Supplemental Information for Suspect Screening Analysis of Pooled Human Serum Samples Using GC×GC/TOF-MS

Katherine A. Phillips[1], Alex Chao[1], Rebecca L. Church[2], Kristin Favela[3], Stavros Garantziotis[2], Kristin K. Isaacs[1], Brian Meyer[1,4], Annette Rice[2], Risa Sayre[1], Barbara A. Wetmore[1], Alice Yau[3], and John F. Wambaugh[1,*]

[1] U.S. Environmental Protection Agency, Office of Research and Development, Center for Computational Toxicology and Exposure, Research Triangle Park, NC 27711
[2] U.S. Department of Health and Human Services, National Institutes of Health, National Institute of Environmental Health Sciences, Clinical Research Unit, Durham, NC 27709
[3] Southwest Research Institute, San Antonio, TX 78238
[4] Deceased April 2023

[*] Correspondence: John F. Wambaugh <wambaugh.john@epa.gov>

Supplemental Information consists of this document (17 Pages and 6 Supplemental Figures) plus 6 Supplemental Tables in separate files.

Supplemental Methods

Study Design

Sample Information and Preparation

Pooled serum samples were obtained from the NIEHS Clinical Research Unit specimen bank. The specimen bank was established to provide a repository of biological specimens from individuals as well as basic demographic information for use by researchers developing or testing assays or other technologies in the environmental health arena. The specimen bank includes blood samples from approximately 3500 adults who have been recruited to participate and donate biological samples. The specimen bank does not constitute a statistically representative sampling of any population but has instead been assembled as a convenience sample. Thus, inferences based on evaluation and results of the analysis of pooled samples from this specimen bank will necessarily be limited. However, the selection of priority analytes has been designed to allow evaluation of the results in the context of NHANES datasets that are representative of the population for selected analytes.

A Materials Transfer Agreement (MTA, for transfer of the blood samples from NIEHS to EPA) was implemented and signed in October 2015 and the required ethics evaluation was completed. Because the project involves analysis of pooled, de-identified serum samples, the project was judged by an EPA Senior Ethics Official not to constitute new human subjects research and therefore was not subject to EPA IRB review or approval.

A variety of criteria and approaches for constructing pools for this pilot project were considered. Children are a population of special interest; however, the NIEHS specimen bank includes only adult participants. Evaluation of existing biomonitoring data from NHANES shows that differences in biomarker concentrations among various demographic groups (by age and gender) are generally relatively small in the context of the uncertainties inherent in the high-throughput testing and modeling strategies considered by EPA's ToxCast and ExpoCast projects. Geometric mean concentrations by age groups almost always

vary by less than a factor of 10 and most often by less than a factor of three[1]. Thus, to streamline the request process and limit the amount of serum taken from the NIEHS specimen bank, only limited stratification criteria were applied in the construction of samples.

As both the National Health and Nutrition Examination Survey (NHANES) in the U.S., the Human Biomonitoring for the European Union (HB4EU) and other prominent population-based biomonitoring studies present their results by stratifying with age and sex, this study has sought to have pooled samples that perform a similar, yet simplified stratification. By looking at both age- and sex-specific results changes in chemical exposures can identified in these sub-populations. Further, age-based trends in exposure have been highlighted repeatedly in scientific literature. While the types of samples available in the NIEHS specimen bank and the number of pools that could be run in this pilot study were limited, stratification in age did occur at 45 years.

Sixteen pooled samples were constructed, four each from four demographic categories: females ages 45 and under, females ages above 45, males ages 45 and under, and males ages above 45. Each pool was constructed from 25 individual one milliliter (ml) samples, selected randomly from the available specimens in that age and sex category. In addition to these sixteen pooled samples, a National Institute of Standards and Technology (NIST) standard reference material (SRM) for human blood was also included in the study (NIST SRM 1958: Organic Contaminants in Fortified Human Serum). The SRM was reconstituted with 10.7 ml of water.

One milliliter (mL) of each of the 16 serum samples and the SRM sample was spiked with 2μL of a 100 μ L solution of the surrogate $^{13}C_{12}$-4,4'-DDT (Cambridge Isotope Laboratories, Inc., Item CLM-1281-5, Lot# SCGH-009) and extracted with 4.0 mL of methyl tertiary butyl ether (MTBE) under both acidic and basic conditions. To 1 mL of sample, 4 mL of MTBE was added, and the pH was adjusted to pH 2 with 1:1 water:sulfuric acid added dropwise (2 drops required). After sonication for 5 minutes, the MTBE layer was removed, and the sample was re-extracted by adding 4mL of MTBE then adjusting to pH 12 with 10M NaOH added dropwise (4 drops required). Following further sonication, the layers were

combined and nitrogen-evaporated to a final volume of 0.1 mL. An initial analysis attempt of the extract directly by GC×GC resulted in severe column overloading in some areas and carry-over of heavy (presumably biological) material to the next run. Therefore, all sample extracts were diluted 5× in dichloromethane (DCM) and three DCM blanks were assayed after each sample and replicate using an accelerated oven ramp.

QC Spikes and Samples

To monitor any interferences or background levels which may have been introduced at the time of extraction, a method blank was generated concurrently with the serum samples. There are two method blanks, one corresponding to each of two extraction batches. The corresponding extraction batch method blank was assayed undiluted with each analytical sequence. These are designated SDB-1, SDB-2, SDB-3, etc. In addition, due to carry-over of heavy material that was observed to build up on the column during the first attempt to assay the samples, three DCM blanks were assayed after each sample and replicate using an accelerated oven ramp.

A solution containing multiple deuterated internal standards were spiked into the calibration standards and sample extracts at a concentration of 1.0 ppm prior to analysis. The internal standards include Naphthalene-d8 (DTXSID10894058), 1,4-Dichlorobenzene-d4 (DTXSID30959416), Acenaphthene-d10 (DTXSID40893473), Phenathrene-d10 (DTXSID60893475), Perylene-d12 (DTXSID60934397), and Chrysene-d12 (DTXSID00893474).

A multi-point calibration curve taken at 10, 25, 50, 100, 250, 500, 1000, and 2000 part per billion (ppb) was assayed with each analytical batch. The wide range of this curve was used due to anticipated varied performance of individual targets in samples. The cocktail contained the following fifteen chemicals: methyleugenol, dimethyl phthalate, diethyl phthalate, simazine, metribuzin, dibutyl phthalate, triclosan, fenamiphos, bisphenol A, carboxin, dieldrin, endrin, o,p’-DDT, mirex and di-n-octyl phthalate. In the case of six chemicals (simazine, metribuzin, triclosan, fenamiphos, BPA, dieldrin), the low-level standard (10 ppb) was inconsistently detected due to being near the instrument detection limit.

Data Acquisition

Chromatography

All samples were analyzed by both low- and high-resolution GC×GC/TOF-MS. High-resolution was used for confirmation analysis, while low-resolution was use for suspect screening. Low-resolution GC×GC/TOF-MS analysis was performed using an Agilent 7890 gas chromatograph coupled to a LECO PEGASUS 4D -TOF (LECO, St. Joseph, MI). Injection volume was 1.0 µL. The inlet temperature was 275 °C and the inlet mode was splitless with a 1-min purge. Separation was achieved using two columns. The primary column (1st Dimension) was an RXi-1MS (30 m × 0.25 mm × 0.25 µm; Restek, Bellefonte, PA) and the second column (2nd Dimension) was an RXi-17SilMS, (1.3 m × 0.18 mm × 0.18 µm; Restek, Bellefonte, PA). The first column was held at 45 °C for three minutes, ramped to 330 °C at a rate of 8 °C/min, and held for five minutes. The second column and modulator were offset by 5 °C and 20 °C, respectively, relative to the primary oven. Helium carrier flow was set to constant flow at 1.0 mL/min. The transfer line temperature was set at 300 °C. The modulation period was five seconds (1.25 s hot, 1.25 s cold with 2 cycles per modulation period).

High-resolution GC×GC/TOF-MS analyses was performed using and Agilent 7890 gas chromatograph coupled to a LECO HRT GC×GC (LECO, St. Joseph, MI). Conditions were the same as for the low-resolution analysis, except the second column was an RXi-17SilMS (1.3m × 0.25 mm × 0.25 µm; Restek, Bellefonte, PA), helium carrier flow was 1.2 ml/min and the modulation period was 4 seconds (1.0 s hot, 1.0 s cold with 2 cycles per modulation period).

While the follow-up, high-resolution analysis was used for confirmation analysis, comparisons were performed between chemicals of potential relevance identified with both. This was done to quantify how many chemicals of Level 1 and a subset of Level 2 identifications could be identified using either method (see Supplemental Information Data Processing section for more information on identification types). It should be noted that the low-resolution analysis was performed in 2017 while the high-resolution analysis was performed on re-extracted samples (using the same method as mentioned previously) in 2020.

Software variable settings used in LECO SSA routine were SN: 57, 1st dimension peak width: 12s, 2nd dimension peak width: 0.1s, and 500 match score required to combine slices with a 0.05s early offset allowed.

Mass Spectrometry

For both low- and high-resolution GC×GC/TOF-MS, the mass spectrometer was operated using electron ionization (EI) at 70eV. The ion source temperature was set at 225 °C. Spectra were collected from 45–650 m/z with a scan time of 100 spectra/sec. Sensitivity was verified according to the manufacturer's specifications on each system.

Quality Control/Quality Assurance Process for Analytical Sequence

One isotopically labeled internal standard was spiked into each sample and replicate before extraction and six isotopically labeled standards were spiked into all extracts just prior to analysis. A reference standard curve was assayed with each analytical sequence and verified at the end of the sequence. Due to the heavy matrix, three system blanks were injected between each sample. The internal standard response was monitored throughout the analysis and samples were re-injected if a drift in response was noted. The sequence was not randomized. However, there are 3 mitigating factors for the lack of sequence randomization:

1. During initial data processing, samples were first screened to identify compounds of potential relevance using a signal-to-noise value of 50. The samples were then re-processed with a signal-to-noise value of 5, which is 10-times lower than the initial screening. This affords a minimum dynamic range of ten between the most abundant sample and the noise floor of the instrument therefore mitigating slight variations in instrument sensitivity.

2. The average internal standard (naphthalene-d8) response was within ± 50% for all samples, exempting one sample which was slightly above that limit. All responses were normalized to the internal standard naphthalene-d8 to account for any drift. After normalization of the internal

standard spiked prior to extraction to naphthalene-d8, the percent relative standard deviation across all analyses was 31%.

3. Most signals were confirmed by re-extracting and re-analyzing the samples on a second instrument.

Samples were run according to the order shown in Supplemental Table 6. Performance was monitored for retention time drift, lack of sensitivity, carry-over, and other non-conforming results. If performance problems were detected throughout a sequence, maintenance and/or troubleshooting occurred and samples re-assayed until no issues were detected. In addition to monitoring external and internal reference standards throughout a sequence to confirm sensitivity, presence of 2pg of hexachlorobenzene was confirmed (per the manufacturer's specifications) on the column with an SN of 10 before each sequence. The injection sequence shows an initial attempt at analysis in which overloading and build-up of heavy material was noted (SI Table 6); various corrective actions such as dilution, blanks and clean-up procedures were tested before arriving at the utilized procedure of diluting each sample 5x and running three blanks after each sample replicate. Maintenance was performed before resuming data collection. Some samples were assayed multiple times, including some of the female ones. The final order of pools were Females greater than 45 years, Males less than or equal 45 years, Males greater than 45 years, and Females less than or equal to 45 years.

Data Processing and Analysis

Data Processing

Data were processed using LECO's Chromatof software to integrate peaks and quantify them against reference standards. For compounds which did not have reference standard, this software was used to match them against the National Institute of Standards and Technology 2017 Mass Spectral library (NIST 2017). Signals found in samples that were also present in solvent and/or dilution blanks were classified as background unless the peak area found in the sample exceeded the blank by at least 10 times. Chromatof

classification regions were used to define areas of the chromatogram where specific classes of compounds tend to elute (that is, hydrocarbons, long chain biologicals, phthalates, sterols, and vitamins)[2]. Classification regions were used to exclude areas of the chromatogram where overloaded biological peaks (such as organic acids, cholesterol, and cholesterol-related peaks) were detected.

After exclusion of overloaded regions of chromatogram, Chromatof then returned 17,380 peaks for 16 pooled samples. First, the commercial software was used to automatically identify peaks matching a retention time and mass spectrum of a reference standard, internal standard, or surrogate. These matching compounds were manually reviewed and classified as "confirmed"; these identifications are equivalent to a Level 1 Identification[3].

For peaks not matching a standard, the 10 highest scoring spectral matches with similarity (i.e., forward-matching) scores greater than or equal to 650 from the NIST 2017 library were reviewed. Based on the quality of the library matches, ion co-maximization, and expert judgement peaks were classified into one of five broad categories:

- tentatively identified or Level 2 Identification equivalent (e.g., a compound with a very good fit to NIST 2017 with at least three fragmentation ions for which there are no other viable library matches),
- tentatively identified as correct chemical class, but specific isomer or conformer not known or a Level 3 Identification equivalent (e.g., a compound with a very good fit to NIST 2017 with at least three fragmentation ions but with one or more other NIST 2017 matches of the same chemical class, such as is encountered with different ring substitution patterns, alkyl chain lengths for organic acids and hydrocarbons, etc.),
- non-specific (e.g., compound has less than three fragment ions and multiple possibilities for chemical class),
- unknown (e.g., no viable match to NIST 2017 or multiple chemical class possibilities), and

- artifact/low quality (e.g., low-quality peaks and artifacts resulting from poor or erroneous deconvolution)[4].

Throughout this manual curation process, expert reviewers had the opportunity to tag peaks as "compounds of potential relevance" (CPRs). Considerations used to tag a peak as a CPRs were based on quality of match, importance to human health, and expert judgement. Compounds tagged as CPRs may not necessarily have been noted in every sample replicate; however, this is not an issue as the CPR processing template was compiled from peaks tagged in the entire sample set, resulting in 130 down-selected compounds with unique retention times and mass spectra, and high match fidelity to the NIST 2017 library identification. While follow-up confirmation was not performed on these compounds, the sample data were re-processed after this initial round of tagging, targeting only the CPR list to resolve any quality issues that could have arisen during the broader peak picking process. This processing method was modified to use a lower signal-to-noise (S/N) threshold to capture peak detection variances (based on concentration) across the replicates analyzed. This is especially important to capture low level compounds near the peak detection threshold.

Follow-up targeted analysis was performed on pooled samples as well as on the SRM material via high-resolution GC×GC/TOF-MS. Target compounds with calibration curves were reviewed against the reference spectra for positive identification based on mass spectrum and retention time. Calibration curves were constructed using a single ion for quantification therefore the peak areas are not directly comparable to the screening results for which the de-convoluted Total Ion Chromatogram (DTIC) is used to estimated concentration (by comparison to the naphthalene-D8 response). Although a single ion was used for quantification, the full scan mass spectrum was used during review for positive identification.

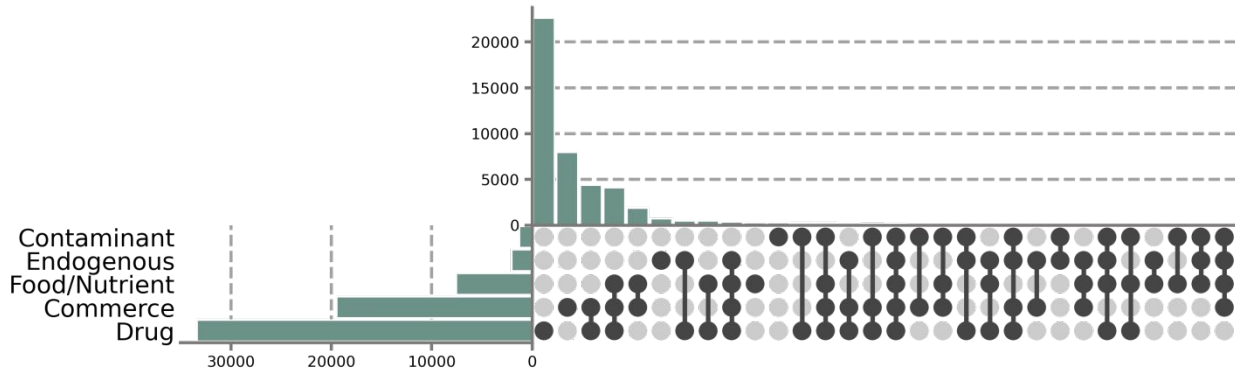
Statistical and Chemometric Analysis

To account for retention time drift of the equipment, the range of both first and second retention times of all internal standards was calculated for both suspect screening and CPR analyses. In addition, the same retention time ranges were calculated for each identification in each analysis across all replicates of a

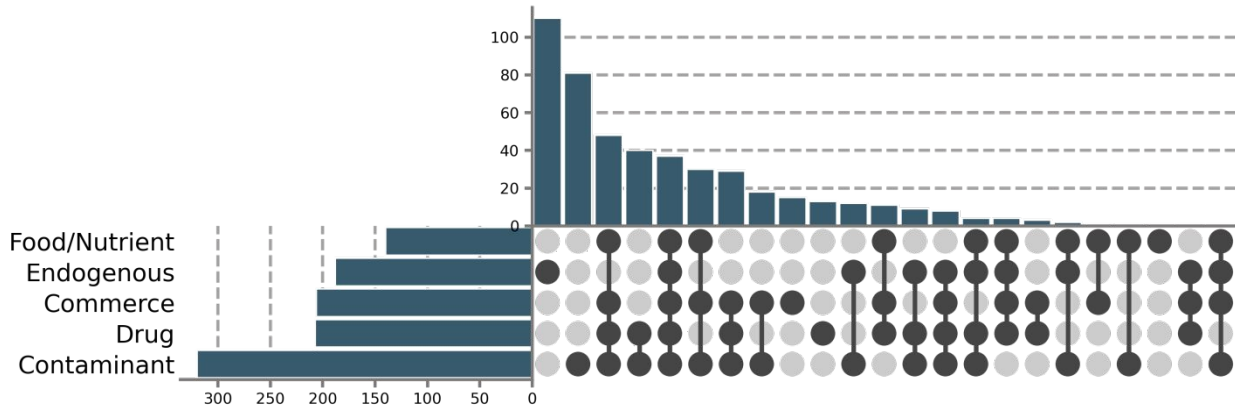
sample. If the range of retention times for an individual chemical was greater than that of the internal standards, then those tentative identifications were discarded. After this step, the number of replicates in which a chemical was tentatively identified was calculated. If a chemical was only tentatively identified in one replicate for a sample, then that tentative identification was removed from further processing.

Separate data files were generated for the suspect screening analysis, the CPR analysis, and the confirmation analysis. The SSA and CPR files were combined, and a "credibility rating" was assigned to each tentative hit in this merged data file. The highest credibility rating was assigned to CPR tentative identifications, tentative chemical identification from the SSA had the second highest, and tentative isomer identification in the SSA had the lowest. If a chemical was tentatively identified multiple times at the same credibility rating in multiple replicates from the same sample, the hits were aggregated with the mean target response being calculated. After this step, if a chemical was tentatively identified multiples times in in multiple sample replicates from the same sample, but at varying credibility ratings, only the most credible, aggregated tentative identification was kept, while the lower credibility tentative identifications were discarded from further analysis. All confirmation, suspect screening and chemicals of potential relevance results are reported in Supplemental Table 2.

Supplemental Figures

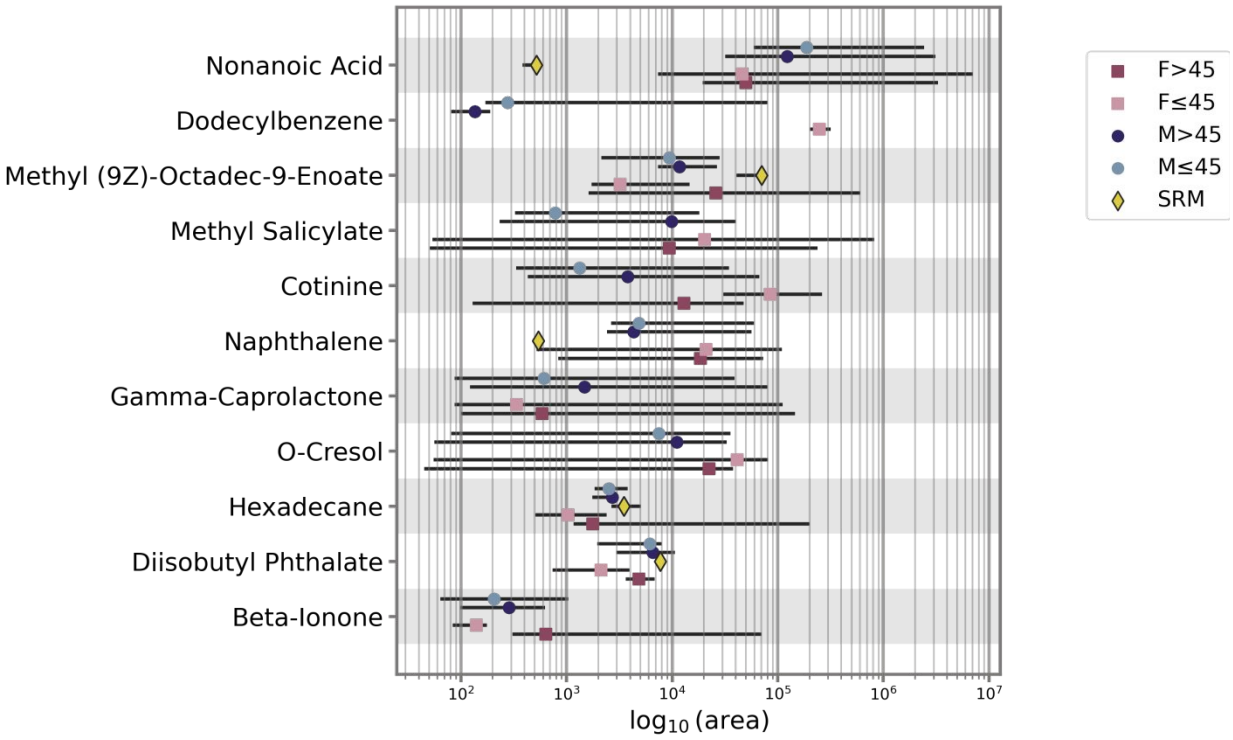


Supplemental Figure 1: Substances in Source Substance Repository. The bar chart in the lower, left panel shows the full tally of compounds identified in each specific substance source type (SST). The lower, right panel shows filled circles which denote the different source types that make up the SST intersection set. The bar chart in the upper, right panel shows the cardinality (or count of compounds identified in SST intersections).



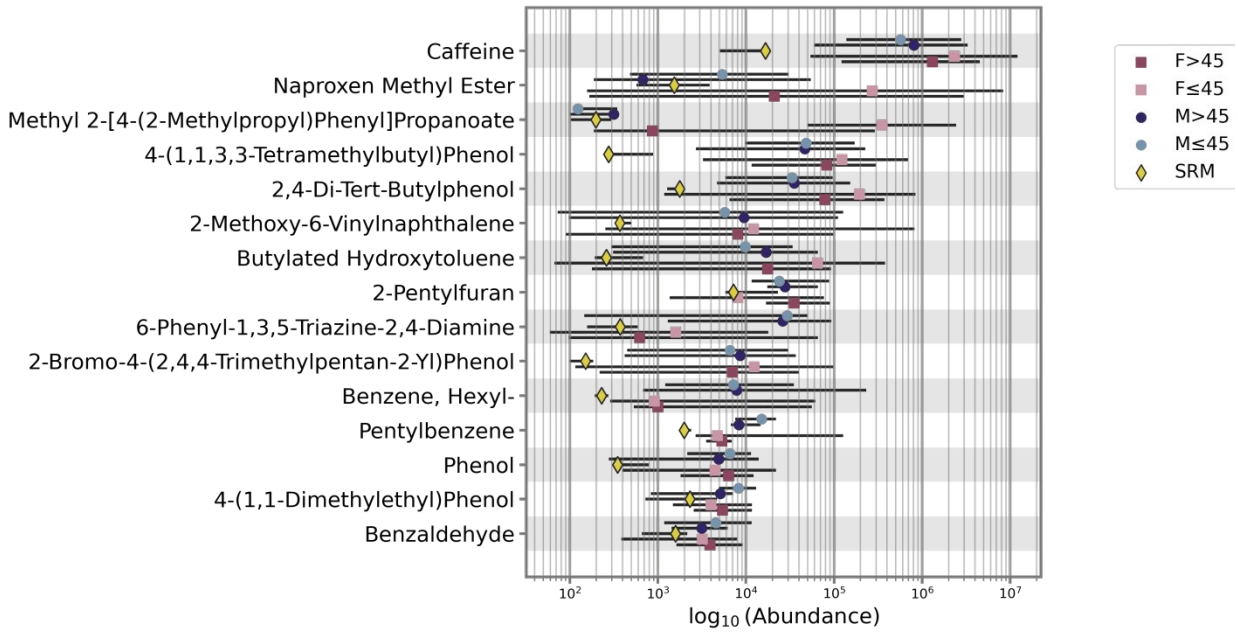
Supplemental Figure 2: Substances identified by confirmation, CPR, and suspect screening analyses. The bar chart in the lower, left panel shows the full tally of compounds identified in each specific substance source type (SST). The lower, right panel shows filled circles which denote the different source types that make up the SST intersection set. The bar chart in the upper, right panel shows the cardinality (or count of compounds identified in SST intersections).

234



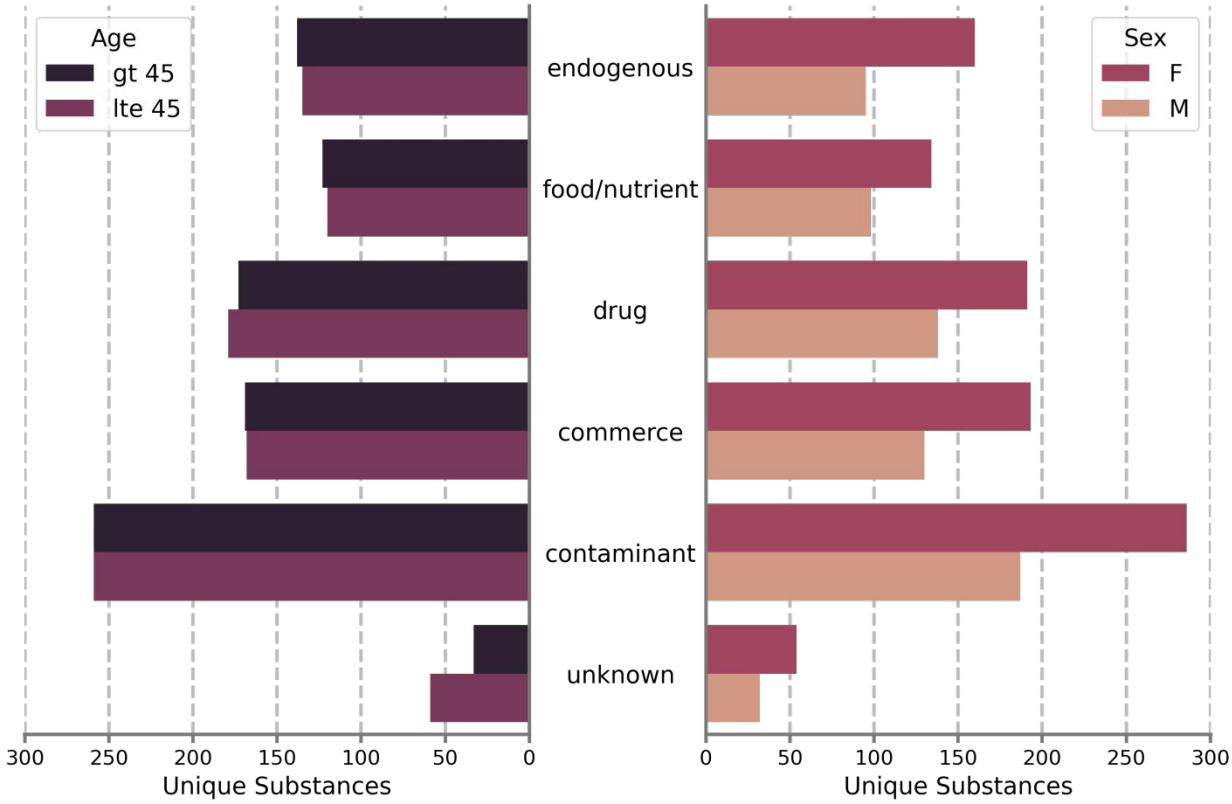
Supplemental Figure 3: The distribution of abundances for confirmed chemical substance in pooled human serum samples. The squares, circles, and diamonds represent the mean abundance across all for Female, Male, and SRM samples, respectively. Blue and red shading of squares and circles represent age stratification. Grey bars span from the minimum to the maximum abundance measured for each substance.

235



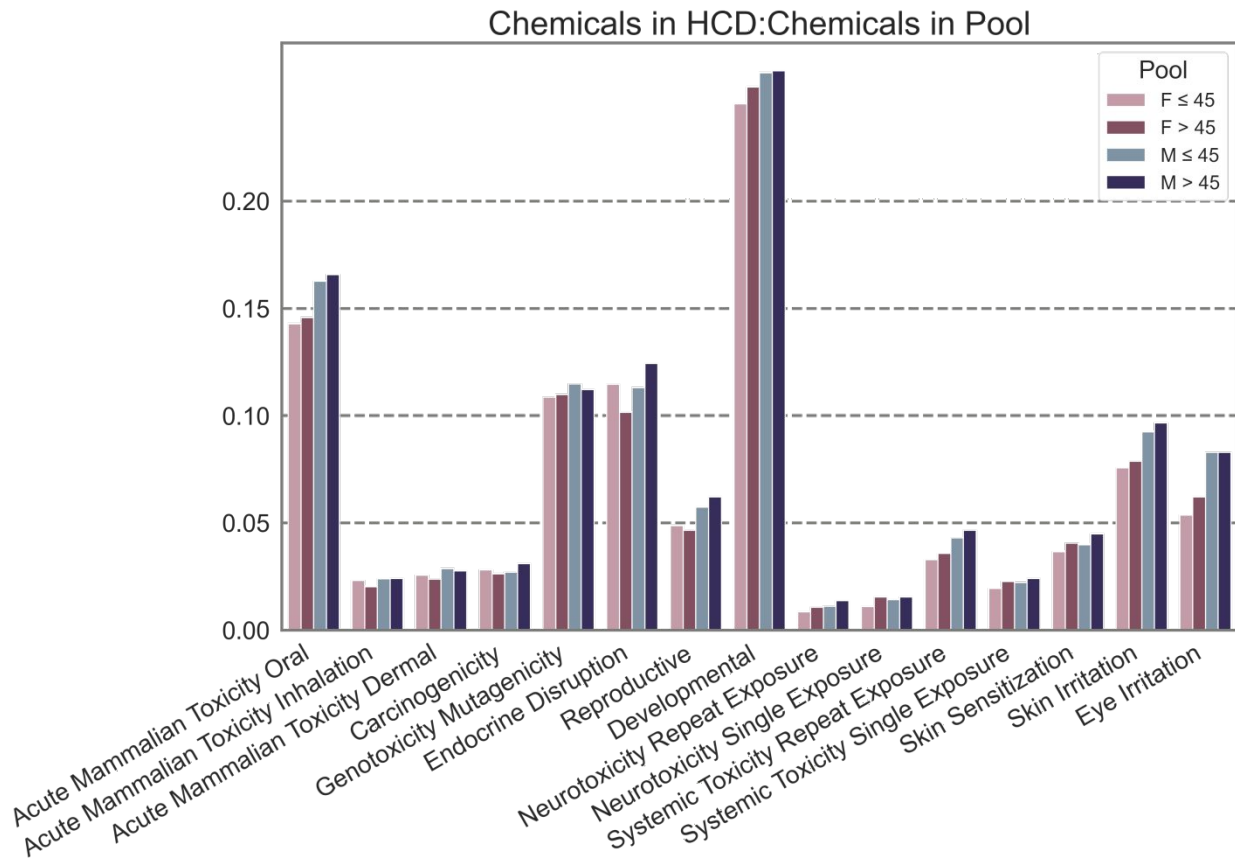
Supplemental Figure 4: The distribution of abundances for tentatively chemical substance in pooled human serum samples that were also tentatively identified in the SRM. The squares, circles, and diamonds represent the mean abundance across all for Female, Male, and SRM samples, respectively. Blue and red shading of squares and circles represent age stratification. Grey bars span from the minimum to the maximum abundance measured for each substance.

236



Supplemental Figure 5: The left panel shows the count of unique substances of each substance source type across the different age groups considered in this study. The right panel shows the count of unique substances of each substance source type across sex.

237



Supplemental Figure 6: Ratio of the number of chemicals identified as “very high”, “high”, or “medium” hazard to the total number of unique chemicals identified in each pool. The hazards of “very high”, “high”, or “medium” were provided from the U.S. EPA’s Hazard Comparison Tool.

238

239

References

(1) Crinnion, W. J. The CDC Fourth National Report on Human Exposure to Environmental Chemicals: What It Tells Us about Our Toxic Burden and How It Assists Environmental Medicine Physicians. *Alternative Medicine Review* **2010**, *15* (2).

(2) Arey, J. S.; Martin Aparicio, A.; Vaiopoulou, E.; Forbes, S.; Lyon, D. Modeling the GCxGC Elution Patterns of a Hydrocarbon Structure Library to Innovate Environmental Risk Assessments of Petroleum Substances. *Environmental Science /& Technology* **2022**, *56* (24), 17913–17923. https://doi.org/10.1021/acs.est.2c06922.

(3) Koelmel, J. P.; Xie, H.; Price, E. J.; Lin, E. Z.; Manz, K. E.; Stelben, P.; Paige, M. K.; Papazian, S.; Okeme, J.; Jones, D. P.; others. An Actionable Annotation Scoring Framework for Gas Chromatography-High-Resolution Mass Spectrometry. *Exposome* **2022**, *2* (1), osac007.

(4) Favela, K. A.; Hartnett, M. J.; Janssen, J. A.; Vickers, D. W.; Schaub, A. J.; Spidle, H. A.; Pickens, K. S. Nontargeted Analysis of Face Masks: Comparison of Manual Curation to Automated GCxGC Processing Tools. *Journal of the American Society for Mass Spectrometry* **2021**, *32* (4), 860–871. https://doi.org/10.1021/jasms.0c00318.