphorus compounds, 1 per- and polyfluoroalkyl substance, 1 phthalate, 7 non-halogenated small polar organics, and 3 surfactants and quaternary ammonium compounds. These compounds therefore had simultaneous evidence of wastewater observability, occurrence within the European monitoring archive, and prior inclusion in a human-relevant neurotoxicity assessment.

2. Construction of "Exposure-Biology-Disease" Knowledge Graphs

We evaluated biological support for neurotoxicity using the ExposomeX E-BIO knowledge system (https://www.exposomex.cn/ebio) [38]. This framework provides a standardized, database-driven approach to link environmental exposures with disease outcomes through curated biological intermediates. The detailed procedures were described previously [39]. For each candidate compound, structured chemical-protein, chemical-phenotype, protein-protein, and phenotype/protein-disease relationships were queried and mapped to neurotoxicity-relevant biological processes. Compounds without identifiable neurotoxicity-related biological evidence were excluded from the primary analytical panel. The full candidate input and identity-matching results are listed in Table S3. Briefly, two complementary graph modes were employed in this study, including Exposure-GO-Disease (EGoD) and Exposure-Protein-Protein-Disease (EPPD) (Fig. 2; Extended Data Fig. 2).

In this network, neurodevelopmental disorders were selected as the target health outcomes based on a structured, multistep strategy integrating biological relevance, epidemiological evidence, and network coverage. First, diseases were screened from standardized disease ontologies (e.g., OMIM: https://www.omim.org, and MESH: https://www.ncbi.nlm.nih.gov/mesh) by retaining terms explicitly related to impaired neurodevelopmental processes, including cognitive, behavioral, and neuropsychiatric phenotypes with established developmental origins. Second, relevance to chemicals was evaluated by mapping candidate diseases to exposure-biological associations curated from toxicological and biomedical databases, ensuring that each selected disease had sufficient links to genes or proteins known to be perturbed by chemicals. Third, diseases were required to exhibit adequate connectivity within the constructed multilayer networks (EGoD and EPPD), avoiding sparse or isolated nodes that could lead to unstable estimates. This approach ensured that disease selection was hypothesis-driven rather than outcome-driven, biologically grounded rather than exposure-only, and methodologically consistent across frameworks.

Mechanism overlap was summarized from the existing disease-specific random-subset analysis without drawing new samples. For each disease, the sampling pool contained only chemicals with a complete path to that disease. The existing results for subsets of 43 chemicals were retrieved, with coverage defined as the deduplicated union of proteins or GO terms reached by a subset divided by the corresponding union for all chemicals associated with that disease. The original analysis evaluated up to 100 distinct subsets per disease and subset size, enumerating all combinations when fewer were available; the random seed was 42 plus the subset size. Mean coverage at 43 chemicals was first calculated within each disease and then averaged with equal weight across all diseases with at least 43 associated chemicals. Diseases with smaller pools were retained in Table S4 with a reason for non-calculation and were not assigned zero or complete coverage. Extended Data Fig. 2c-d shows the 30 diseases with the largest node unions in each

network; these display subsets were not used to define the full eligible-disease averages. Table S4 reports both sets of summaries and the fixed NT43 results separately.

We further adapted the TxGNN framework to predict exposure-disease associations using an ExposomeX-derived heterogeneous knowledge graph (Fig. S1; Tables S5-S6). The background graph was supplemented with EDB-retrieved mechanistic relationships associated with the wastewater chemical inventory and neurological disease query set. ExDGNN combined relation-specific graph representations with pretrained Uni-Mol molecular features and generated association scores using a DistMult decoder (Fig. S2). Graph construction, entity eligibility, model training, evaluation and query selection of ExDGNN are detailed in Supplementary Section 1, with model performance in Fig. S3 and the NT43-disease scores in Table S7. The graph data, molecular features, trained models and analysis code are provided in "Supplementary Code_ExDGNN Chemical-Disease Association Prediction".

3. Reconstruction of Urban Wastewater Chemical Concentrations

Georeferenced wastewater records were assigned to morphological urban areas delineated by the Global Urban Boundary (GUB) Global 2025 dataset [33] using strict point-in-polygon matching, supplemented by documented station-to-city assignment, and records assigned to the same GUB were aggregated as chemical-specific median concentrations. Coordinate validation, spatial inclusion criteria, overlapping-boundary resolution and within-GUB aggregation are described in Supplementary Section 5. NORMAN monitoring records formed a highly sparse urban-area-by-chemical matrix because individual campaigns differed in sampling frequency, analytical coverage, and target chemical panels. We reconstructed this matrix before extracting the final neurotoxicant panel so that city and chemical comparisons were not restricted to repeatedly monitored combinations (Fig. S4a). Reconstruction eligibility required each chemical to occur in at least three urban areas, each urban area to contain at least three retained chemicals, and each chemical to have at least one finite off-diagonal reference Spearman correlation with another retained chemical based on at least five shared original sampling events (Table S12-S13). The input comprised median concentrations for 663 chemicals measured in at least three of 117 GUB urban areas, corresponding to 12,746 observed cells and 64,825 missing cells (83.57% missing). For chemical i in urban area j, let C_{ij} denote its concentration and MW_i its average molecular weight. Positive concentrations reported in μg L^{-1} were converted to nM and then $\log_{10}$-transformed as

$$C_{ij}^{\mathrm{nM}} = \frac{1000 C_{ij}^{\mu\mathrm{g\ L^{-1}}}}{\mathrm{MW_i}}, Y_{ij} = \log_{10}\left(C_{ij}^{\mathrm{nM}}\right)$$

Values that were missing, nonnumeric or nonpositive were treated as unavailable rather than assigned an arbitrary concentration. All model fitting and residual calculations were performed on Y_{ij}, whereas the original concentration scale was retained for reporting and downstream exposure translation.

Each target chemical was reconstructed separately from information available for the other chemicals. For each fit, the 32 non-target chemicals with the greatest current coverage were used as base predictors. Missing base-predictor values were replaced with medians calculated only from the training rows, and a matched binary missingness indicator was added for every base predictor; the number of observed base predictors in each urban area was also included. The reference co-occurrence information was rebuilt from 40,657 positive original NORMAN records representing 1,580 identifiable sampling events and 1,021 chemicals, without restricting the reference to the target panel or urban cohort. Replicate records for the same event and chemical were aggregated by their median. Pairwise Spearman correlations required at least five shared original events and

nonconstant values for both chemicals; unsupported pairs remained missing in the exported reference matrix (Table S14). For target chemical *i*, positive correlations were ranked and the top 5, 10 and 20 neighbours were summarized separately. If ρ_{ik} is the positive Spearman correlation between target *i* and neighbour *k*, and $\tilde{Y}_{kj}$ is the observed or training-median-filled log10 concentration of neighbour *k* in urban area *j*, the correlation-neighbour feature was

$$R_{\mathrm{ij}}^{(\mathrm{K})} = \frac{\sum_{k \in N_i^{(K)}} \rho_{ik} \times \widetilde{Y_{kJ}}}{\sum_{k \in N_i^{(K)}} \rho_{ik}}, K \in \{5,10,20\}$$

Correlation coefficients were also used as chemical fingerprints to represent broader similarity among compounds. The NORMAN correlation profiles were standardized across chemicals and reduced by principal-component analysis to a 16-dimensional embedding (Table S15). For each target, Euclidean distances were calculated to all other chemicals represented in this space, and the ten closest chemicals were retained. Distance d_{ik} was converted to similarity S_{ik}, and similarity-weighted concentration summaries were calculated for the closest 5 and 10 embedding neighbours as

$$S_{ik} = \frac{1}{1 + d_{ik}}, E_{ij}^{(K)} = \frac{\sum_{k \in M_i^{(K)}} S_{ik} \times \widetilde{Y_{kJ}}}{\sum_{k \in M_i^{(K)}} S_{ik}}, K \in \{5,10\}$$

For every embedding neighbour, its training-median-filled concentration, an indicator of whether that concentration was observed, and its similarity to the target were also supplied individually to the model. The embedding features therefore allowed chemically similar compounds to inform reconstruction even when they were not among the highest-coverage predictors, while the availability indicators preserved the distinction between measured and filled predictor values.

The resulting features were supplied to a pretrained TabPFN v2.6 regressor [22]. Chemicals represented in the NORMAN correlation matrix were processed in the reference order, followed by unmatched chemicals ordered by decreasing observed coverage. After each target model was fitted, predictions were written only to positions that were missing in the original matrix. All originally observed Y_{ij} values were then restored before the next chemical was processed. Earlier reconstructions could consequently contribute predictor information to later targets, but no measured concentration could be overwritten by the sequential procedure. This produced a deterministic, observation-anchored complete matrix on the $\log_{10}$ nM scale.

Reconstruction performance was assessed by ten repeated masking experiments. In each repeat, 10% of eligible observed urban-area-by-chemical cells were randomly withheld, while retaining at least one observed chemical per urban area and at least two observed urban areas per chemical. The entire sequential reconstruction was rerun using the remaining observations, and the withheld values were compared with their predictions using Pearson correlation, coefficient of determination (R^2), root-mean-square error (RMSE) and mean absolute (MAE) error on the $\log_{10}$ scale (Fig. S4b). Each repeat contained 1275 masked cells, giving 12,750 held-out predictions in total. The distinct data-supplementation and observation-thinning assessments are described in Supplementary Section 2 and Extended Data Fig. 3. For each chemical, the validation residual was defined as

$$\varepsilon_{ij} = Y_{ij} - \hat{Y}_{ij}$$

Because deterministic reconstruction can compress between-city dispersion, chemical-specific residual pools were used to introduce bounded variability into the missing cells. The setting minimizing imputed values greater than 1.2 times the corresponding chemical-specific observed

maximum was selected, with ties resolved in favour of the larger scaling factor and then the narrower residual interval. The selected configuration retained residuals between the 25th and 75th percentiles and sampled them with replacement using random seed 42. Chemicals with fewer than five usable residuals received no perturbation. For an originally missing cell, the final log10 reconstruction was

$$\hat{Y}_{ij}^{*} = \hat{Y}_{ij} + \alpha\varepsilon_{ij}^{*}, \alpha = 0.2$$

where ε^{*}_{ij} is a sampled residual from the corresponding chemical-specific pool and α=0.2. Residual perturbations were applied only to originally missing cells; measured cells were reset to their observed values. Final concentrations were returned to nM as

$$\hat{C}_{ij}^{\mathrm{nM}} = 10^{\hat{Y}_{ij}^{*}}$$

The completed matrix therefore contained the original measurements wherever available and model-derived concentrations only for previously missing combinations. These reconstructions were used to establish a comparable analytical surface for chemical and urban-area ranking (Fig. S4c).

4. Building the Relationship between Wastewater Chemical Concentrations and Neurotoxic Mixture Burden

WPN translated wastewater chemical concentrations into effect-weighted mixture estimates in two sequential steps: wastewater concentration to estimated daily exposure (WW→DE), followed by toxicokinetic and neurotoxicity-potency translation (DE→BEQ). For the WW→DE model, we assembled 11,217 region-chemical records covering 1,116 substances across 90 countries or regions from the external training dataset (Table S16). Regional observations were first summarized to a chemical-level median to minimize disproportionate influence from intensively sampled regions. Chemicals reported in only one region were excluded, leaving 472 chemicals with multi-region information for model development (Table S17). Each chemical therefore constituted one independent modelling unit.

The five predictors were $\log_{10}$-transformed wastewater concentration in μM; topological polar surface area (TPSA), the $\log_{10}$ octanol-air octanol partition coefficients (logKoa), the $\log_{10}$ wet-dry octanol partition coefficients (logKoo) and the ADMET half-life descriptor ($t_{0.5}$). Wastewater concentrations were converted from nM to μM before $\log_{10}$-transformation. Missing physicochemical descriptors were imputed using medians calculated from the training data, and the four descriptor variables were standardized using training-set means and standard deviations. The response variable was $\log_{10}$-transformed daily exposure [DE, μmol day^{-1}]. DE was derived from the SEEM3 Total_median exposure estimate expressed in mg kg^{-1} day^{-1} [40], assuming a standardized body mass of 70 kg and converting mass-based exposure to molar amount using each chemical's mean molecular mass. Predictions were back-transformed to obtain DE in μmol day^{-1}.

A TabPFNRegressor was fitted to the chemical-level dataset [22]. Predictive performance was assessed using randomly shuffled tenfold cross-validation, with each chemical assigned to a single validation fold and no chemical appearing simultaneously in training and validation sets. Pooled out-of-fold performance was evaluated using R^2, root-mean-square error (RMSE), and mean absolute error (MAE) on the $\log_{10}$ scale. Predictor contributions were evaluated using permutation importance, defined as the increase in prediction error after independently permuting each predictor and normalized across the five predictors. To test whether the model preserved a physically interpretable wastewater-exposure relationship, we performed a one-at-a-time sensitivity analysis. For each of 43 chemicals, the median wastewater concentration across the 117 cities was multiplied by factors ranging from 0.1 to 10 while all chemical descriptors were held

fixed. Predicted DE under each perturbation was divided by the baseline prediction, and the median and interquartile range of these ratios were summarized across chemicals (Fig. S5c).

For application to European cities, reconstructed wastewater concentrations for the 43 target chemicals were entered into the selected model. For chemical i in urban area j, the estimated DE was divided by body mass (BW) and translated under a standardized one-compartment steady-input approximation using $t_{1/2}$ and Vd from HExpPredict [37]. Multiplication by the dimensionless molar REP_i yielded μmol L^{-1} narciclasine-equivalent BEQ. For reporting in μg L^{-1} narciclasine-EQ, this molar equivalent was multiplied by the molecular mass of the reference compound narciclasine, MW_{Narc} = 307.25 g mol^{-1}. The combined equation was

$$BEQ_{ij} = (DE_{ij}/\mathrm{BW}) \times [t_{1/2}/(Vd_i \times \ln 2)] \times \mathrm{MW_{Narc}} \times REP_i$$

and the urban mixture BEQ was obtained by summing chemical-specific equivalents,

$$BEQ_j^{\mathrm{mixture}} = \sum_{i=1}^{43} BEQ_{ij}$$

The internal-equivalent concentration implied by the bracketed toxicokinetic term was retained only as a computational bridge and was not reported as a separate result level. Chemical and urban-area BEQ values were expressed in μg L^{-1} narciclasine-EQ. BEQ was interpreted as an endpoint-specific, screening-level effect-equivalent metric, rather than an estimate of neurological disease risk or a measured human blood concentration.

To quantify how successive biological translations altered chemical priorities, chemicals were ranked independently according to their across-urban-area median WW, DE, and BEQ values. Ranks were further classified into high (rank 1-15), intermediate (16-29), and low (30-43) strata. Across-city arithmetic means were used for the mean chemical rankings and mean BEQ contributions. Both means and medians, their ranks and the median-based strata are reported in Table S10. Spearman correlations quantified concordance between WW, DE, and BEQ rankings, and transitions among strata were used to identify chemicals whose priority changed during WW →DE→BEQ translation.

A complementary comparison used HExpPredict, an alternative model for predicting human blood exposure [37]. Both sets of concentration estimates were weighted by the same DeepToxLab neurotoxicity probabilities [41]. Across 43 matched chemicals, the resulting scores showed moderate concordance (R^2 = 0.547, R = 0.592; Extended Data Fig. 6).

To examine whether national food-supply structure covaried with wastewater chemical profiles, 2015–2019 FAOSTAT Food Balance Sheet quantities for 71 food items were summarized as national medians for the 18 represented countries and paired with country-level median wastewater concentrations of the 43 target chemicals. This analysis concerned national covariation with wastewater composition, not dietary source attribution or validation of the exposure model. Detailed data harmonization, permutation testing, food-module definitions and stratification procedures are provided in Supplementary Section 3.

Acknowledgments:
We would like to express our gratitude to the working group of environmental exposure and human health of the China Cohort Consortium (http://chinacohort.bjmu.edu.cn/)

Funding:
This study was supported by the National Natural Science Foundation of China (Grant No. 42477455; 42677546), National Key Research & Development Program of Ministry of Science and Technology of China (2023YFC3708305)

Author contributions:

X. Li and B. W. conceptualized the study. N. F., X. Li and B. W. developed the methodology, prepared the figures and wrote the manuscript. N. F., X. Li and B. W. contributed to the data analysis. N. F., S. A., S. Y., Q. W., T. Wu, T. Wang, Z. C., X. Lou, H. Z., M. R., J. Xu, J. Xue, X. Lu, Y. H., M. F., P. G.-F., N. A., F. M., X. Li and B. W. critically reviewed and revised the manuscript. X. Li and B. W. managed the project.

Competing interest:

Authors declare no competing interests.

Data and Code materials availability:

The data and code used in this study can be accessed on Zenodo (see: https://doi.org/10.5281/zenodo.23914160). The repository comprises seven supplementary code packages and an additional local web platform:

(1) Wastewater Imputation and Sensitivity Analysis
Reconstruction of a wastewater concentration matrix covering 117 urban areas and 663 chemicals, including reference-feature construction, repeated masking validation, reconstruction comparisons and observation-removal sensitivity analyses.

(2) WW-DE Model Development and Validation
Development of the wastewater-to-daily-exposure model using 472 chemicals, including tenfold cross-validation, predictor importance, fitted model files and wastewater concentration response analyses.

(3) WPN Effect-Weighted Chemical Prioritization
Application of the exposure model to the 43 target chemicals, BEQ calculation using HExpPredict toxicokinetic parameters and relative effect potencies, chemical ranking, city-level mixture summaries, chemical-class contributions and comparison with HExpPredict.

(4) E-BIO: Linking All Toxic Chemicals to Neurobehavioral Disorders
Processing and analysis of exported ExposomeX E-BIO networks for the literature-derived inventory of 1,554 candidate chemicals, including protein- and Gene Ontology-mediated disease associations, network visualization and mechanism-overlap analyses.

(5) E-BIO: Linking NT43 to Neurobehavioral Disorders
Corresponding E-BIO network analyses and visualizations for the 43 target chemicals.

(6) ExDGNN Chemical-Disease Association Prediction
Molecular representation, graph neural network training and evaluation, candidate-disease ranking and prediction of 1,849 associations between 43 chemicals and 43 neurological disease endpoints.

(7) Food Supply-Wastewater Association Analysis
Analysis of national food-supply patterns and wastewater chemical concentrations across 18 countries, including Spearman correlations, false-discovery-rate correction, country-label permutation, leave-one-country-out assessment and food-module summaries.

(8) WPN-Local_Web-Windows-macOS-CPU
A locally runnable, browser-based WPN platform. Its Overview, Wastewater, WPN and Rank modules provide access to study methods, wastewater data, reconstruction validation and chemical prioritization results. Users can browse and filter data, export tables and figures, and enter wastewater concentrations in nM for any subset of 1–43 supported NT43 chemicals. The platform applies the saved exposure model, calculates BEQs and chemical contributions, and displays WW and BEQ ranking plots with cumulative contribution curves and logarithmic or linear axes. The package includes website files, models, preprocessing files, supporting data

and numbered setup and operation instructions. CPU launchers are provided for Windows 10/11 (64-bit) and Apple Silicon Macs running macOS 14 or later, using Python 3.12. Initial installation requires internet access; subsequent local use can operate offline. See 0_README.md for instructions.

Readers can contact Dr. Bin Wang for more detailed information, Email: binwang@pku.edu.cn.

Supplementary Materials:

Supplementary Sections 1 to 5
Figs. S1 to S8
Tables S1 to S17

References and Notes:

1 Steinmetz, J. D. *et al.* Global, regional, and national burden of disorders affecting the nervous system, 1990-2021: a systematic analysis for the Global Burden of Disease Study 2021. *Lancet Neurol* **23**, 344-381 (2024). https://doi.org:10.1016/S1474-4422(24)00038-3

2 Grandjean, P. & Landrigan, P. J. Neurobehavioural effects of developmental toxicity. *Lancet Neurol* **13**, 330-338 (2014). https://doi.org:10.1016/S1474-4422(13)70278-3

3 Escher, B. I., Stapleton, H. M. & Schymanski, E. L. Tracking complex mixtures of chemicals in our changing environment. *Science* **367**, 388-392 (2020). https://doi.org:10.1126/science.aay6636

4 Vermeulen, R., Schymanski, E. L., Barabasi, A. L. & Miller, G. W. The exposome and health: Where chemistry meets biology. *Science* **367**, 392-396 (2020). https://doi.org:10.1126/science.aay3164

5 Backhaus, T. *et al.* Include a mixture allocation factor to improve EU chemical risk management. *Science* **390**, 678-680 (2025). https://doi.org:10.1126/science.aeb6374

6 Gilles, L. *et al.* HBM4EU combines and harmonises human biomonitoring data across the EU, building on existing capacity - The HBM4EU survey. *Int J Hyg Environ Health* **237**, 113809 (2021). https://doi.org:10.1016/j.ijheh.2021.113809

7 Braun, G. *et al.* Neurotoxic mixture effects of chemicals extracted from blood of pregnant women. *Science* **386**, 301-309 (2024). https://doi.org:10.1126/science.adq0336

8 Carneiro, R. B. *et al.* A critical review of wastewater-based epidemiology as a tool to evaluate the unintentional human exposure to potentially harmful chemicals. *Anal Bioanal Chem* **417**, 495-511 (2025). https://doi.org:10.1007/s00216-024-05596-z

9 Andrianou, X. D. & Makris, K. C. The framework of urban exposome: Application of the exposome concept in urban health studies. *Sci Total Environ* **636**, 963-967 (2018). https://doi.org:10.1016/j.scitotenv.2018.04.329

10 Gracia-Lor, E. *et al.* Measuring biomarkers in wastewater as a new source of epidemiological information: Current state and future perspectives. *Environ Int* **99**, 131-150 (2017). https://doi.org:10.1016/j.envint.2016.12.016

11 Kasprzyk-Hordern, B. *et al.* Wastewater-based epidemiology for the assessment of population exposure to chemicals: The need for integration with human biomonitoring for global One Health actions. *J Hazard Mater* **450**, 131009 (2023). https://doi.org:10.1016/j.jhazmat.2023.131009

12 Baz-Lomba, J. A. *et al.* Comparison of pharmaceutical, illicit drug, alcohol, nicotine and caffeine levels in wastewater with sale, seizure and consumption data for 8 European cities.

BMC Public Health **16**, 1035 (2016). https://doi.org:10.1186/s12889-016-3686-5

13 Been, F. *et al.* Mining the Chemical Information on Urban Wastewater: Monitoring Human Exposure to Phosphorus Flame Retardants and Plasticizers. *Environ Sci Technol* **52**, 6996-7005 (2018). https://doi.org:10.1021/acs.est.8b01279

14 Choi, P. M. *et al.* Social, demographic, and economic correlates of food and chemical consumption measured by wastewater-based epidemiology. *Proc Natl Acad Sci U S A* **116**, 21864-21873 (2019). https://doi.org:10.1073/pnas.1910242116

15 Duan, L. *et al.* Wastewater surveillance for 168 pharmaceuticals and metabolites in a WWTP: Occurrence, temporal variations and feasibility of metabolic biomarkers for intake estimation. *Water Res* **216**, 118321 (2022). https://doi.org:10.1016/j.watres.2022.118321

16 Kasprzyk-Hordern, B. *et al.* Estimation of community-wide multi-chemical exposure via water-based chemical mining: Key research gaps drawn from a comprehensive multi-biomarker multi-city dataset. *Environ Int* **147**, 106331 (2021). https://doi.org:10.1016/j.envint.2020.106331

17 Li, D. *et al.* Trends in artificial sweetener consumption: A 7-year wastewater-based epidemiology study in Queensland, Australia. *Sci Total Environ* **754**, 142438 (2021). https://doi.org:10.1016/j.scitotenv.2020.142438

18 Rousis, N. I. *et al.* Wastewater-based epidemiology to assess pan-European pesticide exposure. *Water Res* **121**, 270-279 (2017). https://doi.org:10.1016/j.watres.2017.05.044

19 Tang, S. *et al.* Urinary Concentrations of Bisphenols in the Australian Population and Their Association with the Per Capita Mass Loads in Wastewater. *Environ Sci Technol* **54**, 10141-10148 (2020). https://doi.org:10.1021/acs.est.0c00921

20 Castiglioni, S. *et al.* Evaluation of uncertainties associated with the determination of community drug use through the measurement of sewage drug biomarkers. *Environ Sci Technol* **47**, 1452-1460 (2013). https://doi.org:10.1021/es302722f

21 Dulio, V. *et al.* Beyond target chemicals: updating the NORMAN prioritisation scheme to support the EU chemicals strategy with semi-quantitative suspect/non-target screening data. *Environmental Sciences Europe* **36**, 113 (2024). https://doi.org:10.1186/s12302-024-00936-3

22 Hollmann, N. *et al.* Accurate predictions on small data with a tabular foundation model. *Nature* **637**, 319-326 (2025). https://doi.org:10.1038/s41586-024-08328-6

23 Braun, G. & Escher, B. I. Prioritization of mixtures of neurotoxic chemicals for biomonitoring using high-throughput toxicokinetics and mixture toxicity modeling. *Environ Int* **171**, 107680 (2023). https://doi.org:10.1016/j.envint.2022.107680

24 Wetmore, B. A. *et al.* Integration of dosimetry, exposure, and high-throughput screening data in chemical toxicity assessment. *Toxicol Sci* **125**, 157-174 (2012). https://doi.org:10.1093/toxsci/kfr254

25 Chen, J. *et al.* A machine-learning approach clarifies interactions between contaminants of emerging concern. *One Earth* **5**, 1239-1249 (2022). https://doi.org:10.1016/j.oneear.2022.10.006

26 Chen, J. *et al.* Mitigating pesticide mixture hazard in global surface waters through agricultural management. *One Earth* **8**, 101163 (2025). https://doi.org:10.1016/j.oneear.2024.11.017

27 Brack, W. *et al.* Effect-based methods are key. The European Collaborative Project SOLUTIONS recommends integrating effect-based methods for diagnosis and monitoring of water quality. *Environmental Sciences Europe* **31**, 10 (2019).

https://doi.org:10.1186/s12302-019-0192-2

28 Escher, B., Braun, G. & Zarfl, C. Exploring the Concepts of Concentration Addition and Independent Action Using a Linear Low-Effect Mixture Model. *Environ Toxicol Chem* **39**, 2552-2559 (2020). https://doi.org:10.1002/etc.4868

29 Escher, B. I. *et al.* Mixtures of organic micropollutants exacerbated in vitro neurotoxicity of prymnesins and contributed to aquatic toxicity during a toxic algal bloom. *Nature Water* **2**, 889-898 (2024). https://doi.org:10.1038/s44221-024-00297-4

30 Escher, B. I. *et al.* Benchmarking organic micropollutants in wastewater, recycled water and drinking water with in vitro bioassays. *Environ Sci Technol* **48**, 1940-1956 (2014). https://doi.org:10.1021/es403899t

31 Lee, J. *et al.* Monitoring Mixture Effects of Neurotoxicants in Surface Water and Wastewater Treatment Plant Effluents with Neurite Outgrowth Inhibition in SH-SY5Y Cells. *ACS Environ Au* **2**, 523-535 (2022). https://doi.org:10.1021/acsenvironau.2c00026

32 Neale, P. A. *et al.* Wastewater treatment efficacy evaluated with in vitro bioassays. *Water Res X* **9**, 100072 (2020). https://doi.org:10.1016/j.wroa.2020.100072

33 Li, X. *et al.* Mapping global urban boundaries from the global artificial impervious area (GAIA) data. *Environmental Research Letters* **15**, 094044 (2020). https://doi.org:10.1088/1748-9326/ab9be3

34 Hommel, E. *et al.* Following the Mixtures of Organic Micropollutants with In Vitro Bioassays in a Large Lowland River from Source to Sea. *ACS Environ Au* **5**, 363-375 (2025). https://doi.org:10.1021/acsenvironau.4c00059

35 Gutierrez-Martin, D. *et al.* Comprehensive profiling and semi-quantification of exogenous chemicals in human urine using HRMS-based strategies. *Anal Bioanal Chem* **415**, 7297-7313 (2023). https://doi.org:10.1007/s00216-023-04998-9

36 Been, F. *et al.* Population normalization with ammonium in wastewater-based epidemiology: application to illicit drug monitoring. *Environ Sci Technol* **48**, 8162-8169 (2014). https://doi.org:10.1021/es5008388

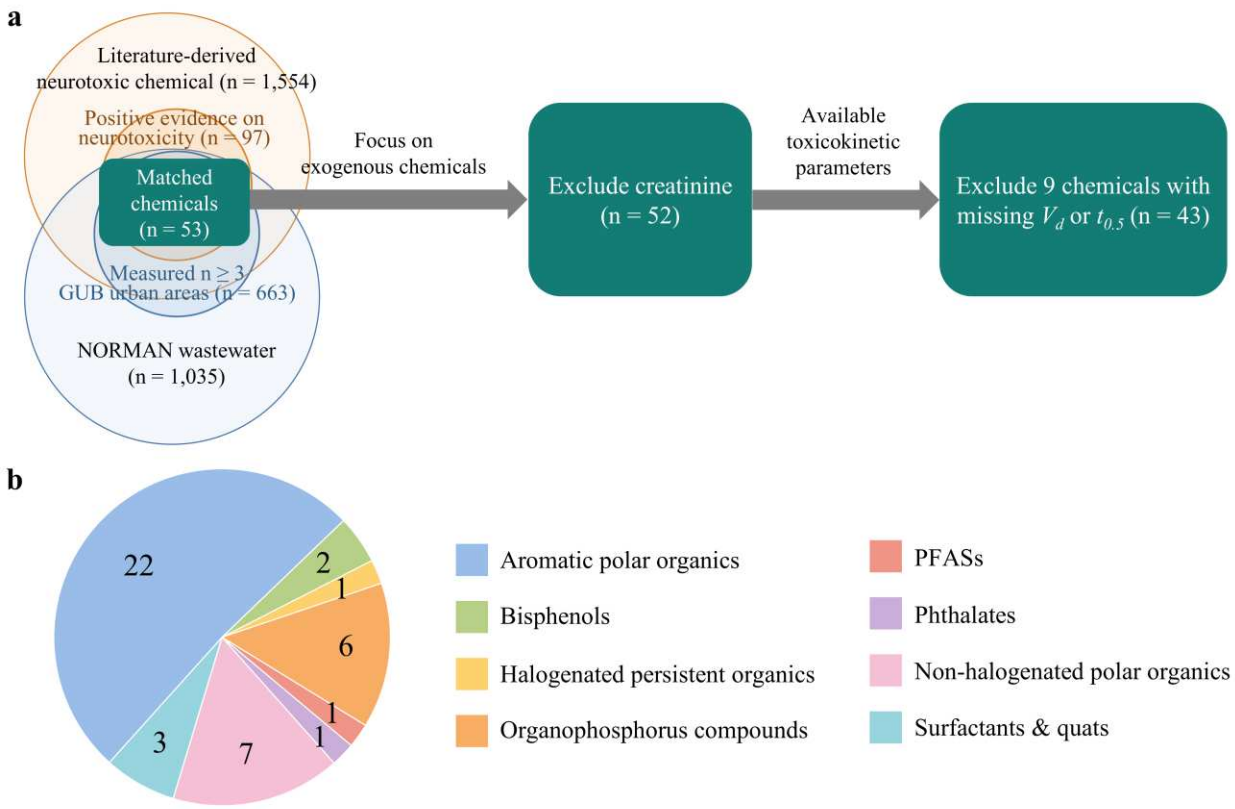
37 Zhao, F. *et al.* HExpPredict: In Vivo Exposure Prediction of Human Blood Exposome Using a Random Forest Model and Its Application in Chemical Risk Prioritization. *Environ Health Perspect* **131**, 37009 (2023). https://doi.org:10.1289/EHP11305

38 Wang, B. *et al.* ExposomeX: Development of an Integrative Exposomic Platform to Expedite Discovery of the "Exposure-Biology-Disease" Nexus. *Environ Sci Technol* **59**, 13251-13263 (2025). https://doi.org:10.1021/acs.est.5c05956

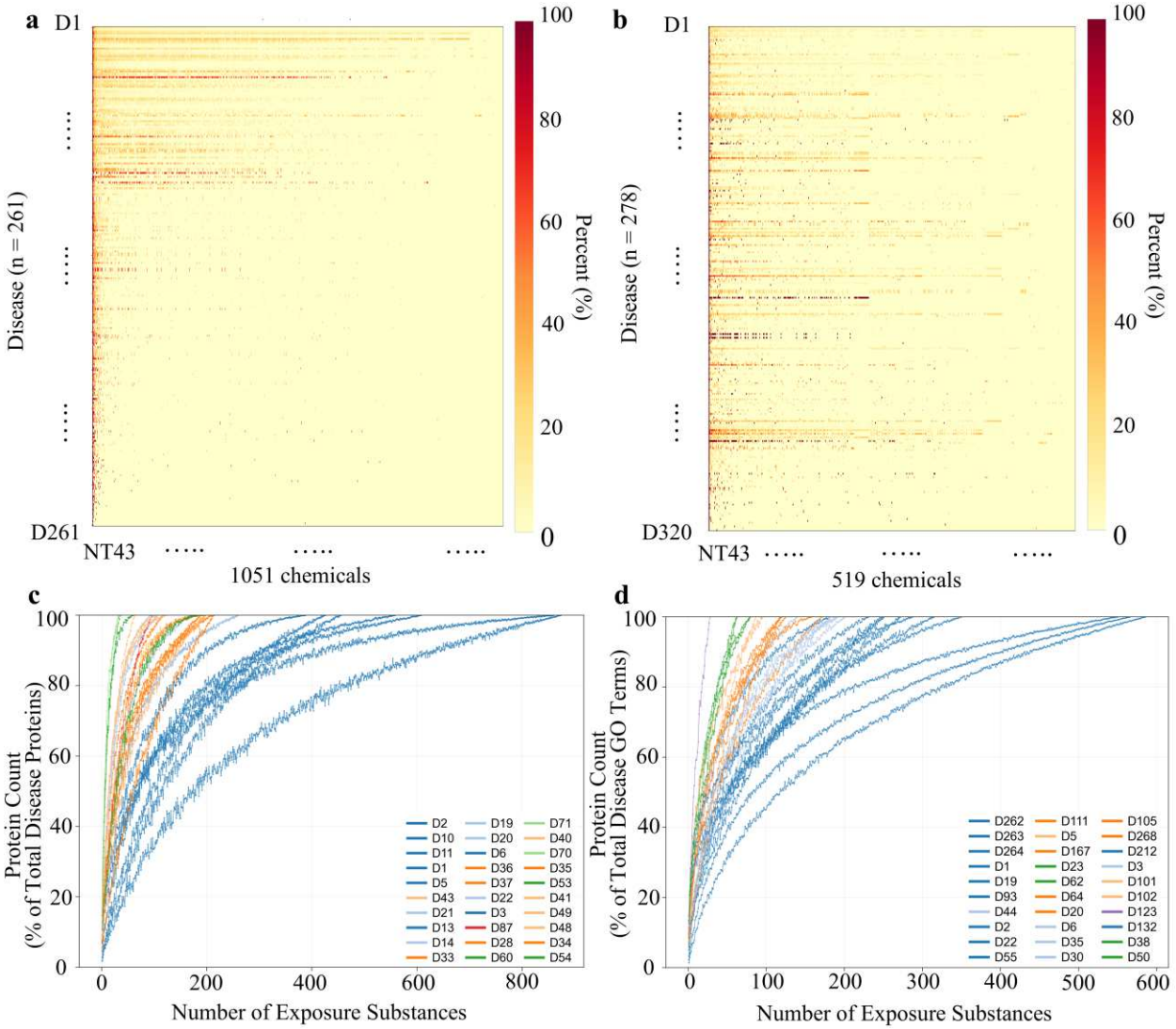
39 Chen, Y., Liu, G., Yang, S. & Wang, B. Control Priorities for Persistent Organic Pollutant Mixture Based on Neurodevelopmental Disorders. *Environ Sci Technol* **60**, 21440-21452 (2026). https://doi.org:10.1021/acs.est.6c00498

40 Ring, C. L. *et al.* Consensus Modeling of Median Chemical Intake for the U.S. Population Based on Predictions of Exposure Pathways. *Environ Sci Technol* **53**, 719-732 (2019). https://doi.org:10.1021/acs.est.8b04056

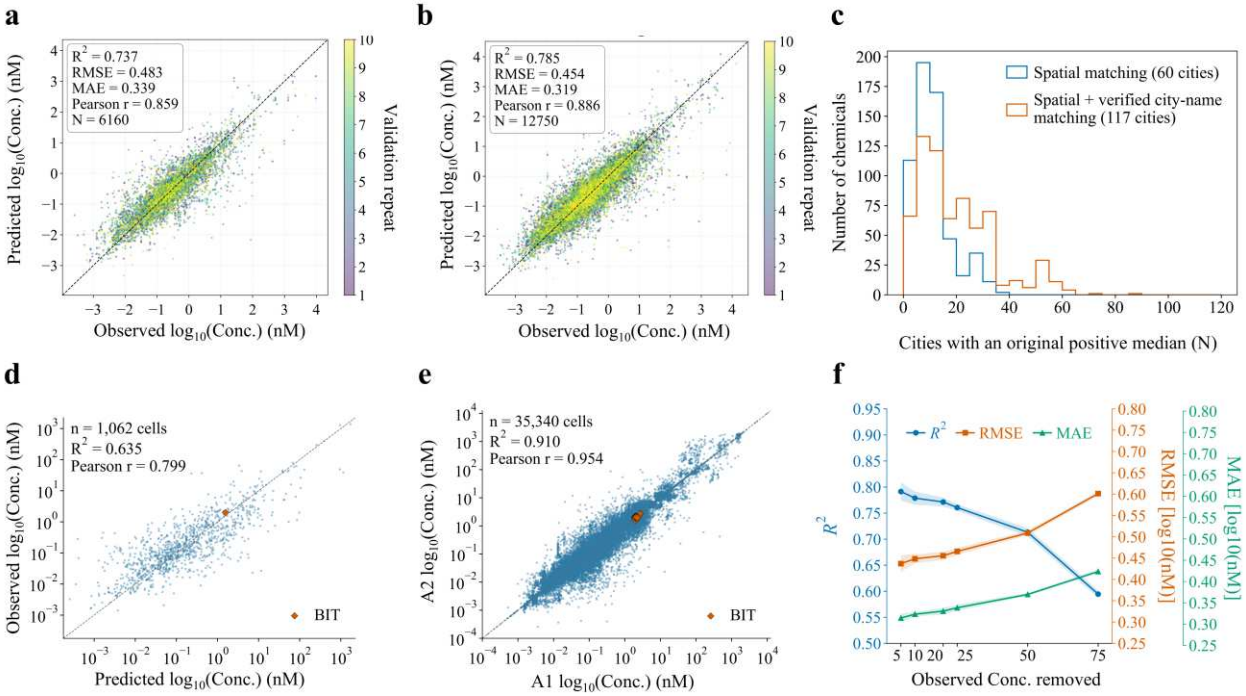
41 Fu, L. *et al.* ADMETlab 3.0: an updated comprehensive online ADMET prediction platform enhanced with broader coverage, improved performance, API functionality and decision support. *Nucleic Acids Res* **52**, W422-W431 (2024). https://doi.org:10.1093/nar/gkae236



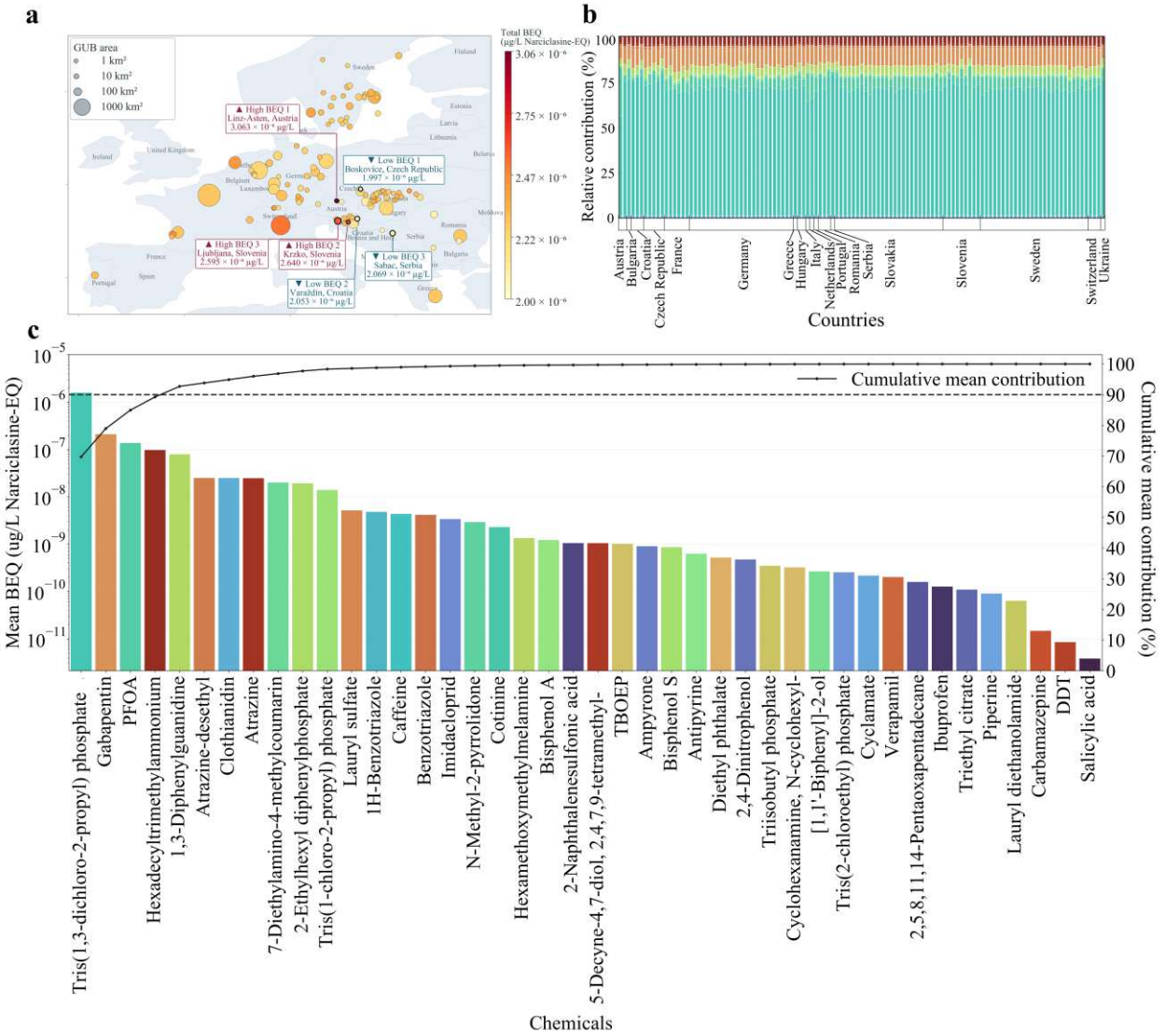
Extended Data Fig. 1 | Selecting the target 43 chemicals from the exposome database **a**, Implemented chemical-selection chain. The literature-derived candidate set comprised 1,554 chemicals. Among these candidates, 97 had positive numerical relative-effect-potency (REP) values for neurite-outgrowth inhibition. The NORMAN wastewater archive contained 1,035 standardized chemicals, of which 663 were measured in at least 3 European Global Urban Boundary (GUB) urban areas. Intersecting the 97 positive-REP chemicals with the 663 sufficiently monitored wastewater chemicals yielded 53 candidates. Exclusion of the endogenous marker creatinine left 52 exogenous target chemicals. Among them, 43 target chemicals (NT43) had the available toxicokinetic parameters (*Vd* and $t_{0.5}$). **b**, Distribution of the 43 chemicals among eight prespecified classes.



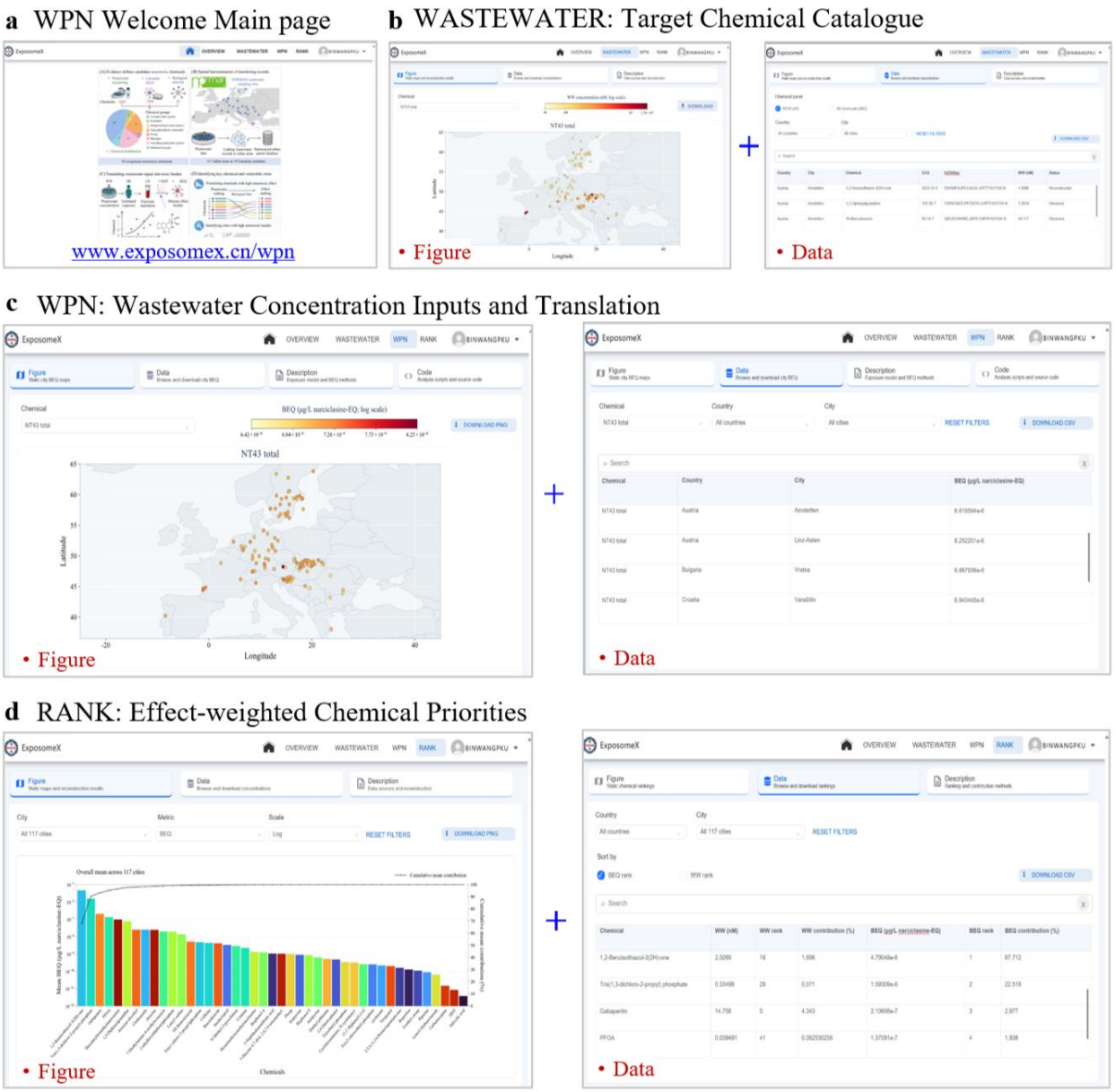
Extended Data Fig. 2 | Disease-associated coverage in the EPPD and EGoD networks. **a**, Disease-specific protein coverage in the EPPD network for the literature-derived candidate set. Individual-chemical columns show returned associations. **b**, GO-term coverage in the EGoD network or the literature-derived candidate set. Individual-chemical columns show returned associations. The first column shows the fixed NT43 union. Each percentage uses all associated nodes for that disease in the corresponding full network as its denominator. Coverage across chemicals overlaps and is not additive. **c**, Mean protein coverage for the 30 diseases with the largest network node unions as the number of sampled chemicals increases. **d**, Mean GO-term coverage for the 30 diseases with the largest network node unions as the number of sampled chemicals increases. Subsets are drawn from the chemicals already linked to each disease, with up to 100 distinct combinations at each subset size. The detailed disease IDs' information of various were provided in Table S4.



Extended Data Fig. 3 | Sensitivity and stability of wastewater concentration reconstruction. **a**, Observed versus deterministic reconstructed $\log_{10}$(nM) concentrations from ten approximately 10% masking repeats: A1 (589 chemicals, 60 urban areas; 6,160 predictions from 3,981 distinct cells). **b**, Observed versus deterministic reconstructed $\log_{10}$(nM) concentrations from ten approximately 10% masking repeats: A2 (663 chemicals, 117 urban areas; 12,750 predictions from 8,339 distinct cells). Colours identify repeats; dashed lines indicate 1:1 agreement. **c**, Original observed-city coverage per chemical. **d**, A1 predictions compared with 1,062 newly observed A2 medians. **e**, Final residual-sampled A1-A2 matrix consistency over 35,340 shared combinations, including retained measurements and reconstructed values. Orange diamonds identify BIT. **f**, Deterministic reconstruction after removing 5%, 10%, 20%, 25%, 50% or 75% of original A2 observations. Lines and shading show ten-repeat means and pointwise 95% t intervals of the conditional random-masking mean.

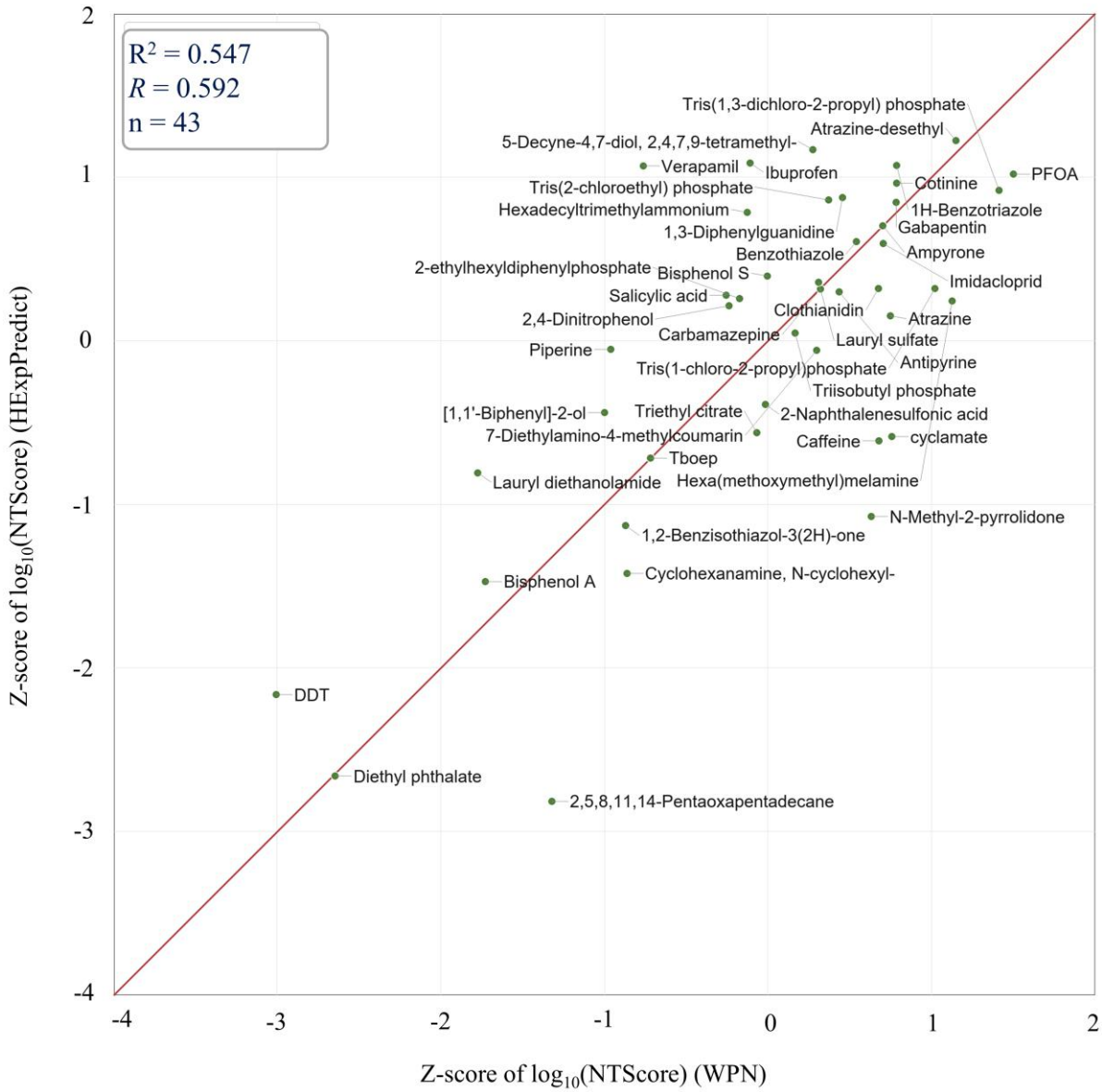


Extended Data Fig. 4 | City-level BEQ profiles after excluding BIT.
a, Total BEQ across 117 urban areas after excluding BIT. Symbol size represents GUB area and colour represents total residual BEQ. **b**, Relative contributions of the remaining 42 chemicals within each urban area. **c**, Across-city mean BEQ by chemical and cumulative contribution, ordered by mean BEQ. Shares are normalized to the total after BIT exclusion.



Extended Data Fig. 5 | Overview of the WPN web platform.

a, Welcome page presenting the overall WPN framework. **b**, WASTEWATER module, providing influent concentrations by city and chemical in map and table views. The data table distinguishes measured from reconstructed concentrations. **c**, WPN module, displaying BEQs by city and chemical in map and table views and provides downloadable code for calculating BEQs from users' own influent measurements. **d**, RANK module, comparing chemical rankings by influent concentration and BEQ for individual cities or the across-city mean profile. In the profile shown, coloured bars represent chemical-specific BEQs, and the black line shows their cumulative contribution to total BEQ.



Extended Data Fig. 6 | External consistency of effect-weighted chemical ordering using WPN and HExpPredict.

For 43 matched chemicals, each concentration estimate was multiplied by the same DeepToxLab neurotoxicity probability; log10 scores were standardized separately before comparison.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarytables.zip
- 2SM20261110.docx