

Comprehensive GC×GC-MS-based Non-Target Screening to Unveil Chemical Profiles of Indoor Dust from Seven European Countries

Andriy Rebryk

Umeå University

Lisa Melymuk

RECETOX

Linda Schenk

Karolinska Institutet

Eva Sugeng

Vrije Universiteit Amsterdam

Pim Leonards

Vrije Universiteit Amsterdam

Marta Gabriel

INEGI

Anna Maria Ingelido

ISS, Italian National Institute of Health

Kristina Aidla

Republic of Estonia Health Board

Stamatina Violaki

Vrije Universiteit Amsterdam

Joana Azeredo

INEGI

Rok Novak

Jožef Stefan Institute

Neža Palir

Jožef Stefan Institute

Valentina Marra

ISS, Italian National Institute of Health

Marco Inglessis

ISS, Italian National Institute of Health

Anna-Sofia Preece

Karlstad University

Konstantin Kislitsõn

Republic of Estonia Health Board

Riin Tarto

Republic of Estonia Health Board

Hans Orru

University of Tartu

Ene Indermitte

University of Tartu

Peter Haglund

`peter.haglund@umu.se`

Umeå University

Research Article

Keywords: Indoor dust, Organic chemical contaminants, GC×GC-MS, Non-target screening, Functional use classification, Geographical trends, Human exposure, Risk assessment

Posted Date: September 24th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10956989/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

People spend most of their time indoors, where dust serves both as an important reservoir of organic chemicals released from building materials, consumer products, activities, and outdoor sources, and as a significant human exposure pathway, particularly through inadvertent dust ingestion. Previous studies have primarily focused on targeted analysis or limited suspect screening of selected contaminant classes, providing only a partial characterization of the chemical complexity of dust. This study – as a part of the INQUIRE project – aimed to comprehensively characterize the lipophilic organic chemical composition of indoor dust collected from households in seven European countries using comprehensive two-dimensional gas chromatography–mass spectrometry (GC×GC-MS) and an advanced non-target screening (NTS) workflow.

Indoor dust samples from 174 households were analyzed using novel data processing workflows that enabled comprehensive feature extraction, alignment, and annotation. Over 2300 anthropogenic and naturally occurring compounds were tentatively identified and classified according to their functional uses using multiple complementary chemical databases. (Semi-)quantitative concentration estimates revealed pronounced differences in chemical composition among countries and identified distinct national chemical fingerprints. Country-specific signature compounds included organophosphates, plasticizers, personal care product ingredients, pharmaceuticals, pesticides, fragrances, and naturally derived compounds, reflecting differences in consumer product use, building materials, lifestyle, and environmental conditions across the examined countries. The inclusion of natural compounds alongside anthropogenic contaminants provided a more comprehensive characterization of indoor dust than has been reported in previous large-scale studies. The risk from dust ingestion exposure was estimated via hazard quotient (HQ) and cancer risk (CR) calculations. No median HQs exceeded 1 and CRs exceeded 10^{-6} only in a small subset of households for several polycyclic aromatic hydrocarbons and dyes.

This work represents one of the most comprehensive NTS studies of organic contaminants in European indoor dust to date. The results demonstrate the complexity of indoor dust chemistry and highlight the value of combining comprehensive chromatographic separation and advanced data processing for large-scale environmental screening. Although no median HQs exceeded 1, the assessment considering only dust ingestion likely underestimates overall exposure. CRs exceeding 10^{-6} indicate that dust ingestion may contribute to non-negligible lifetime cancer risk in some indoor environments. The generated dataset substantially expands current knowledge of chemicals present in indoor dust and provides a foundation for future studies investigating exposure sources and the compounds contributing to biological activity and potential health effects within the INQUIRE project.

1. Introduction

According to the World Health Organization (WHO), air pollution remains one of the leading environmental health risks worldwide, contributing to approximately 9 million premature deaths each year (1). Because people spend most of their time indoors – typically more than 50% of each day (13–16 hours/day) and, in some populations, up to 90% (over 21 hours/day) following the COVID-19 pandemic (2, 3) – indoor environments have become an increasingly important contributor to overall chemical exposure. Among indoor exposure media, settled house dust represents a particularly relevant matrix because it accumulates chemicals released from building materials, furnishings, electronic equipment, consumer products, occupant activities, and outdoor infiltration over extended periods (4–6). Consequently, indoor dust provides an integrated record of the indoor chemical environment while simultaneously serving as a significant exposure pathway, particularly through inadvertent dust ingestion, which is considered one of the dominant routes of exposure for many organic contaminants, especially in infants and young children due to their frequent hand-to-mouth behavior and close contact with indoor surfaces (4–8).

Indoor dust contains a complex mixture of volatile and semi-volatile organic compounds (VOCs and SVOCs), including both anthropogenic contaminants and naturally occurring substances originating from a wide range of indoor and outdoor sources (4–6). Historically, many studies of dust contamination have been restricted to targeted analyses or limited suspect screening with a focus on legacy or well-known compound classes such as pesticides, phthalates, plasticizers, per- and polyfluoroalkyl substances (PFASs), brominated flame retardants (BFRs), and organophosphorus substances (OPs) (9–11), constraining the range of compounds that can be detected.

Over time, the chemical space of interest has broadened to include contaminants of emerging concern (CECs), substitution compounds, and more polar contaminants, but comprehensive characterization remains challenging. In an early exemplary work, Moschet *et al.* proposed an integrated protocol combining target, suspect, and non-target screening (NTS) of Californian house dust using both liquid and gas chromatography (LC and GC, respectively) as well as *in-silico* approach, addressing the need to widen analytical coverage beyond preselected compounds (4). In that effort, NTS served to identify previously unmonitored and unexpected compound classes, complementing targeted analysis and quantification. In 2023, Zhu *et al.* used GC-Orbitrap-MS to reveal 22 additional SVOCs in Danish kindergarten dust beyond those in their suspect list as well as previously unreported antioxidant Irganox 1076 (12). In the most recent study – a 2024 NORMAN Network collaborative trial on indoor settled dust – Haglund *et al.* applied GC-MS, two-dimensional GC (GC×GC)-MS, and LC-MS based target analysis, suspect screening, and NTS approaches across multiple European homes to quantify and semi-quantify levels of a wide range of contaminants, resulting in tentative identification and (semi-)quantification of more than 1200 anthropogenic compounds (6). That study draws on the experiences gained through an inter-laboratory study conducted in 2019 by the NORMAN research network, in which suspect screening and NTS were utilized to create an organic indoor contaminants database (5). Such studies underscore both the potential of complementary analytical approaches and remaining gaps in chemical coverage of indoor dust matrices.

Despite these advances, most existing studies focus on anthropogenic chemical classes (e.g., additives, flame retardants, UV filters, pesticides) and seldom include naturally derived compounds (e.g., plant secondary metabolites, natural organic transformation products). A retrospective dust-analysis study conducted in 2022 by Dubocq *et al.* demonstrated that in 30 indoor dust samples, nearly 298 compounds could be tentatively identified, including plant natural products, pharmaceuticals, food additives, colorants, and UV stabilizers, indicating that organic dust composition extends far beyond classical contaminants (13). Thus, a comprehensive NTS analytical approach can reveal a more holistic view of dust composition, including both anthropogenic and natural constituents. Ideally, such a comprehensive chemical characterization of indoor dust should be complemented with an equally broad characterization of the biological potency of such samples and extracts thereof.

Within this larger framework, our work implements a GC×GC-MS-based NTS of organic compounds in settled indoor dust from seven European countries (Czechia, Estonia, Italy, the Netherlands, Portugal, Slovenia, and the United Kingdom). Moreover, in contrast to previous efforts that largely concentrate on anthropogenic chemical classes, our approach integrates detection of both anthropogenic and naturally occurring organic molecules (e.g., bioactive natural products and various metabolites and terpene-related compounds). INQUIRE also includes a broad characterization of the biological potency of contaminants in house dust. This broader coverage offers new insight into the full complexity of organic contaminant mixtures in indoor environments and may help bridge the gap between measured chemical concentrations and observed bioactivity or health-related effects.

Methodologically, this study leverages state-of-the-art comprehensive GC×GC-MS, which substantially increases the number of detected compounds as the additional separation in the 2nd chromatographic dimension can potentially resolve numerous chemicals co-eluting in the 1st dimension, revealing minor or hidden chemicals (15). Moreover,

sample preparation is designed to minimize selective losses (e.g., use of solvents of various polarities and minimal cleanup). This was done to preserve as many chemicals as possible – identified and unknown – to enable the broadest possible chemical and biological characterization and the establishment of causative links between individual chemicals or groups of chemicals and observed bioactivity and toxicological effects, obtained *via*, e.g., *in-vitro* toxicity assays, virtual effect directed analysis (vEDA), and quantitative structure–activity relationship (QSAR) modeling by the INQUIRE partners.

By deploying this approach, our study aims to i) provide one of the most extensive datasets of organic house dust contaminants across Europe to date, ii) reveal previously unreported compound classes (anthropogenic or natural) in indoor dust, iii) examine cross-country similarities and differences in indoor dust chemical composition, and iv) help guide prioritization of chemicals for toxicological evaluation. The unique contribution of the INQUIRE indoor dust screening lies in its wide geographic range and the inclusion of both comprehensive contaminant characterization and biological potency characterization, enabling a deeper understanding of indoor organic contaminant burdens and their potential implications for human exposure and health.

2. Materials and methods

2.1. Chemicals and reagents

SupraSolv grade (for gas chromatography) dichloromethane (DCM), *n*-hexane, methanol, and toluene were purchased from Merck KGaA (Darmstadt, Germany).

A mixture of isotopically labeled polycyclic aromatic hydrocarbons (PAHs), containing naphthalene-D₈, acenaphthene-D₁₀, phenanthrene-D₁₀, chrysene-D₁₂, perylene-D₁₂, and dibenz[a,h]anthracene-D₁₄, was purchased from Cambridge Isotope Laboratories, Inc. (CIL; Tewksbury, MA, USA) and used as an internal standard (IS); see Table S1 in the Supplementary Information (SI).

EPA 8270 MegaMix (containing VOCs/SVOCs), SVOC Additions Standard, and Terpenes MegaMix #1 (containing terpene-related compounds) were purchased from Restek Corp. (Bellefonte, PA, USA) and combined into a single reference standard mixture INQUIRE Mix #1 (see SI, Table S2). EPA 8270 Appendix IX Mix #1 and Appendix IX Mix #2 (VOCs/SVOCs), EPA 8260B Calibration Mix #1 (VOCs) and Terpenes MegaMix #2 were also purchased from Restek Corp. and combined into a single reference standard mixture INQUIRE Mix #2 (see SI, Table S2). Both reference standard mixtures were spiked with 10 µL of the IS mixture.

An *n*-alkane C7–C40 mixture was purchased from Supelco Inc. (Bellefonte, PA, USA) and used for linear retention index (RI) calculation (SI, Table S3).

2.2. Samples

House dust samples were collected from 174 households with a child ≥ 5 years old across seven European countries during years 2023–2024 by the INQUIRE project partners (see Table S4 in the SI). For seven countries (CZ, IT, PT, UK, SE, NL and EE) samples consisted of settled dust from accessible locations within the main living area (e.g., in children's play areas, on carpets, and around furniture, while surfaces directly adjacent to cooktops, fireplaces, or electronic equipment were avoided), collected as a composite sample. Participants were asked to refrain from cleaning or vacuuming their floors in the main living area for 2–3 days before sampling. Dust was collected using a conventional vacuum cleaner fitted with a forensic filter holder sampling head (Vacuumark, BVDA International B.V., the Netherlands) containing a pre-cleaned glass fiber filter (Whatman GF/F, 0.7 µm pore size). Additionally, Slovenian

samples were supplemented by dust from the vacuum-cleaner bag because of low dust volumes collected on the filter. Although it was ensured that the vacuum cleaner was only used indoors and in the immediate living spaces, it cannot be ruled out that the chemicals/chemical burden detected are from other parts of the household or originate from the bag itself. After collection, filters with collected dust were sealed in their holders, wrapped in aluminum foil, stored in zip-lock bags, transported to the laboratory, and frozen at -18°C until analysis. In each home, dust sampling was repeated a second time approximately four weeks after the first sample collection, and both samples were combined and used in further processing.

Field blanks (FBs), consisting of cleaned glass fiber filters transported in the same forensic filter housings, were acquired and analyzed in parallel to dust samples for each country. Additionally, procedural blanks (PBs) and one portion of the reference material "SRM 2585 – Organic Contaminants in House Dust" from the National Institute of Standards and Technology (NIST) were analyzed in parallel with the samples from each country (see SI, Table S4).

In addition, a composite sample from the NORMAN Network European dust trial (6) was analyzed.

See Table S4 in the SI for detailed sample information.

2.3. Extraction, clean-up, and fractionation

House dust was extracted by the INQUIRE project partners using a modified protocol by Moschet *et al.* (2018) (4). In short, the filters were grinded, and the mixtures were transferred to Pyrex tubes, weighed, pooled (per household), an aliquot was taken for biological analysis by an INQUIRE partner, and the remaining was extracted as follows. Five mL *n*-hexane:acetone (3:1, v/v) were added and the samples were vortexed for 30 sec. The samples were ultrasonicated for 15 min and then centrifuged at 3000 rpm for 5 min. The supernatants were collected into a clean tube. Thereafter, 5 mL of acetone were added to the remaining dust, and the procedure was repeated. The supernatants were combined with the corresponding *n*-hexane:acetone fraction. Then, the combined extracts were evaporated using TurboVap (Biotage Sweden AB, Uppsala, Sweden) until approximately 0.5 mL remained. The tubes were centrifuged again at 3000 rpm for 5 minutes and the extracts were split into four parts for various types of analyses carried out by the INQUIRE project partners. One part of the extracts was shipped to Umeå University for further GC×GC-MS based NTS.

The extracts were spiked with the IS mixture, then passed through ISOLUTE PSA (primary-secondary amine) solid-phase extraction (SPE) cartridges (500 mg, 6 mL; Biotage Sweden AB, Uppsala, Sweden) to remove fatty acids and other highly polar compounds. The SPE cartridges were preconditioned with 5 mL of dichloromethane (DCM) using the Büchi SyncorePlus Analyst's SPE setup (Büchi Labortechnik GmbH, Essen, Germany) and a gentle vacuum. After that, the extracts were loaded onto the SPE cartridges and eluted with 20 mL of DCM, again applying gentle vacuum. Then, the extracts were evaporated, and the solvent was exchanged to toluene using Büchi SyncorePlus Analyst in the evaporation mode. The purified samples were transferred to GC vials, and the volumes were adjusted to 0.3 mL or 1 mL, depending on the respective dust equivalent weight of the initial samples. See Table S4 in the SI for detailed sample information.

The field and procedural blanks were run in parallel to the dust samples. In addition, seven country-composite samples were created by pooling aliquots of individual samples per country. Then, 10 seven-state composite samples (7SCS) were prepared by pooling a 10- μL aliquot of each country's composite sample and used for quality control (QC).

2.4. Two-dimensional gas chromatography–time-of-flight mass spectrometry (GC×GC-TOFMS)

Analyses were carried out using an Agilent 8890 GC System (Agilent Technologies, Santa Clara, CA, USA) coupled to a LECO Pegasus BT 4D GC×GC-TOFMS (LECO Corp., St. Joseph, MI, USA). A secondary oven and a quad-jet dual stage modulator were built into the main GC oven. Two serially connected 30 m × 0.25 mm × 0.25 μm 5%-phenyl methylpolysiloxane DB-5ms-UI columns (Agilent Technologies, Santa Clara, CA, USA) were used for the first-dimension separation and a 2 m × 0.25 mm × 0.25 μm 35%-trifluoropropyl methylpolysiloxane column (Rtx-200; Restek, Bellefonte, PA, USA) was used for the second-dimension separation. Helium was used as the carrier gas at a constant flow rate of 1.6 mL/min.

Samples were split-less injected at 280°C. Injection volume depended on the dust equivalent weights of the initial samples and ranged from 0.5 μL to 3 μL (Table S4 in the SI). The primary GC oven was initially maintained at 80°C for 1 min, then ramped at a rate of 3°C/min to 320°C, which was held for 5 min. The secondary oven had an offset of + 5°C relative to the first oven and the modulator had an offset of + 15°C relative to the secondary oven. The transfer line temperature was 320°C. The modulation timing was set as follows: i) start of run to 1300 s: 2.50 s (second dimension time), 0.40 s (hot pulse time), 0.85 s (cool time between stages) followed by ii) 1300 s to end of run: 2.50 s (second dimension time), 0.75 s (hot pulse time), 0.50 s (cool time between stages).

MS analyses were carried out in electron ionization (EI) mode at 70 eV electron energy, 1.0 μA emission current, and an ion source temperature of 250°C. Data were stored over the mass range of 47 to 750 amu, at an acquisition rate of 100 spectra/s. The system was tuned using perfluorotributylamine (PFTBA). Data were acquired using ChromaTOF Software for Pegasus BT (version 5.51.50.0; LECO Corp., St. Joseph, MI, USA).

Sample set, including field and procedural blanks, was divided into five batches based on the dust equivalent weights of the initial samples. Each batch was randomized and QC samples, such as alkane mixture for retention index (RI) calculation, reference standard mixtures, seven-state composite sample, and solvent (toluene) blanks, were added at the beginning and the end of each batch. Source and analyzer focus, mass calibration, gain optimization, and tune validation were run at the beginning of each sample batch. Then, two injections of toluene were made before the sample injections.

2.5. Data processing workflow

First, data were processed in ChromaTOF (version 5.51.50.0), where several types of features were classified into alkanes (C10-C36; 26 classes), high-molecular-weight phthalates (branched C8-C11; four classes), and DINCH compounds (1,2-cyclohexane dicarboxylic acid diisononyl ester and derivatives; single class) with a data processing method that utilizes chromatographic retention time regions and spectral peak filters based on diagnostic ions and ion intensity ratios (see Figure S1 and Workflow S1, section 1 in the SI). Features belonging to these compound classes, except for the main C8 phthalate, di(2-ethylhexyl) phthalate (DEHP), which was treated as an individual compound, were summed up and presented as classes. Those were processed separately from all other individual features, including DEHP.

The data were exported as separate TXT files (one for each sample). After export, features classified as alkanes, high-molecular-weight phthalates, and DINCH compounds and corresponding data, including retention index (RI) range, ranges for retention times in the 1st (RT1) and in the 2nd dimensions (RT2), formulae, base masses, and class total areas, were retrieved from the exported TXT files using a custom Python script (Python version 3.11; Python Software Foundation, Wilmington, DE, USA) Class_Retriever.py (available at https://github.com/r3bryk/Class_Retriever) and aligned into single alignment result table (ART) using Class_Aligner.py (https://github.com/r3bryk/Class_Aligner). Then, semi-quantification using internal standards was

carried out using INQUIRE_Quantifier.py (https://github.com/r3bryk/INQUIRE_Quantifier) on total component areas calculated by the script. The final concentrations were normalized to 1 g of dust.

For the (semi-)quantification of the individual compounds, the acquired GC×GC-MS data were exported as two separate SMP raw data files including data from the retention time regions, 9–61 min and 58–96 min, respectively (see Figure S2 and Workflow S1 in the SI). This step was undertaken to enable inter-sample peak alignment using tighter peak width tolerance windows for the early eluting compounds and looser tolerance windows for late eluting compounds. This helps to prevent misalignments of closely eluting VOCs with similar spectra, such as terpenes. It also helps to prevent SVOC compounds with broad peaks, such as high-molecular-weight PAHs, from splitting into duplicate or multiple features. This is a known problem of inter-sample peak alignment of comprehensive GC×GC-MS data, discussed in detail in previous work (16).

ChromaTOF Sync 2D (version 2.0.15.0; LECO Corp., St Joseph, MI, USA) was used for further data processing. It was used together with the NIST 2023 EI mass spectral library (the National Institute of Standards and Technology, Gaithersburg, MD, USA). The main data processing steps included: i) peak finding, ii) inter-sample peak alignment, iii) retention indexing, iv) library searching using the NIST 2023 EI mass spectral library, and v) data filtration using Advanced Rules of the built-in Filters functionality. Details and software settings for the workflow developed and used in this study are available in the SI, Workflow S1.

Then, the compounds/features in the two alignment result tables (ARTs) were curated manually and tentative structures were kept (good library hits), re-assigned (bad library hits), or assigned from scratch if possible (unknown features). As part of this curation process, the ARTs were exported as TSV (tab-separated values) files and additional library searching was done using JLibSearch.jl (<https://github.com/r3bryk/jLibSearch>) (16). This was done as various spectral match algorithms and settings sometimes return different tentative structures (16). Then, identification confidence levels (CLs) were assigned manually based on the scheme for GC×GC-MS proposed by Röhler and colleagues (17) with a slight modification (subdivision of CL 3 into 3a and 3b: CL 3a was assigned when positional isomers were available for the feature, whereas CL 3b was assigned when functional isomers were available, i.e. proposed compound identities could belong to different compound classes and, additionally, could include positional isomers).

After compound identification review, the two ARTs were merged into a single ART in Microsoft Excel (Microsoft, Redmond, WA, USA). After merging, the ART was examined for the presence of duplicate features coming from overlapping chromatographic region 58–61 min, and duplicates were removed from the ART.

Next, features that were associated with the compound classes captured in the first data processing step (alkanes, high-molecular-weight phthalates, and DINCH compounds) were removed using GC retention time regions and EI-MS spectra filters by the script RT_Regions_&_Spectral_Filters.py (https://github.com/r3bryk/RT_Regions_and_Spectral_Filters) to avoid overlaps between datasets. Only DEHP was kept as an individual compound.

After that, compound names in the ART were checked for unusual or broken characters, such as a German letter “ß” (*Eszett* or *scharfes S*) or “İ” instead of a Greek letter “β” (*beta*), and those were replaced with Latin analogues in the NIST library format (where applicable), e.g., “.beta.” instead of just “beta”, using a custom VBA script (<https://github.com/r3bryk/ReplaceCorruptedCharacters>) in Microsoft Excel.

Subsequently, the ART was treated with a series of custom Python scripts. First, Sync2D_Postprocessor.py (https://github.com/r3bryk/Sync2D_Postprocessor) was applied for detection frequency (DF) prioritization. This step

was needed as ChromaTOF Sync 2D rarely returns 0 as peak area value in the ART, which makes detection frequency estimations problematic (16). Hence, Sync2D_Postprocessor.py was applied for limit of detection (LOD) computation, utilizing unsupervised machine-learning (ML)-based modelling approach such as a probabilistic unsupervised ML model the Gaussian mixture model (GMM) as well as the kernel density estimation (KDE), a non-parametric method used for the probability density function estimation. The models were used to describe the bimodal distribution associated with low level noise features (artifacts) and true analyte features, respectively, and to find the inflection point. This corresponds to a peak area cutoff value that can be used to distinguish between the noise peaks and real peaks. All areas under the cutoff value were replaced with -cutoff (- LOD) values. The minus sign was used to differentiate between detected compounds (positive) and not-detected compounds (negative) during subsequent data processing and evaluation. Sync2D_Postprocessor.py also merged duplicates based on RI tolerance window and spectral similarity as well as calculated DFs and DF-based ranks applying a set of rules to facilitate feature filtration (see SI, Workflow S1, section 7).

Additional scripts were used to retrieve compound metadata from various sources. NIST_Data_Retriever.py (https://github.com/r3bryk/NIST_Data_Retriever) retrieves InChIKey, CAS#, formulae, molecular weights, and retention indices from NIST 2023 library based on compound names and PubChem_Retriever.py (https://github.com/r3bryk/PubChem_Retriever) retrieves InChIKey, SMILES, CAS#, DTXSID (Distributed Structure-Searchable Toxicity Substance Identifier), uses, and use classification values from PubChem database (18) using compound names and existing NIST InChIKey values as input. In addition, it retrieved chemical class data from ClassyFire Batch (19). Then, INQUIRE_Classfinder.py (https://github.com/r3bryk/INQUIRE_Classfinder) was used to retrieve functional use data from various databases, including the US EPA Functional Use Database (FUse) (20), NORMAN Network European dust trial data (21), the US EPA ChemExpo Function Data (22), and in-house harmonized dust functional use database, based on compound name, InChIKey, and DTXSID.

Finally, INQUIRE_Quantifier.py was applied to quantify features, for which reference standards were available (CL 1 features), and semi-quantify all other features (CLs 2–5, including unknowns) using internal standards. The calculations were made using total component areas computed by the script and the final concentrations were normalized to 1 g of dust.

2.6. Multivariate analysis

Multivariate analysis (MVA) using SIMCA 17 software (Sartorius, Umeå, Sweden) was applied for data overview, filtration, and country-specific cluster discovery.

First, principal component analysis (PCA) was utilized as a primary tool for data overview, exploring sample and contaminant clustering and patterns, and outlier discovery. Prior to PCA, the signs for all - LOD values were changed to positive (for MVA only), and the data were log-transformed and scaled to unit variance to approximate normality, minimize the dominance of high-abundance contaminants, and accentuate compositional differences between samples. In the first PCA run, all detected features (identified and unknown) in all samples, QC samples, and field and procedural blanks (denoted as FB and PB, respectively) were analyzed. The results of the PCA were further evaluated by hierarchical cluster analysis (HCA) to examine what samples form distinct clusters.

Then, features abundant in PBs (DF for PBs $\geq 50\%$ and geomean $\log_{10}(\text{concentration})$ samples vs. PBs ratio < 5) were removed from the initial dataset and PCA and HCA were re-run. The results were evaluated and samples that clustered together with FBs were removed. Thereafter, all features abundant in the FBs (ratio $\log_{10}(\text{concentration})$ samples vs. FBs < 5 and DF $\geq 50\%$ or DF $\geq 30\%$ for CL 2–4 and CL 5 compounds, respectively) were removed. PCA was re-run on all remaining samples, QC samples, and PBs to get an overview of the data and check its quality

(distribution and clustering of PBs and QC samples that should form two clusters). Next, PCA was re-run again on reduced datasets that i) included only samples (excluding PBs and QC samples) and all features (CL 1 to CL 5 (unknowns)), and ii) included only samples and CL 1 to CL 4 features (excluding alkanes and CL 5 (unknown) features), to examine score plots for clear difference between the two datasets – if there is no difference, unknowns and alkanes can be skipped from further analysis.

A multiclass orthogonal partial least squares discriminant analysis (OPLS-DA) was applied to the final reduced dataset (only samples; $n = 168$) in the SIMCA software to further explore relationships between samples from different countries and search for key country-specific and/or dominant contaminants. The results were exported as Excel files and combined into a single table. A combined analysis of OPLS-DA coefficients, VIP (variable importance of the projection) values, loadings, and detection frequencies were used to reveal compound class–country and individual signature compound–country relationships. The OPLS-DA coefficients indicate how strongly a variable (signature compound) contributes to discrimination of a given class (country); they are significant if their absolute values are greater than the corresponding confidence interval (CI). VIP values are measures of the importance of the predictor variables' contributions to the OPLS model across all latent components, i.e. how strongly a variable (compound concentration) contributes to the latent structure to discriminate between/separate countries; VIP values ≥ 1 indicate important variables. The analysis and interpretation were carried out using a Python script `PLSDA_Interpreter.py` (available at https://github.com/r3bryk/PLSDA_Interpreter) and consisted of the following steps:

i) computing loading strength (LS) to remove variables not contributing to the latent structure ($LS < 0.005$) using Eq. (1):

$$LS = \sqrt{pq1^2 + pq2^2} \text{ Eq. (1)}$$

where $pq1$ and $pq2$ are the projected values of a sample in the space of the latent variables; scores along the first and the second predictive latent variables, respectively

ii) determining maximum positive country coefficient ($MaxCoef$) to identify which country the variable (compound) most strongly predicts using Eq. (2):

$$MaxCoef = \max(CZ_{coef}, EE_{coef}, IT_{coef}, NL_{coef}, PT_{coef}, SI_{coef}, UK_{coef}) \text{ Eq. (2)}$$

where CZ_{coef} , EE_{coef} , IT_{coef} , NL_{coef} , PT_{coef} , SI_{coef} , and UK_{coef} are compound's OPLS-DA coefficients for Czechia, Estonia, Italy, the Netherlands, Portugal, Slovenia, and the United Kingdom, respectively.

1. filtering away compounds whose coefficients are insignificant ($|Coefficient| < CI$, where CI represents the confidence interval associated with the coefficient estimate); compounds passing this criterion were assigned a country significance value (CS) of 1
2. filtering away compounds with $MaxCoef \leq 0$; compounds passing this criterion were assigned a positive specificity value (PCS) of 1
3. filtering away compounds with $VIP < 1$ (standard chemometric importance threshold)
4. filtering away compounds with overall detection frequency (DF) $< 10\%$ (number of detections (DN) < 17 ; $n = 168$)
5. computing detection robustness (DR) to prioritize compounds reproducibly detected across samples using Eq. (3):

$$DR = \log_{10}(DN + 1) \text{ Eq. (3)}$$

viii) computing compound's priority score (*PS*) to prioritize compounds with strong coefficients, high VIPs, and frequent detections using Eq. (4):

$$PS = MaxCoef \times VIP \times \log_{10}(DN + 1) \text{ Eq. (4)}$$

ix) computing country specificity score (*CSS*) as a difference between the maximum positive country coefficient (*MaxCoef*) and the second-largest positive country coefficient (*SecondCoef*, a measure of how strongly each feature contributes to the model) using Eq. (5):

$$CSS = MaxCoef - SecondCoef \text{ Eq. (5)}$$

x) computing coefficient stability score (*Stability*) to prioritize compounds with strong and statistically reliable coefficients using Eq. (6):

$$Stability = \frac{MaxCoef}{AssignedCI} \text{ Eq. (6)}$$

where *AssignedCI* is the confidence interval corresponding to the assigned-country coefficient

xi) standardizing the *CSS*, *PS*, and *Stability* metrics using z-score transformation to prevent scale dominance among scoring components using Eq. (7):

$$z = \frac{x - \mu}{\sigma} \text{ Eq. (7)}$$

where *x* is the metric value, μ is the mean, and σ is the standard deviation of the metric across all compounds

1. applying class-priority weighting to reduce the ranking of compounds belonging to classes with no specific use information or poorly interpretable classes (e.g., miscellaneous compounds, generic esters, ethers, or carbonyl compounds), while preserving environmentally relevant anthropogenic contaminants; strongly deprioritized classes received a weighting factor of 0.35, whereas borderline classes received a weighting factor of 0.70
2. computing compound's final score (*FS*) using weighted standardized metrics and class-priority adjustment according to Eq. (8A) and Eq. (8B):

$$FS_{raw} = CSS_z \times 5 + Stability_z \times 2 + PS_z \text{ Eq. (8A)}$$

$$FS = FS_{raw} \times ClassPriority \text{ Eq. (8B)}$$

where CSS_z is the standardized country specificity score, $Stability_z$ is the standardized coefficient stability score, PS_z is the standardized priority score, *ClassPriority* is the class-based weighting factor. The weighting was implemented by multiplying each standardized metric by a predefined numerical coefficient (5, 2, and 1, respectively) before summation, thereby increasing or decreasing its contribution to the final score. Greater weighting (5) was assigned to country specificity because the primary aim was identification of geographically distinctive signature compounds

xiv) selecting compounds with the highest *FS* for each country and visualizing geographical differences using box-and-whisker plots and heatmaps.

2.7. Estimation of risk from dust ingestion exposure

Possible health risks associated with detected chemical contaminants were evaluated through calculation of the hazard quotients (*HQ*) and cancer risks (*CR*) following the Agency for Toxic Substances and Disease Registry (ATSDR) guidelines (23) and exposure factors relevant for the European child population (SI, Table S5). Using a

central tendency dust ingestion rate for children aged 1–6 years and a mean body weight for 2-3-year-olds, while conservatively assuming 100% bioaccessibility (24–26), the exposure assessment provided a conservative upper-central exposure estimate.

Chronic oral reference doses (*RfD*) values and cancer slope factors (*CSF*) were obtained for compounds identified at confidence level (CL) 1, where available. Only agency-published health-based guidance values were considered, e.g., acceptable daily intakes (*ADI*), tolerable daily intakes (*TDI*), reference doses (*RfD*), minimal risk levels (*MRL*), which were treated as equivalent benchmarks. Of the 120 CL 1 compounds, 49 had oral reference doses available, and 42 had oral cancer slope factors (SI, Table S6); 48 compounds had no available hazard reference values. In addition, fourteen of these 48 substances had harmonized or industry notified classifications indicating concern for systemic health effects after long-term exposure (SI, Table S6).

Dust exposure and associated risk were estimated based on standard exposure and risk equations (23):

1. Hazard quotient (*HQ*):

$$HQ = \frac{C \times IR}{BW} / RfD \text{ Eq. (9)}$$

1. Cancer risk (*CR*):

$$CR = \frac{C \times IR}{BW} \times CSF \times \frac{ED}{LY} \text{ Eq. (10)}$$

where *HQ* is the hazard quotient, *C* is the concentration measured in dust (e.g., in ng/g), *IR* is the ingestion rate of dust (mg/day), and *RfD* is a reference dose (mg/kg/day). (24–26) The *HQ* provides a screening-level metric for non-carcinogenic risk assessment of individual chemicals. An *HQ* value ≤ 1 indicates that adverse health effects are unlikely, whereas an *HQ* value > 1 suggests a potential health concern warranting additional assessment (23). *CSF* is the cancer slope factor, a measure of the potential carcinogenic effect of a given chemical. We considered an excess life-time cancer risk greater than 10^{-6} as a non-negligible risk of cancer. *ED* is the exposure duration, and *LY* is the average lifetime in years.

Indicative risks for compounds from CLs 2–3 were also estimated following the same methodology, with the acknowledgement of higher uncertainty associated with the semi-quantification and specific compound identities. For the CLs 2–3 compounds, only the subset of compounds with anthropogenic sources and high concentration in dust ($> 0.1 \mu\text{g/g}$ for CL 2 compounds and $> 1 \mu\text{g/g}$ for CL 3 compounds) was selected, e.g., compounds with use in polymers and as polymer additives, industrial chemicals, chemical intermediates, personal care products, pharmaceuticals, pesticides and biocides, surfactants, solvents, and dyes (86 features; Table S7 in the SI). Of these 86 features, 72 had no available hazard reference values (SI, Table S6). Fourteen of the 72 had harmonized or industry notified classifications indicating concern for systemic health effects after long-term exposure (SI, Table S6).

Exposure and risks were estimated individually for each compound/sample, and interpretations are based on medians and maxima. The resulting *HQ* and cancer risk estimates are discussed in section 4. Results for the CL 1 and CLs 2–3 compounds are presented separately due to differing uncertainties.

3. Results and discussion

3.1. Overview: Chemical contaminants

After alignment, 39247 features were detected initially, 86% of which were filtered away at different stages of data processing, resulting in 5372 features (Fig. 1). Of the initial features, approximately 6% (2352 features) were tentatively identified at various confidence levels (CLs): CL 1 (structure confirmed with reference standard) to CL 4 (known compound class). Of these, 0.4% (17 features) were identified at CL 4, 35% (1879) at CL 3, 6% (336) at CL 2, and 2% (120 compounds) at CL 1 (Fig. 1).

All tentatively identified compounds (CL 1 to CL 4) were classified into 26 classes by reported functional use or chemical class information (Fig. 2). Since the number of tentatively detected compounds is rather high, several classes had to be merged to keep the number of classes reasonable. Dominant classes (over 5% of the total number of tentatively identified contaminants) comprise PAHs (11%, 265 compounds), personal care products (PCPs; 11%, 257 compounds), contaminants belonging to the intermediate and research compounds (8%, 181 compounds), various esters (7%, 170 compounds; excluding fatty acid methyl esters, FAMES), biogenic terpenes and terpene-related compounds (7%, 169 compounds), and various industrial compounds, such as plasticizers (excluding phthalates) and dyes (6%, 142 compounds). These data demonstrate that the detected contaminants in the most abundant classes originate either from PCPs (indoor sources) or from a combination of indoor and outdoor sources (compound classes like PAHs; from industry, traffic, and indoor combustion such as fireplaces and candles).

3.2. Overview: Multivariate data analysis results

From the initial PCA (Figure S3 in the SI), it can be seen that i) procedural blanks (PB, black) cluster together apart from the main sample cluster, ii) field blanks (FB, blue) form a separate cluster in the lower right corner of the plot, iii) NIST SRM samples (NIST, red) and 7-state composite samples (7SCS, dark green) cluster within the main sample cluster, and iv) no clear geographical separation is obtained, although some country-specific groupings are observed. Four FBs form a sub-cluster, which covaries with petroleum hydrocarbons, i.e. they were probably petroleum contaminated during sampling or transportation. There are also four outliers among the collected dust samples: one from Czechia (pink), one from the Netherlands (turquoise), two from Slovenia (light green). These cluster close to the previously mentioned FB sub-cluster, which suggests that these may have been petroleum contaminated. Additionally, two samples from Italy are found between the FB cluster and the central sample cluster and those may also be affected by contamination, but to a lower degree.

The associated HCA (single linkage method and tree sorted by size; Figure S4 in the SI) shows several small, tight clusters at the bottom left-hand side. Those include (from left to right): i) PBs (black), ii) the four outlier samples (one from Czechia (pink), one from the Netherlands (turquoise), and two from Slovenia (light green)), iii) FBs (blue; including four potentially petroleum tainted FBs), and iv) two potentially petroleum tainted samples from Italy (brown). These clusters merge at very low heights, which means those samples are very similar to each other and probably have low overall contaminant load (as most of them are blanks) and trace levels of petroleum hydrocarbons (petroleum tainted FBs and samples).

The PCA pattern and HCA clustering remained the same when 145 features abundant in PBs were removed from the dataset, which indicates a stable clustering of the investigated dust samples. Hence, the four FBs and six dust samples (one from Czechia, one from the Netherlands, two from Slovenia, and two from Italy) that were strongly associated with the petroleum hydrocarbon background were removed from the dataset, hence, 168 out of initial 174 dust samples remained for further data analyses. Additionally, features abundant in the remaining FBs (296 features) were removed. PCA was re-run and the resulting score plot (Figure S5 in the SI) showed relatively tight clustering of PBs, FBs, NIST SRM samples, and 7SCS samples, respectively, which suggests good repeatability of the GC×GC-MS analysis and data processing procedure.

Next, the remaining FBs, PBs, NIST, and 7SCS samples were removed and PCA was re-run using the 168 house dust samples from the seven European countries (including unknown features and alkanes). The scores and loadings plots are shown in SI, Figure S6A. Here, the country-wise separation can be seen in PC2 and PC3: samples from Portugal and Slovenia are clustered more at the bottom-left of the score plot, whereas samples from Estonia, Czechia, and the Netherlands are located towards the upper-right corner (Figure S6B). The proximity of Slovenian samples to those from the southern European countries, i.e. Portugal and Italy, may reflect similarities in building design, inhabitant behavior, and consumption patterns. A southern-to-northern Europe trend (left-to-right) can be seen clearer when country centroids are computed and plotted (SI, Figure S6B, four-point stars): Portugal and Slovenia are separated from the remaining countries, with their centroids located in the lower-left region of the plot. In contrast, the centroids for Czechia, Estonia, and the UK are positioned toward the upper-right, while the Netherlands occupies the upper-left region. Italy is located in the upper-left part of the plot, intermediate between the Portugal–Slovenia cluster and the cluster formed by the Netherlands, Czechia, Estonia, and the UK, indicating a chemical profile that shares characteristics with both groups.

Nevertheless, PCA demonstrates that variations between samples from different countries are smaller than variations between households. This is expected for multivariate data analysis of big complex datasets like the one generated in this study, as PCs often can reveal weak but real relationships between biological/environmental samples. This outcome can potentially be explained by i) the unified regulation framework across the EU member countries and the UK (a former EU country) and ii) a common global consumer product market with certain – mostly small – national differences.

Finally, OPLS-DA was applied to the same data, containing 168 dust samples, to further explore the relationships between samples from different countries and search for key country-specific contaminant signatures, i.e. contaminants enriched in the samples from a certain country. In the score plot (Fig. 3A), country-wise sample separation can be clearly seen in most cases: Portugal and Slovenia cluster close, as well as Netherlands and Italy. Samples from Czechia, the UK and Estonia are separated from all other countries, which indicate that they have specific compound profiles. In the loadings plot (Fig. 3B), it is seen that the UK samples are associated with high-molecular-weight PAHs, organophosphates (OPs), and glycols and polymers (less pronounced), and the Estonian samples with biogenic compounds, such as carboxylic acids and terpene-related chemicals (in the PCPs and flavors and food-related classes). In addition, several pesticides are located at the top (in the direction of the Italian, Slovenian, and Portuguese samples; Fig. 3B), particularly terbutryn, but also permethrin, other pyrethroids (cyfluthrin and fenvalerate), an herbicide diflufenican, and a fungicide dimethomorph. These observations can potentially be explained by i) higher rates of agricultural activities in countries with a warmer climate, ii) climatic influences on the use of mold- and algae-preventing biocides, iii) the timing of the sample collection, and iv) locations of the sampled households, such as rural vs. urban, e.g., Portuguese samples were collected mostly from household located in rural areas.

Further OPLS-DA interpretation using `PLSDA_Interpreter.py` revealed additional compound class–country and individual signature compound–country relationships (see section 3.3.2 for the latter). A heatmap of the compound class–country relationships is shown in Fig. 4. Only the 50 compounds that were most specific for each country (with the highest final score values, c.f. section 2.6) were considered. The heatmap confirms strong domination of PAHs in the UK samples and PCPs and cosmetics and flavors and food-related classes in the samples from Estonia. It is worth mentioning that PCPs and cosmetics is also a dominant class for Slovenia, the Netherlands, and Portugal, whereas PAHs and fused aromatics class is also prominent in samples from Italy. In addition, the Italian samples are rich in plasticizers and industrial chemicals (top class) and PCPs and cosmetics classes. Finally, samples from Czechia are dominated by fatty acids and FAMES, followed by flavors and food-related compounds.

3.3. Concentrations of detected contaminants

3.3.1. Total contaminant class concentrations

The total contaminant class concentrations (based on the median value for each class; excluding values lower than the limit of detection (< LOD)) are presented in Fig. 5. As can be seen, the total concentration difference between the country with the highest contaminant burden (Slovenia, ca. 1750 µg/g) and the country with the lowest one (Italy, ca. 1100 µg/g) is close to a factor of 1.6. Moreover, low sample weights were collected in some countries, in particular Italy (SI, Table S4), which may result in negative bias in total contaminant levels, as levels of some contaminants would be below LODs. Thus, overall, the total contaminant levels are very similar across the seven countries investigated.

This can be an outcome of i) the unified regulations and common product market across Europe, mentioned in section 3.2, and/or ii) combining several classes into one class, e.g., combining antioxidants and UV filters into one class could erase potential trend in higher use of the latter in southern European countries where higher quantity of sun protection products is used.

Nonetheless, samples from central European countries (the Netherlands and Czechia) as well as the two southern European countries (Portugal and Italy) show similar contaminant class distribution both pairwise and when compared altogether. It is worth noting that samples from all countries have similar levels of sterols and steroids. In addition, dust from all countries, except for Slovenia, have similar levels of PCP chemicals. Additionally, samples from the UK and Slovenia have high total levels of phthalates, and samples from Czechia and Estonia have high levels of surfactants and solvents. Finally, the UK samples contain the highest levels of POPs and halogenated compounds, while the dust from Portugal is rich in pharmaceuticals and stimulants.

3.3.2. Country-specific signature compound concentration comparison

The average, median, and maximum concentrations of the (semi-)quantified tentatively identified compounds, as well as chemical identifiers and information on functional use/chemical classes, are presented in SI, Table S8.

House dust samples contain a wide variety of chemical classes spanning several orders of magnitude in concentration, ranging from PFAS (ca. 0.46 µg/g dust) and POPs (ca. 1.2 µg/g dust) to phthalates (ca. 360 µg/g dust) and PCP and cosmetic-related chemicals (ca. 600 µg/g dust).

Individual signature compounds, i.e. compounds enriched in the samples from a certain country but also potentially present in the samples from any other country at lower level, and their levels in the respective countries, as well as other countries, are shown in Fig. 6. The top 5 signature compounds per country were chosen based on the final score of the OPLS-DA data evaluation (section 3.2). The discussion is based on enrichment factors (EFs), ratios $C_{median}(Country) : C_{median}(All\ other\ countries)$, where $C_{median}(Country)$ is a median concentration of a signature compound in the samples from country for which this compound is a specific signature according to the OPLS-DA evaluation and $C_{median}(All\ other\ countries)$ is a median concentration of the same compound in the samples from all other countries.

The identified signature compounds showed varying degrees of enrichment, with EFs spanning from approximately 2-fold (dibutyl maleate (the Netherlands, NL) and etofenamate (Portugal, PT)) to more than 20,000-fold (branched isoundecyl phthalates (Slovenia, SI)). While some signature compounds reflected moderate concentration elevations

between the countries, others exhibited highly pronounced country-specific enrichments, suggesting localized or highly specific emission sources, e.g., branched isoundecyl phthalates in Slovenian samples and organophosphates (OPs) in the samples from the UK.

Among all countries, Slovenia exhibited the most extreme enrichment pattern, driven primarily by branched isoundecyl phthalates, which showed an exceptionally high EF (> 23,000). Such a strong enrichment likely reflects regional use of specific plasticizer-containing materials or products. Additional Slovenian signature compounds included other two phthalates (bis(2-hydroxyethyl) phthalate (EF 3.6) and hexyl neopentyl phthalate (EF 3.0)), further supporting the importance of plastic-associated emissions. Finally, benzocaine (pesticides and biocides class) and 2,3-dihydroxypropyl decanoate (PCPs and cosmetics) were the remaining signature compounds with EFs 4.9 and 3.6, respectively.

The UK demonstrated a distinctly OP-dominated signature profile, with three OPs, including tris(1-chloro-2-propyl) phosphate (TCIPP, EF 40) and tris(3-chloropropyl) phosphate (TCPP, EF 8.0), among the strongest national signature compounds. The elevated concentrations of these compounds are consistent with the historically extensive use of flame-retarded furniture and polymeric materials in the UK, partly associated with furniture flammability regulations (27). Elevated levels of OPs in UK dust align with earlier studies reporting comparatively high indoor concentrations of OPs in British residential environments (28). In addition to OPs, the presence of an industrial compound azacyclotridecan-2-one (EF 22) and fragrance-related frambinone (EF 12) suggests contributions from consumer products and polymer-associated sources.

Estonian signature compounds were dominated by flavor-related and PCP-associated compounds, including dl-isomenthone (EF 62), octyl octanoate (EF 16), *p*-menthan-3-one (EF 71), and benzyl acetate (EF 7.4). These substances are commonly associated with fragrances, essential oils, flavorings, and cosmetic formulations, suggesting that indoor dust in Estonian households may be more strongly influenced by volatile consumer product ingredients or terpene-derived compounds. Finally, industrial photoinitiator Irgacure 651 (2,2-dimethoxy-2-phenylacetophenone) showed EF 17.6.

Czechia exhibited a comparatively heterogeneous group of signature compounds composed of intermediates, food-related compounds, PCP-associated compounds, and OP-related chemicals. Diethyl 2,5-pyridinedicarboxylate showed particularly strong enrichment and achieved the highest overall score among Czech compounds with EF 24. Several Czech signature compounds, including ethyl oleate, methyl α -linolenate, and methyl laurate (EFs 8.0, 7.9, and 6.9, respectively), are lipid-related or naturally derived compounds that may originate from cooking activities, food residues, or cosmetic formulations. This observation supports the broader finding that indoor dust contains not only anthropogenic contaminants but also naturally derived organic compounds.

Portuguese signature compounds included pharmaceuticals, UV stabilizers, antioxidants, pesticides, and PCP-associated compounds. Isonox 132, a UV stabilizer, showed the strongest Portuguese enrichment (EF 6.1), followed by a PCP-related chemical *p*-tert-butylacetophenone (EF 5.5) and another UV stabilizer 2,4,6-tris(1-phenylethyl)phenol (EF 4.9). The occurrence of UV stabilizers may reflect increased use or degradation of polymeric materials exposed to higher solar radiation and temperature conditions in southern Europe. Similarly, the presence of terbutryn, a biocidal compound commonly used in coatings and treated materials, may indicate climatic influences on biocide use associated with mold prevention and material preservation. Finally, etofenamate, a nonsteroidal anti-inflammatory pharmaceutical ingredient, showed EF 2.3.

Italian signature compounds were primarily associated with phthalates, industrial additives, dyes, and polymer-related chemicals. The occurrence of multiple industrially associated compounds, including an additive styrene-

acrylonitrile trimer (4-(1-cyanoethyl)-1,2,3,4-tetrahydro-1-naphthalenecarbonitrile; EF 4.1) and a glycol ether hexaethylene glycol monomethyl ether, acetate (EF 4.5), may reflect emissions from treated materials, coatings, adhesives, or plastic products. The Italian profile overall appeared more chemically diverse and less dominated by a single compound class.

In contrast to the more strongly differentiated profiles observed in some countries, the Netherlands exhibited comparatively moderate enrichment factors overall, with EFs < 6. The Dutch signature compounds included dipropylene glycol *n*-butyl ether (EF 5.4), monoallyl phthalate (EF 4.1), a solvent *N*-methylbenzenamide (EF 4.7), and PCP-related ketoisophorone (EF 2.9) and dibutyl maleate (EF 2.3), suggesting contributions from coatings, cleaning products, industrial formulations, and consumer products. The relatively lower enrichment factors may indicate that these compounds are more widely distributed across Europe and therefore less geographically specific despite still contributing to the Dutch chemical fingerprint.

Overall, the identified signature compounds and the OPLS-DA results demonstrate that indoor dust composition across Europe reflects a combination of regional consumer habits and material usage patterns, climatic influences (i.e. warmer climate and more agricultural activities and pesticide use in the southern countries), and regulatory differences (i.e. OPs in the UK). The observed country-specific signatures were not driven exclusively by classical contaminants such as phthalates or OP-based flame retardants, but also included fragrances, naturally derived compounds, pharmaceuticals, food-related substances, and industrial intermediates. This highlights the value of comprehensive NTS approaches for capturing the full complexity of indoor chemical exposures. Moreover, the coexistence of anthropogenic and natural signature compounds emphasizes that indoor dust represents an integrated chemical archive of residents' behaviors and activities, product usage, and building material emissions.

3.3.3. Comparison with existing data

The concentrations (median) of contaminants detected were compared with the values previously reported by the NORMAN Network (6); see Fig. 7. It is worth mentioning that the NORMAN Network collaborative dust trial samples were collected in 2020, whereas samples for this study were collected during the years of 2023 and 2024, i.e. 3–4 years later.

The levels of phthalates in the current study are mostly in a similar range or slightly lower than in the NORMAN trial; only dimethyl phthalate (DMP) demonstrates ca. 7 times higher values. From the class of plasticizers, dyes, and industrial chemicals, only alternative plasticizers hexanedioic acid, bis(2-ethylhexyl) ester (DEHA) and tri(2-ethylhexyl) trimellitate (TOTM) demonstrate higher levels in this work, ca. 3 and ca. 7 times higher. These observations for both plasticizer classes can potentially be explained by phasing out phthalates and higher usage of the alternatives.

The levels of the PCPs and cosmetic products reported in this work are in a similar range or lower. The levels of most of the glycols and polymeric compounds are substantially lower in this study. The exceptions are nonaethylene glycol (PEG-9), whose median level is ca. 2 times higher, and, especially, tetraethylene glycol monododecyl ether (PEG4:C12), whose level is > 100 times higher in this study.

The levels of PAHs and fused aromatics are in a similar range in both studies, except for dibenz[*a,h*]anthracene (DBaA), which is ca. 45 times higher in the current work. DBaA is formed during incomplete combustion of organic matter, including fuels, which, in this case, can indicate increasing industrial emissions and engine exhaust. Another possible explanation can be the ongoing Russian invasion of Ukraine.

Surprisingly, the levels of POPs, like CB153, BDE47, and BDE99, are approximately 3 times higher in dust from this study. This can be explained by the presence of those in a low percentage of samples on one hand and higher concentrations in individual samples on the other hand. More samples were included in the dataset from this study ($n = 168$) as compared to the NORMAN study ($n = 75$), which increases the likelihood of including one or more high-concentration samples among the collected sample set. Also, different (semi-)quantification procedures (some of the data in the NORMAN studies were from targeted analyses) were used in the NORMAN study.

For all other classes, i.e. UV filters and antioxidants, pharmaceuticals and stimulants, intermediates and research chemicals, pesticides and biocides, surfactants and solvents, and organophosphates (OPs), the levels in the current study are lower than in the NORMAN trial.

4. Estimated risk from dust ingestion exposure

Exposure estimates were performed for all compounds, and hazard quotients (HQs) and cancer risks (CRs) were summarized for compounds detected in at least 10% of households (Table S9 and Table S10 in the SI).

HQs were below 1 for all compounds and samples with available data. The CL 1 compounds with the highest HQ values were for aniline, benzo[a]pyrene, hexachlorobenzene, dibutyl phthalate (DBP), and di(2-ethylhexyl) phthalate (DEHP), with DEHP and DBP HQs notably higher than other compounds (Fig. 8).

For the CLs 2–3 compounds (not shown in Fig. 8), the highest HQ values were observed for isononylphthalate (DINP) and isodecylphthalate (DIDP) mixtures, as well as for tris(1-chloro-2-propyl)phosphate (TCIPP), an organophosphate, and 4-(1-cyanoethyl)-1,2,3,4-tetrahydronaphthalene-1-carbonitrile, an industrial chemical used in the production of styrene plastics (29) (SI, Table S9). The median estimated risk for DINP is similar to that of DEHP and the maximum estimated risk for DINP, was notably high, with an HQ of 12, exceeding that of DEHP, suggesting excess risk from phthalate exposure in a small number of households.

Although no median HQs exceeded 1, it should be noted that these estimates consider only dust ingestion, and with the inclusion of dermal and air inhalation pathways, it is possible that exposure to these compounds may reach a concerning level. The HQs are also derived substance by substance, not considering potential dose-addition for substances with similar mode of action. In addition, 120 of the substances included in the risk estimation lacked any hazard reference value. While risk has hence not been evaluated for these, it can be noted that for 28 of the 120, the harmonized or industry notified classifications indicate concern for systemic health effects after long-term exposure.

Estimated cancer risks (CRs) exceeded 10^{-6} for seven compounds in a small subset of households (SI, Table S10): 4-aminobiphenyl (4-ABP), 2-acetylaminofluorene (2-AAF), 3-methylcholanthrene (MCA), dibenz[a,h]anthracene (DBahA), methyl yellow, 7,12-dimethylbenz[a]anthracene (DMBA), and benzidine (Fig. 9). Nine households had estimated CRs exceeding 10^{-6} for multiple compounds, with a maximum of 4 compounds in one Czech home. All are documented carcinogens and two source categories dominated: dyes/dye intermediates, and formation of polycyclic aromatic compounds from incomplete combustion processes. For most compounds, exceedance of the cancer threshold was limited to 1 or 2 homes, up to a maximum of 22 homes for benzidine, used in the production of dyes. It should be noted that this is a conservative estimate as we assumed that the compounds are fully bioaccessible from the ingested dust.

5. Conclusions

In this study, we present comprehensive data on chemical contaminants present in indoor dust across Europe (168 households from seven countries). This extensive dataset covers over 2300 anthropogenic and natural compounds, especially those that have found their use in industry and/or consumer products. Furthermore, this work provides valuable retention information (retention indices) for identified compounds that can help facilitate compound identification in other studies.

Dust contains a huge variety of chemical classes at various levels: from PFAS (ca. 0.46 µg/g dust) and POPs (ca. 1.2 µg/g dust) to phthalates (ca. 360 µg/g dust) and PCPs and cosmetics chemicals (ca. 600 µg/g dust). The total chemical loads vary per country from ca. 1100 µg/g in dust from Italy to ca. 1750 µg/g in dust from Slovenia. The levels of most chemicals are of similar magnitude to those in a previous European study.

The detected compounds were also evaluated (when possible) in terms of potential health-related risks. Median levels did not exceed risk thresholds for any compounds; however, a challenge in fully characterizing the risk of dust ingestion exposure is the lack of reference values for most compounds detected in dust, and the lack of a framework to address the dust chemical mixture as a whole.

In addition, this work presents a useful comprehensive analytical workflow for GC×GC-MS-based non-target screening of dust samples. The data processing part consists of vendor/commercial software and bespoke scripts and is designed to align large sample sets, detect chemical compounds using several algorithms for a fuller coverage, quantify detected chemicals, and retrieve various important chemical data – from chemical identifiers to functional use information – from several online databases.

Finally, the outcomes of this work are expected to extend beyond chemical characterization by helping prioritize compounds or classes driving observed biological effects. Within the INQUIRE framework, these findings will be combined with effect-based methods, including *in-vitro* bioassays, vEDA, and QSAR modeling, to clarify exposure–effect relationships. This integrated approach aims to inform targeted mitigation strategies and ultimately improve indoor air quality and human health in residential settings.

Declarations

Competing interests

The authors declare no competing interests.

Funding

This work was funded by the European Union under grant agreement Project No. 101057499 – INQUIRE. This work was also co-funded by UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee [grant numbers 10038689 and 10042425] and the Australian National Health and Medical Research Council (NHMRC) [NHMRC; 2022/GNT2017837]. Open access funding was provided by Umeå University.

Ethics approval

Ethical approval for this study was granted by the local ethical committee at each sample collection location (CZ: Ethics committee of (C)ELSPAC, Ref. No.:(C)ELSPAC/EK/3/2023; EE: Tartu Ülikooli inimüuringute eetika komitee;

Ref. No.: 380/T-1; IT: Comitato Etico Nazionale, Ref. No.: 0020314; NL: BETHCIE, Research Ethics Review Committee, Faculty of Science, Ref. No.: 23-008; PT: Ethics committee of the University of Porto, Ref. No.: 135/CEUP/202; SI: Komisija Republike Slovenije za medicinsko etiko, Ref. No.: 0120-398/2023/3; SE: Etikprövningsmyndigheten, Ref. No.: 2024-01249-01-581358; UK: Yorkshire & The Humber - Bradford Leeds Research Ethics Committee, Ref. No.: 322462), and informed consent was obtained from a representative member of each household.

Data availability

The data underlying this article cannot be shared publicly due to ethical restrictions. Data may be made available upon reasonable request to the corresponding author, subject to approval by the appropriate ethics committees and in accordance with institutional and legal data protection requirements.

Acknowledgements

The authors would like to thank the rest of the INQUIRE team and all the study participants for their invaluable contributions to this research. The views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Health and Digital Executive Agency (HaDEA). Neither the European Union nor the granting authorities can be held responsible for them.

The authors would also like to thank LECO Corp. for providing a trial license for the ChromaTOF Sync 2D software (version 2.0.15.0). Additionally, the authors would like to thank Kristina Stetsiv for the visual artwork.

OpenAI ChatGPT and Microsoft Copilot were used to aid development of the scripts used in this work.

Author contributions

Andriy Rebryk: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Sample Analysis, Data Analysis, Data Curation, Writing – Original Draft, Writing – Review and Editing, Supervision, Visualization.

Peter Haglund: Conceptualization, Methodology, Validation, Resources, Data Analysis, Data Curation, Writing – Review and Editing, Supervision, Project Administration, Funding Acquisition.

Lisa Melymuk: Conceptualization, Methodology, Data Analysis, Visualization, Writing – Review and Editing.

Linda Schenk: Data Analysis, Visualization, Writing – Review and Editing, Funding Acquisition.

Eva Sugeng: Conceptualization, Investigation, Writing – Review and Editing, Funding Acquisition.

Pim Leonards: Conceptualization, Investigation, Writing – Review and Editing, Funding Acquisition.

Marta Gabriel: Conceptualization, Investigation, Writing – Review and Editing, Funding Acquisition.

Anna Maria Ingelido: Investigation, Writing – Review and Editing, Funding Acquisition.

Kristina Aidla: Project Administration, Investigation, Writing – Review and Editing.

Stamatina Violaki: Investigation, Writing – Review and Editing.

Joana Azeredo: Investigation, Writing – Review and Editing.

Rok Novak: Investigation, Writing – Review and Editing.

Neža Palir: Investigation, Writing – Review and Editing.

Valentina Marra: Investigation, Writing – Review and Editing.

Marco Inglessis: Investigation, Writing – Review and Editing.

Anna-Sofia Preece: Investigation, Writing – Review and Editing.

Konstantin Kislitsõn: Investigation, Writing – Review and Editing.

Riin Tarto: Investigation, Writing – Review and Editing.

Hans Orru: Resources, Investigation, Writing – Review and Editing.

Ene Indermitte: Resources, Investigation, Writing – Review and Editing.

References

1. WHO. Optimizing brain health across the life course: WHO position paper: World Health Organization; 2022. <https://iris.who.int/bitstream/handle/10665/361251/9789240054561-eng.pdf>.
2. Schweizer C, Edwards RD, Bayer-Oglesby L, Gauderman WJ, Ilacqua V, Juhani Jantunen M, et al. Indoor time–microenvironment–activity patterns in seven regions of Europe. *Journal of Exposure Science & Environmental Epidemiology*. 2007;17(2):170–81.
3. Mannan M, Al-Ghamdi SG. Indoor Air Quality in Buildings: A Comprehensive Review on the Factors Influencing Air Pollution in Residential and Commercial Structure. *International Journal of Environmental Research and Public Health*. 2021;18(6):3276.
4. Moschet C, Anumol T, Lew BM, Bennett DH, Young TM. Household dust as a repository of chemical accumulation: new insights from a comprehensive high-resolution mass spectrometric study. *Environmental Science & Technology*. 2018;52(5):2878–87.
5. Rostkowski P, Haglund P, Aalizadeh R, Alygizakis N, Thomaidis N, Arandes JB, et al. The strength in numbers: comprehensive characterization of house dust using complementary mass spectrometric techniques. *Analytical and Bioanalytical Chemistry*. 2019;411(10):1957–77.
6. Haglund P, Alygizakis NA, Covaci A, Melymuk L, Nizzetto PB, Rostkowski P, et al. Comprehensive characterization of European house dust contaminants: Concentrations and profiles, geographical variability, and implications for chemical regulation and health risk. *Sci Total Environ*. 2024;957:177639.
7. Sugeng EJ, Leonards PE, van de Bor M. Brominated and organophosphorus flame retardants in body wipes and house dust, and an estimation of house dust hand-loadings in Dutch toddlers. *Environmental Research*. 2017;158:789–97.
8. EPA. Exposure Factors Handbook: 2011 Edition (Final). Washington, DC: U.S. EPA; 2011. <https://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=236252>.

9. Brandsma SH, de Boer J, van Velzen MJ, Leonards PEJ. Organophosphorus flame retardants (PFRs) and plasticizers in house and car dust and the influence of electronic equipment. *Chemosphere*. 2014;116:3–9.
10. Mitro SD, Dodson RE, Singla V, Adamkiewicz G, Elmi AF, Tilly MK, et al. Consumer product chemicals in indoor dust: a quantitative meta-analysis of US studies. *Environmental Science & Technology*. 2016;50(19):10661–72.
11. Rudel RA, Camann DE, Spengler JD, Korn LR, Brody JGJ. Phthalates, alkylphenols, pesticides, polybrominated diphenyl ethers, and other endocrine-disrupting compounds in indoor air and dust. *Environmental Science & Technology*. 2003;37(20):4543–53.
12. Zhu L, Fauser P, Mikkelsen L, Sanderson H, Vorkamp K. Suspect and non-target screening of semi-volatile emerging contaminants in indoor dust from Danish kindergartens. *Chemosphere*. 2023;345:140451.
13. Dubocq F, Wang TJ. Extended characterization of the indoor dust chemical composition by retrospective suspect and non-target analysis of high resolution mass spectrometric data. *Journal of Environmental Exposure Assessment*. 2022;1(4).
14. INQUIRE. Improving indoor air quality for a healthier home and Europe: INQUIRE; 2022. <https://inquire-he.eu/>.
15. Mondello L, Tranchida PQ, Dugo P, Dugo G. Comprehensive two-dimensional gas chromatography-mass spectrometry: A review. *Mass Spectrometry Reviews*. 2008;27(2):101–24.
16. Rebryk A, van Herwerden D, Samanipour S, Haglund P. Evaluation of peak alignment performance of commercial and open-source tools for GC×GC–MS based non-targeted screening. *Briefings In Bioinformatics*. 2026;27(5):bbag458.
17. Röhler L, Bohlin-Nizzetto P, Rostkowski P, Kallenborn R, Schlabach M, Discussions P. Non-target and suspect characterisation of organic contaminants in ambient air, Part I: Combining a novel sample clean-up method with comprehensive two-dimensional gas chromatography. *Atmospheric Chemistry and Physics*. 2020;2020:1–33.
18. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, et al. PubChem 2023 update. *Nucleic acids research*. 2023;51(D1):D1373–d80.
19. ClassyFireBatch. ClassyFire Batch by Fiehn Lab: Fiehn Lab; 2019. <https://cfb.fiehnlab.ucdavis.edu/#/>.
20. Phillips KA, Wambaugh JF, Grulke CM, Dionisio KL, Isaacs KJ. High-throughput screening of chemicals as functional substitutes using structure-based classification models. *Green Chemistry*. 2017;19(4):1063–74.
21. Haglund P, Schymanski E. S120 | DUSTCT2024 | Substances from Second NORMAN Collaborative Dust Trial (NORMAN-SLE-S120.0.1.0). Zenodo2024. <https://www.doi.org/10.5281/zenodo.13835255>.
22. EPA. ChemExpo - Get Data: United States Environmental Protection Agency; 2025. https://comptox.epa.gov/chemexpo/get_data/.
23. ATSDR. Calculating Hazard Quotients and Cancer Risk Estimates 2025. https://www.atsdr.cdc.gov/pha-guidance/conducting_scientific_evaluations/epcs_and_exposure_calculations/hazardquotients_cancerrisk.html.
24. Pinto-Vidal FA, Novák J, Jílková SR, Rusina T, Vrana B, Melymuk L, Hilscherová K. Endocrine disrupting potential of total and bioaccessible extracts of dust from seven different types of indoor environment. *Journal of Hazardous Materials*. 2024;469:133778.
25. Juhasz AL, Keith A, Jones R, Kastury F. Impact of precursors and bioaccessibility on childhood PFAS exposure from house dust. *Science of the Total Environment*. 2023;889:164306.
26. He R, Li Y, Xiang P, Li C, Zhou C, Zhang S, et al. Organophosphorus flame retardants and phthalate esters in indoor dust from different microenvironments: bioaccessibility and risk assessment. *Chemosphere*. 2016;150:528–35.
27. Abdallah MA-E, Harrad S. Dermal uptake of chlorinated organophosphate flame retardants via contact with furniture fabrics; implications for human exposure. *Environmental Research*. 2022;209:112847.

28. Brommer S, Harrad S. Sources and human exposure implications of concentrations of organophosphate flame retardants in dust from UK cars, classrooms, living rooms, and offices. *Environment International*. 2015;83:202–7.

29. National Toxicology Program. Toxicology and carcinogenesis study of styrene-acrylonitrile trimer in F344/N rats (perinatal and postnatal feed studies). National Toxicology Program technical report series. 2012(573):1-155. <https://ntp.niehs.nih.gov/publications/reports/tr/tr573>.

Figures

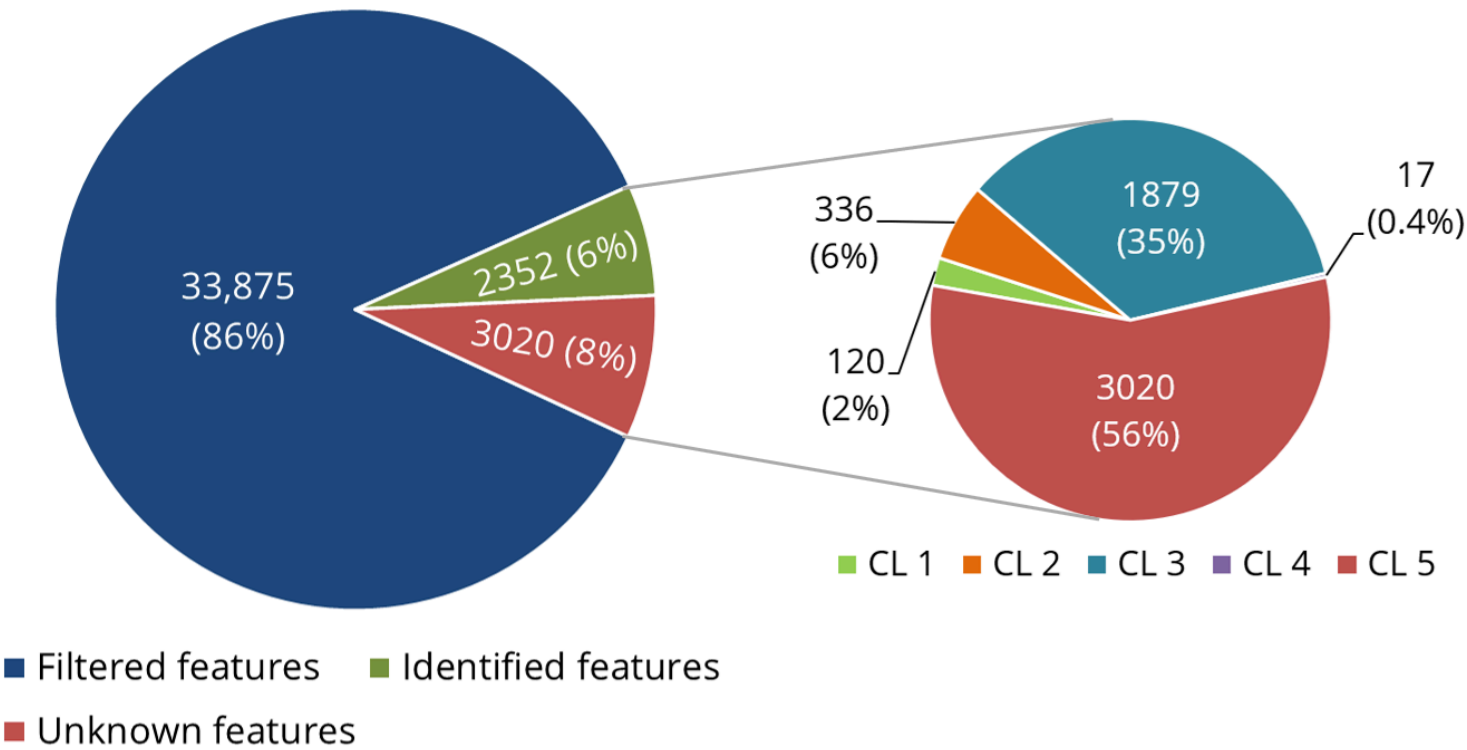


Figure 1

Legend not included with this version

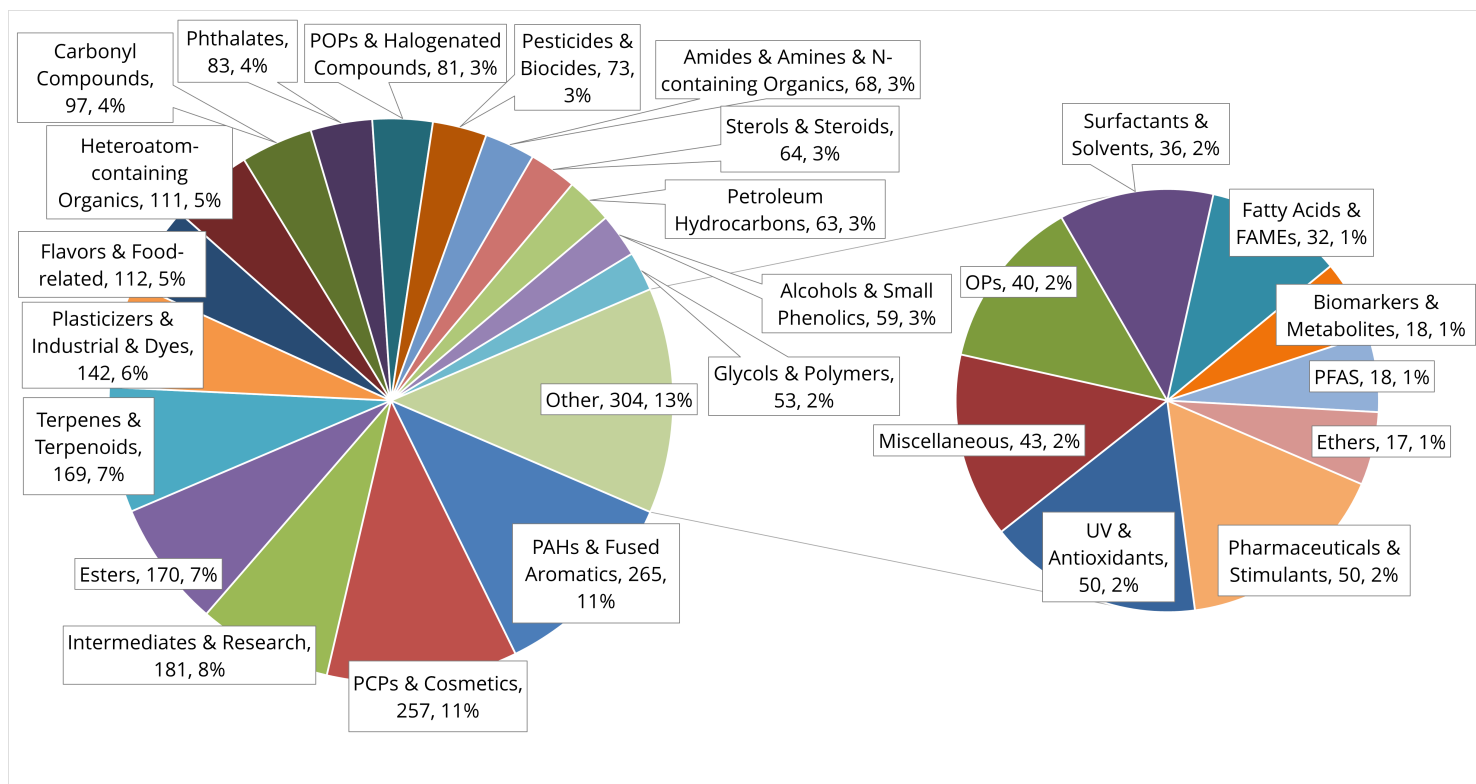


Figure 2

Legend not included with this version

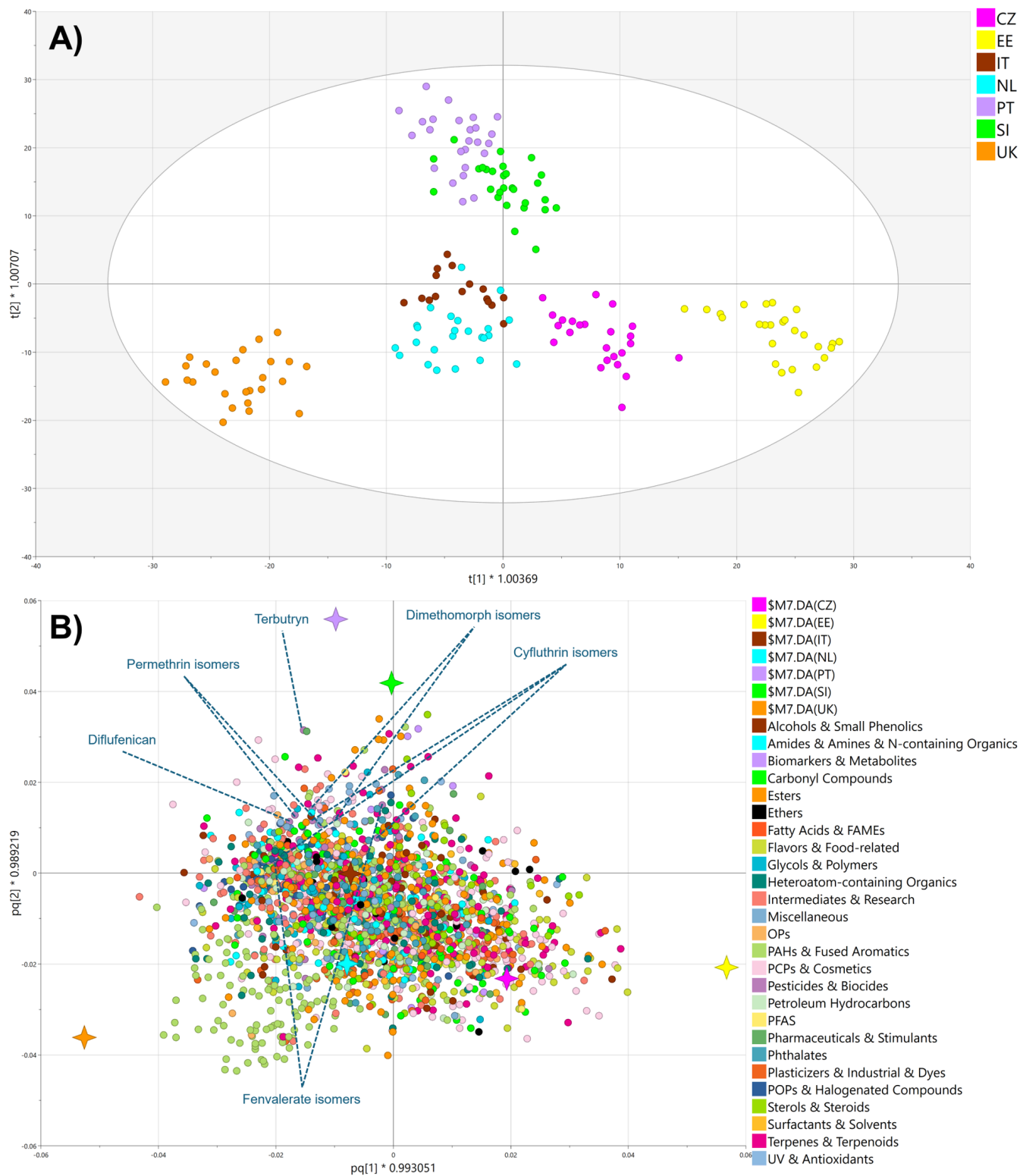


Figure 3

Legend not included with this version

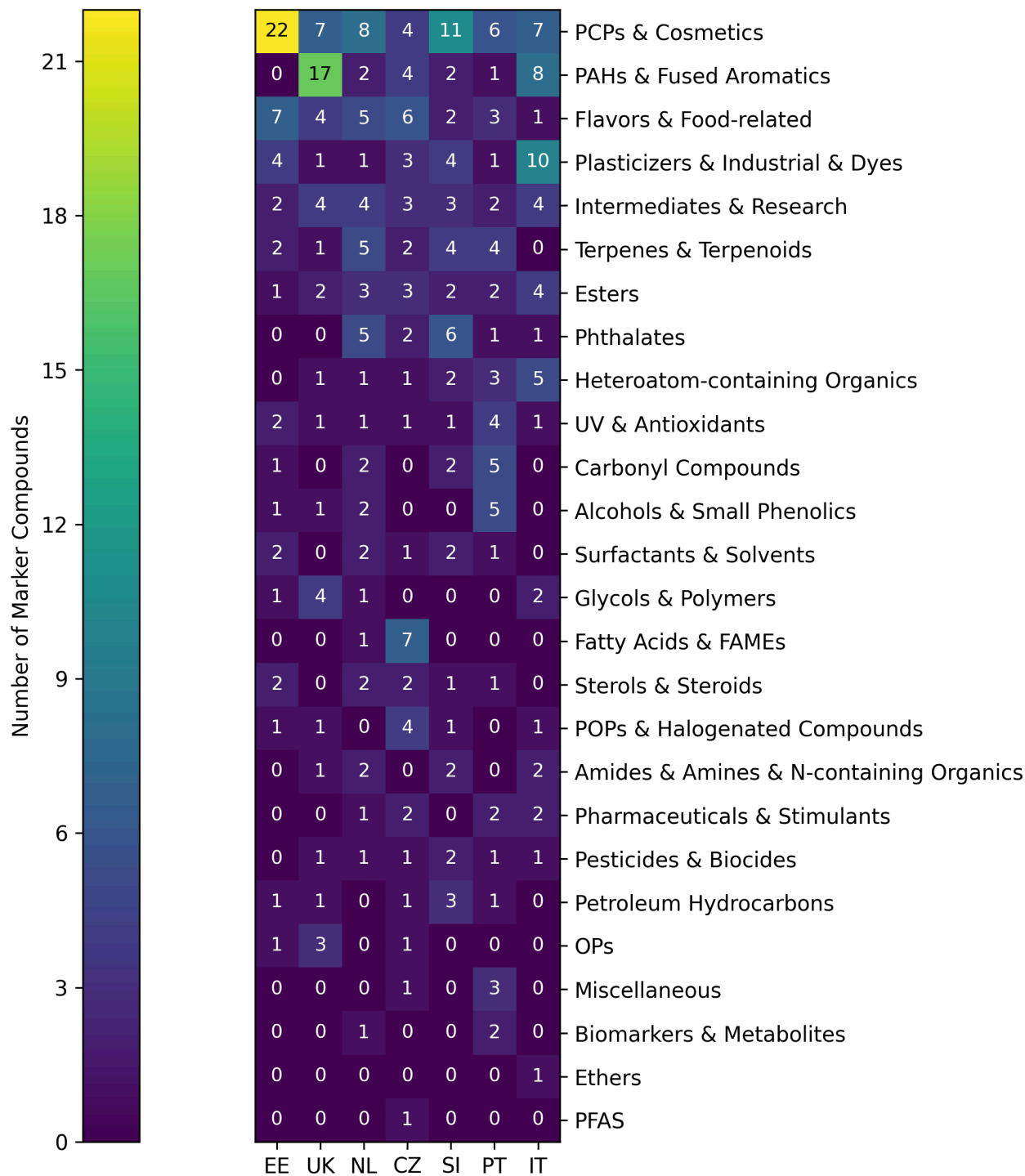


Figure 4

Legend not included with this version

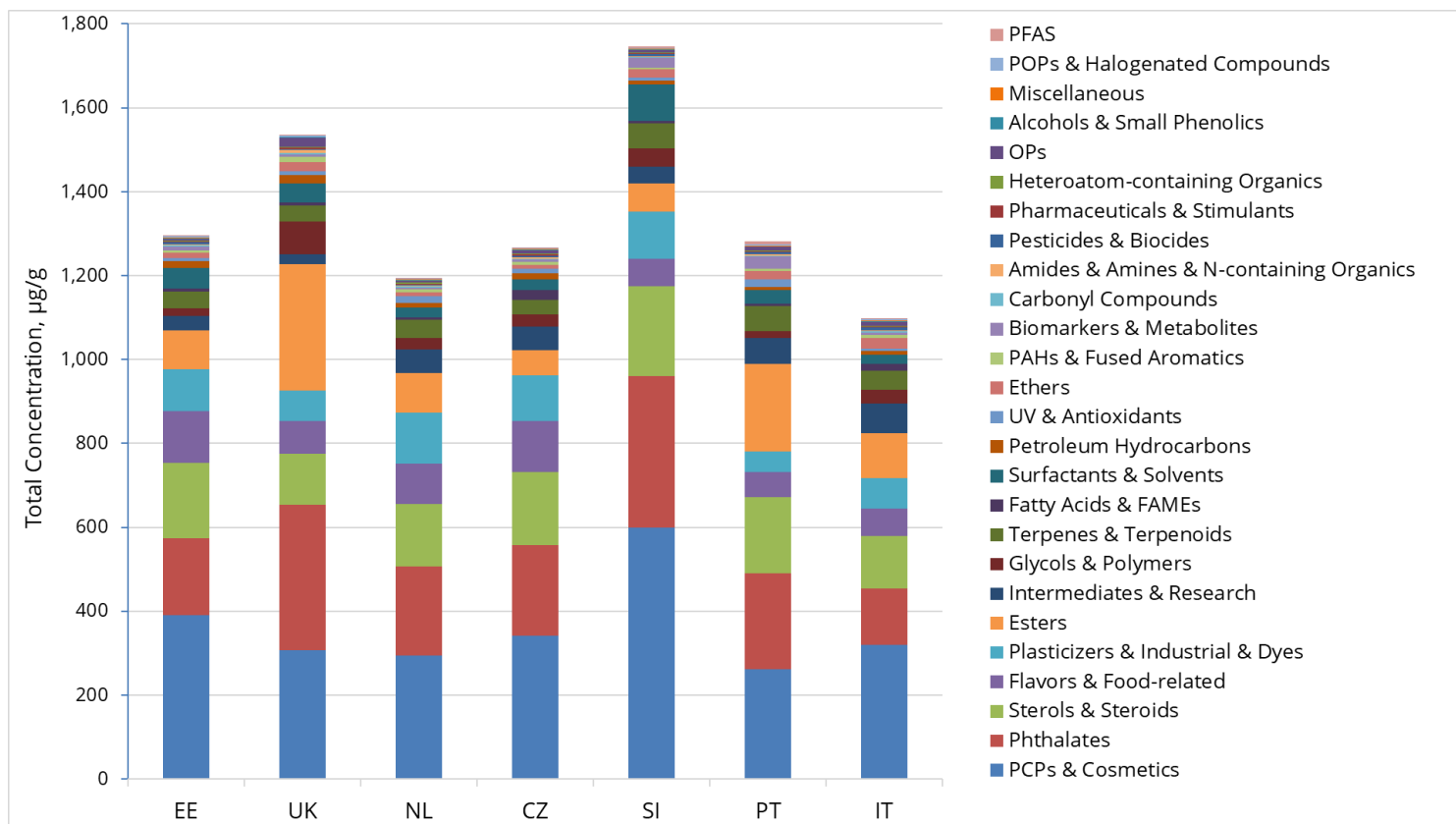


Figure 5

Legend not included with this version

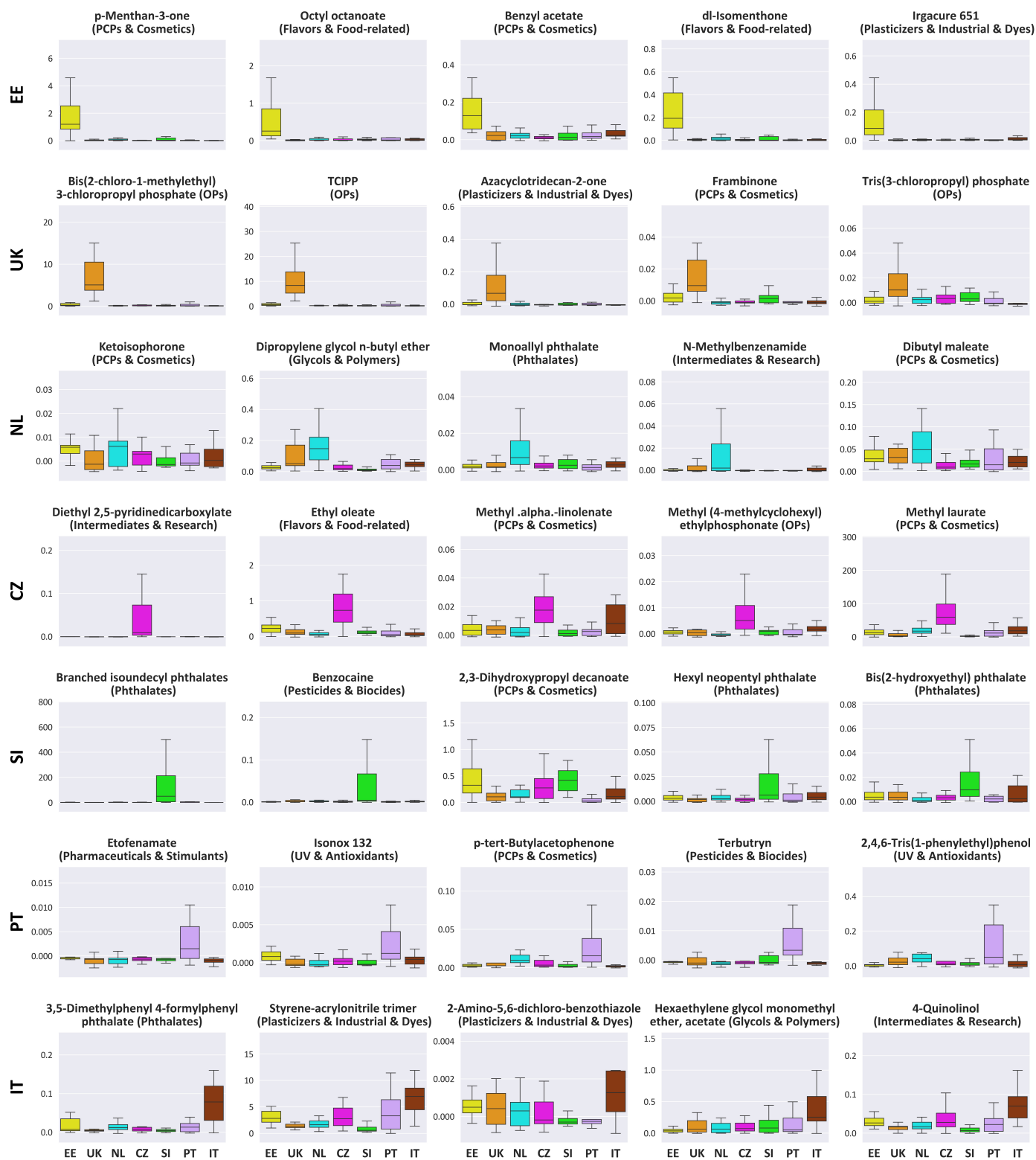


Figure 6

Legend not included with this version

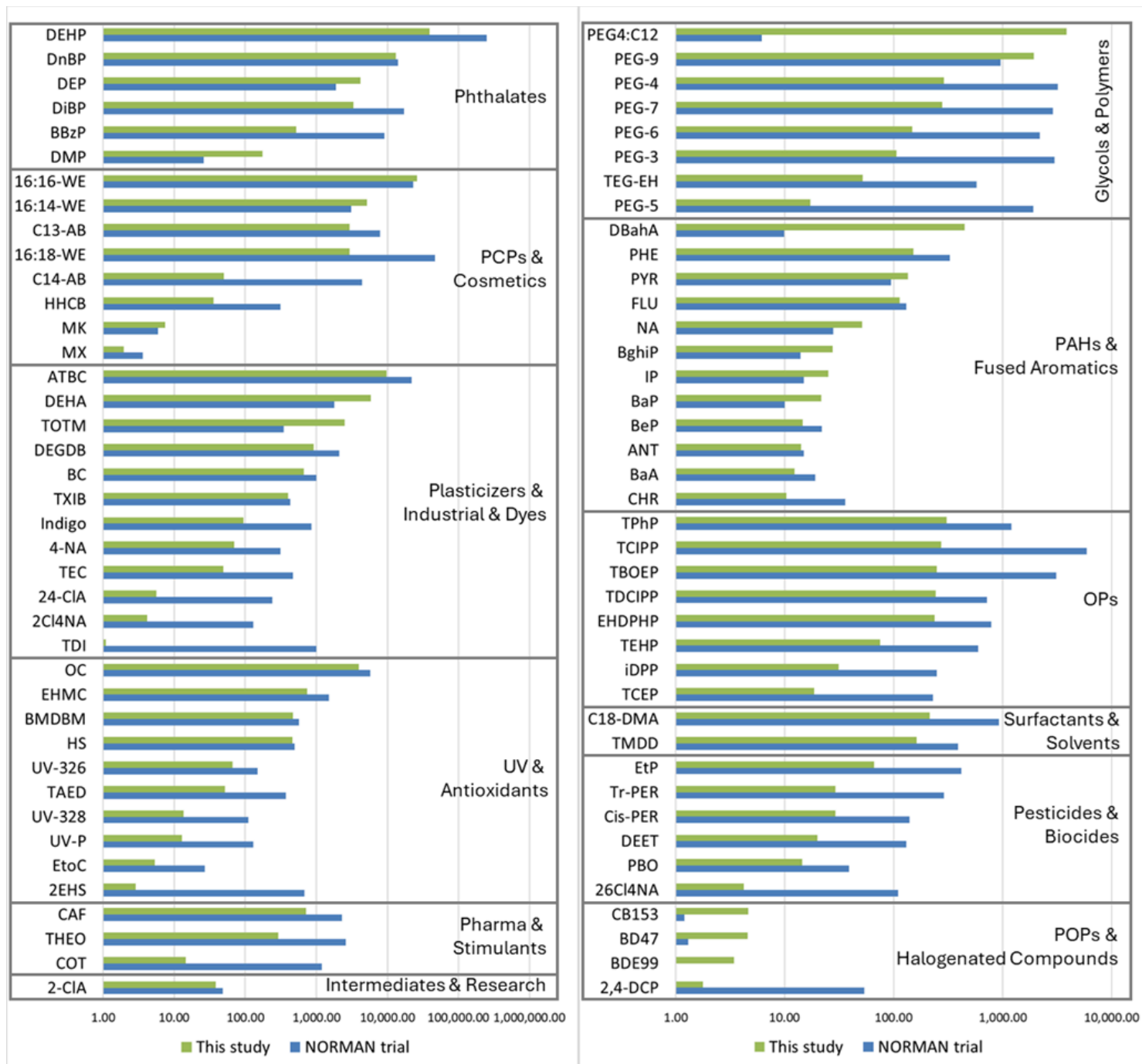


Figure 7

Legend not included with this version

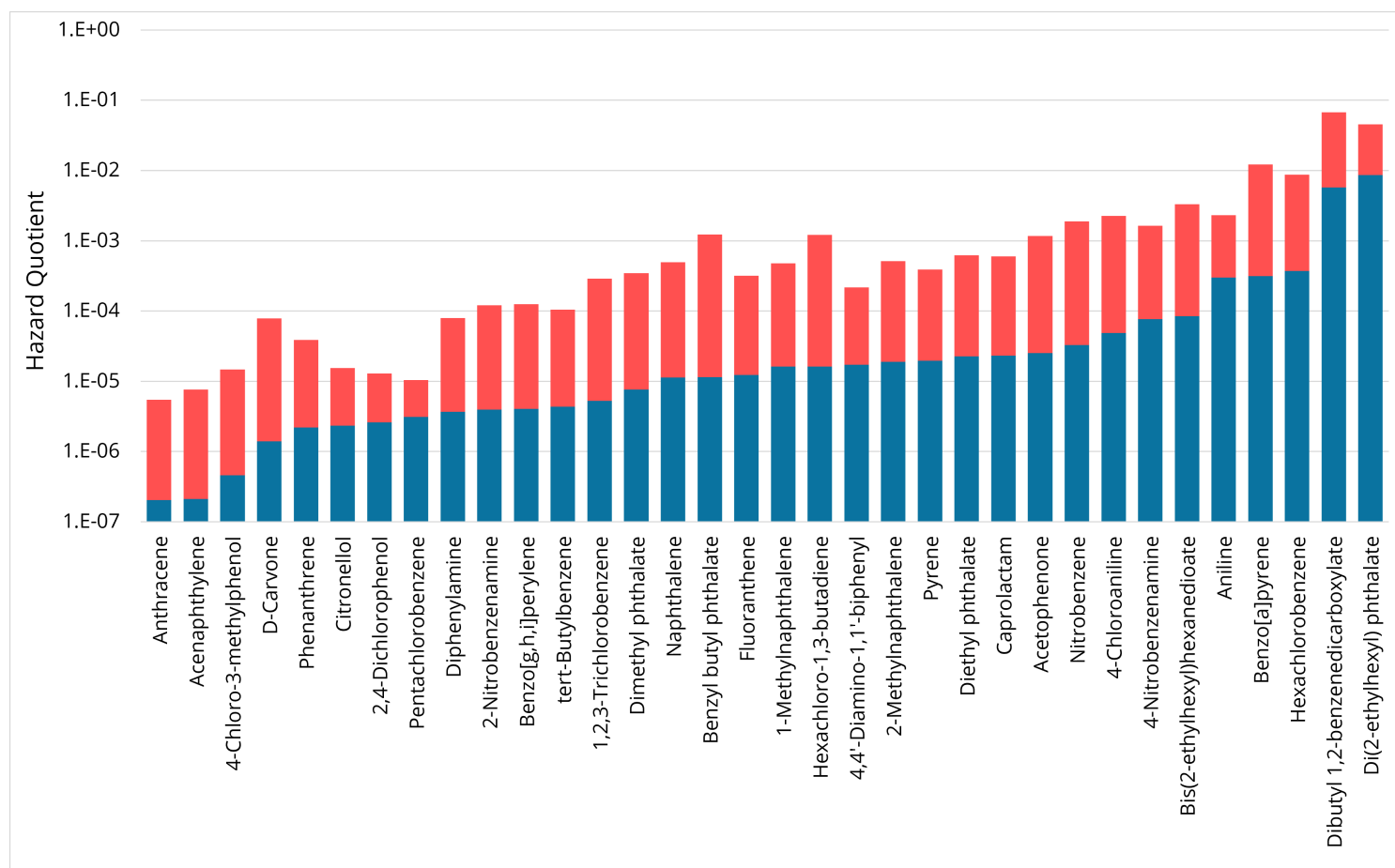


Figure 8

Legend not included with this version

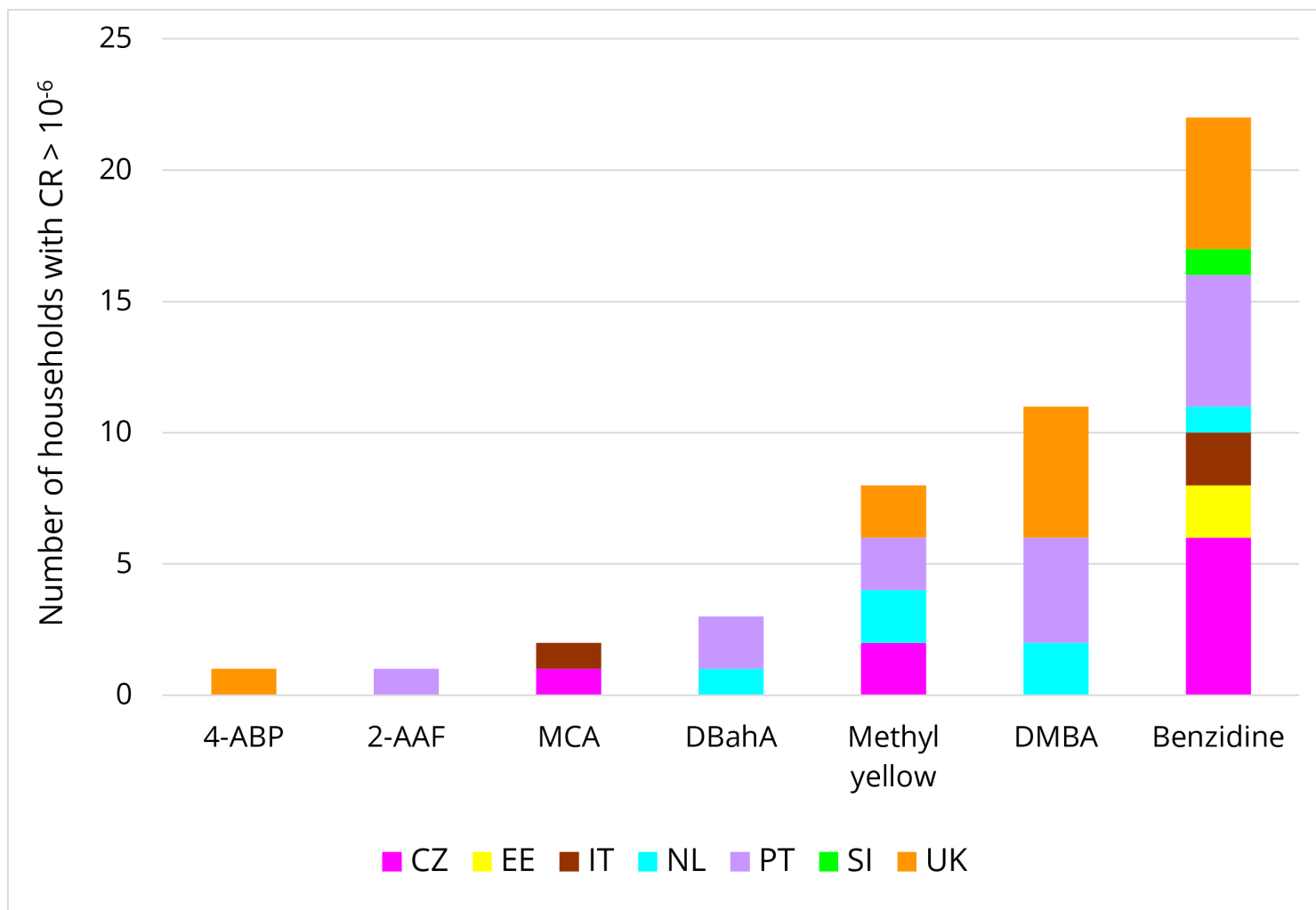


Figure 9

Legend not included with this version

Supplementary Files

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

- [260905INQUIREGCxGCDustPaperSITables.xlsx](#)
- [260905INQUIREGCxGCDustPaperSI.docx](#)