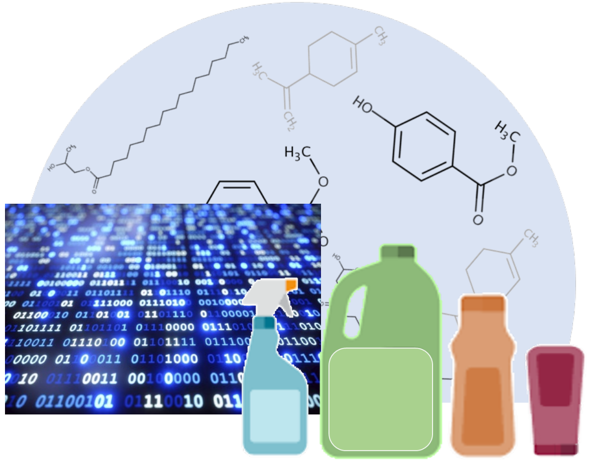
https://orcid.org/0000-0001-9547-1654
Isaacs.kristin@epa.gov

United States Environmental Protection Agency

Mining Potential Chemical Co-exposures from Consumer Product Purchasing and Ingredient Data



Kristin Isaacs and Zachary Stanfield

Center for Computational Toxicology and Exposure, US-EPA, RTP, NC

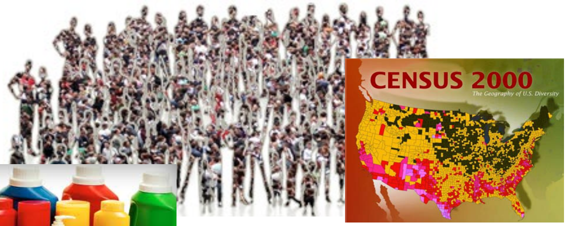
Introduction

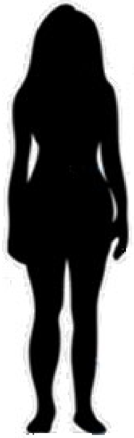
United States Environmental Protection Agency

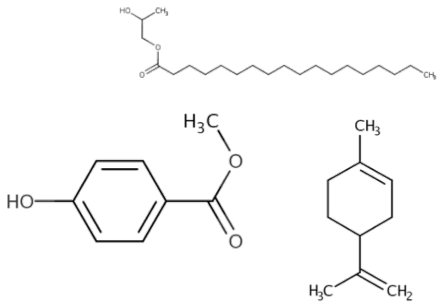
- Addressing risks associated with chemical mixtures is a challenge
  - Too many chemicals and too many co-exposures
- EPA's ToxCast Program has screened thousands of chemicals for bioactivity in high-throughput *in vitro* assays
- HTS and mixtures
  1. Predict activity from component chemical responses using modeling
  2. Test whole mixtures (can inform #1)
- But which mixtures to test?
- **In ExpoCast, we are developing tools that allow us to identify relevant chemicals with potential real-world co-exposures**

Approaches for Indentifying Co-Exposures

Prediction of Human Behavior and Resulting Exposure





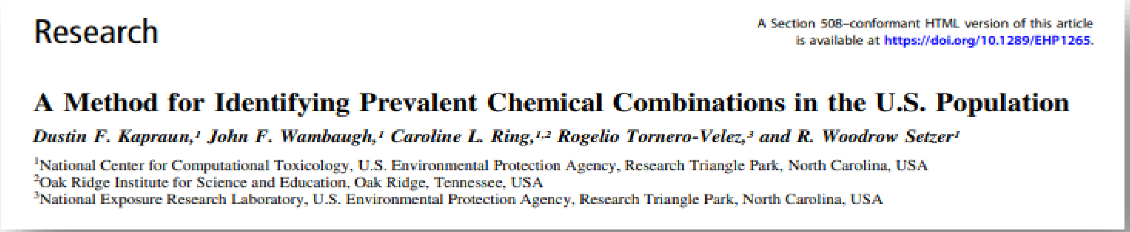


True Co-exposures

Biomonitoring

- **Modeling approaches**
  - Multiple sources, pathways, and routes of exposure can be considered
  - Uncertainties associated with estimating external versus internal exposure (dose) – timing of exposures and consideration of absorption, distribution, metabolism, and excretion (ADME) processes
  - Impacted by data gaps in behavior (e.g., consumer habits and practices), source information (e.g., chemical use or ingredient data), or toxicokinetics
- **Biomonitoring**
  - Can identify both parent chemicals and metabolites in blood or urine
  - Aggregate over time (e.g., bioaccumulating compounds)
  - Limited number of chemicals (expensive, need standard analytical methods)

Mining Human Biomonitoring Data to Identify Prevalent Chemical Mixtures

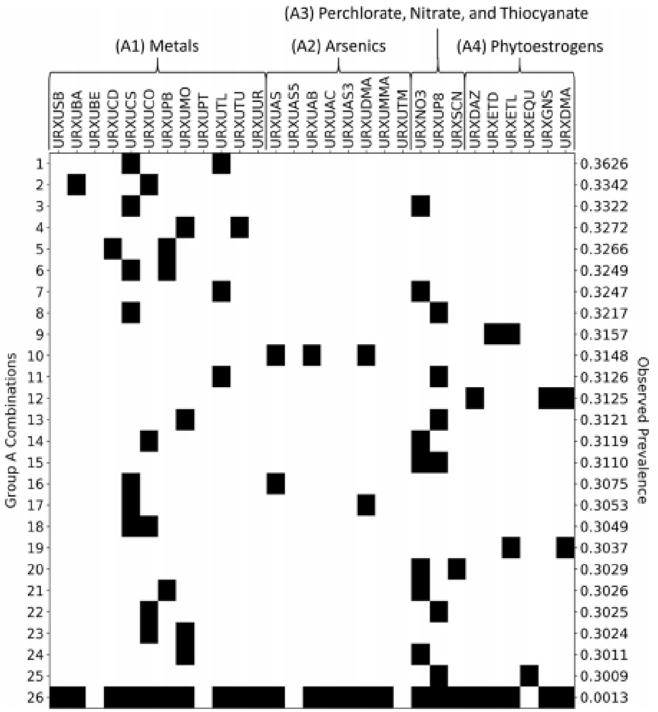

Research

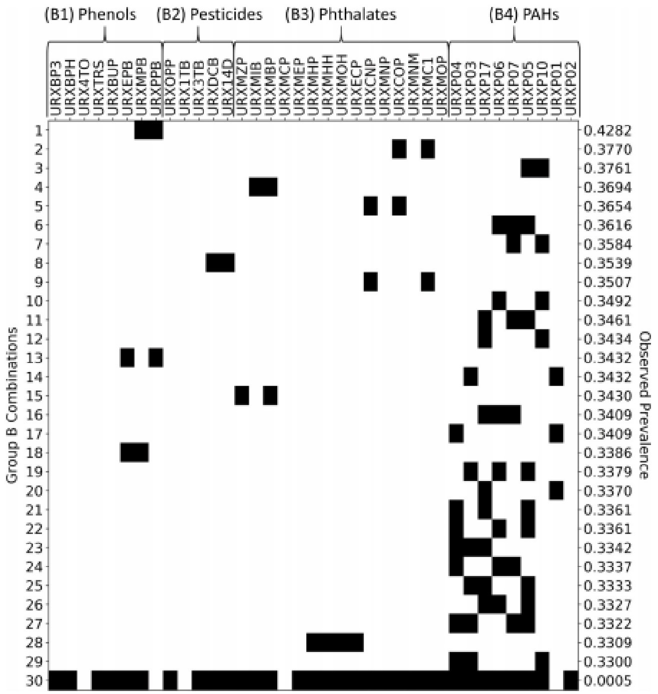
A Section 508–conformant HTML version of this article is available at https://doi.org/10.1289/EHP1265.

A Method for Identifying Prevalent Chemical Combinations in the U.S. Population

Dustin F. Kapraun,[1] John F. Wambaugh,[1] Caroline L. Ring,[1,2] Rogelio Tornero-Velez,[3] and R. Woodrow Setzer[1]

[1]National Center for Computational Toxicology, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, USA
[2]Oak Ridge Institute for Science and Education, Oak Ridge, Tennessee, USA
[3]National Exposure Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, USA



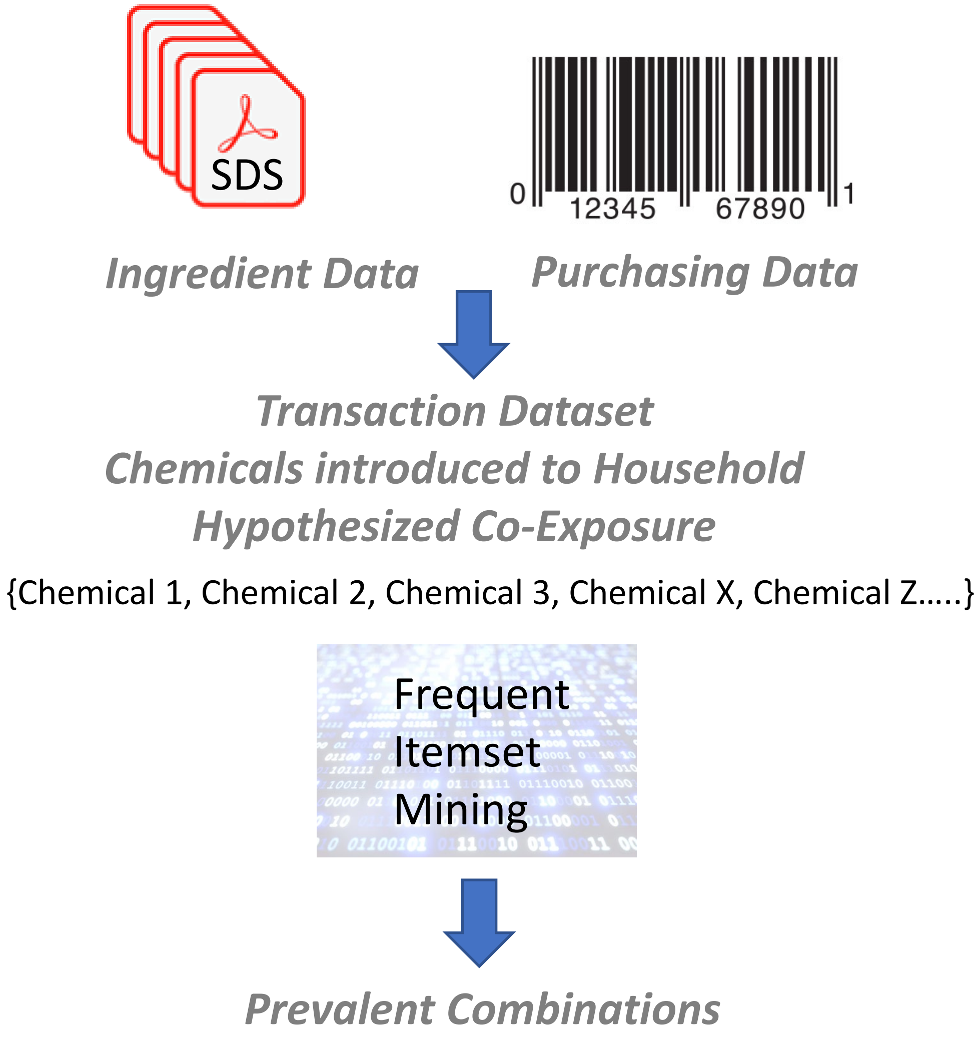


- Kapraun et al. (2017) mined biomonitoring data from the NHANES study to identify prevalent chemical combinations
- Measured concentrations were discretized to presence/absence using a fixed threshold
- Examined co-occurrence within three groups of chemicals measured in unique subsamples of the study population, using frequent itemset mining (FIM)
- Identified 90 chemical combinations consisting of relatively few chemicals that occur in at least 30% of the U.S. population
- Identified three “supercombinations” of chemicals that occurred in a smaller fraction of the population

Kapraun et al., Environmental Health Perspectives, 125:8, 2017

Current Approach

- Integrate large datasets of consumer product ingredient and product purchasing information to develop a dataset that can be mined for chemical co-exposures
- Apply FIM to identify prevalent co-occurring chemicals within household-months
- Stratified results by household demographics to characterize variability in co-exposure patterns and identify potential chemical combinations associated with sensitive populations, such as families with young children and women of childbearing age



EPA-ORD's Chemicals and Products Database (CPDat)

United States Environmental Protection Agency

- EPA ORD database containing curated chemical use and consumer product ingredient data
- Public version of the dataset contains ingredient data for over 60,000 products, mapped to standardized product categories for use in exposure assessment and modeling
- Also recently extracted ingredient data from 230,407 retailer-provided product safety data sheets (SDSs), including product name, category, universal product code (UPC), and chemical identifiers

www.nature.com/scientificdata

SCIENTIFIC DATA

OPEN

Data Descriptor: The Chemical and Products Database, a resource for exposure-relevant data on chemicals in consumer products

Received: 16 October 2017
Accepted: 30 April 2018
Published: 10 July 2018

Kathie L. Dionisio[1], Katherine Phillips[1], Paul S. Price[1], Christopher M. Grulke[2], Antony Williams[2], Derya Biryol[1,3], Tao Hong[4] & Kristin K. Isaacs[1]

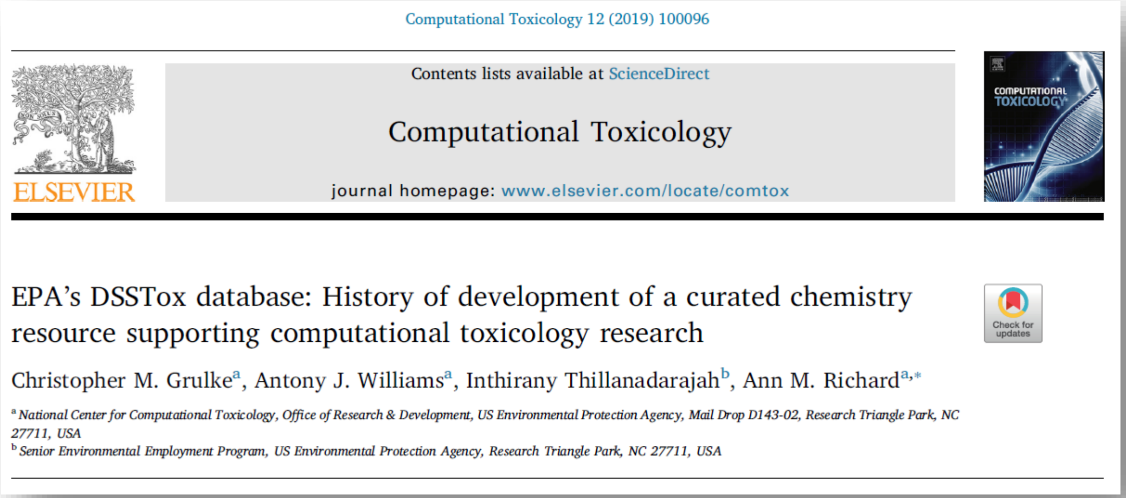
Dionisio et al. Sci Data 5:180125 (2018).

https://www.epa.gov/chemical-research/chemical-and-products-database-cpdat

EPA-ORD's Chemicals and Products Database (CPDat)

United States Environmental Protection Agency

- EPA ORD database containing curated chemical use and consumer product ingredient data
- Public version of the dataset contains ingredient data for over 60,000 products, mapped to standardized product categories for use in exposure assessment and modeling
- Also recently extracted ingredient data from 230,407 retailer-provided product safety data sheets (SDSs), including product name, category, universal product code (UPC), and chemical identifiers
- Chemical identifiers curated to harmonized EPA Distributed Structure-Searchable Toxicity (DSSTox) Substance Identifiers (DTXSIDs)


Computational Toxicology 12 (2019) 100096

Contents lists available at ScienceDirect

Computational Toxicology

journal homepage: www.elsevier.com/locate/comtox

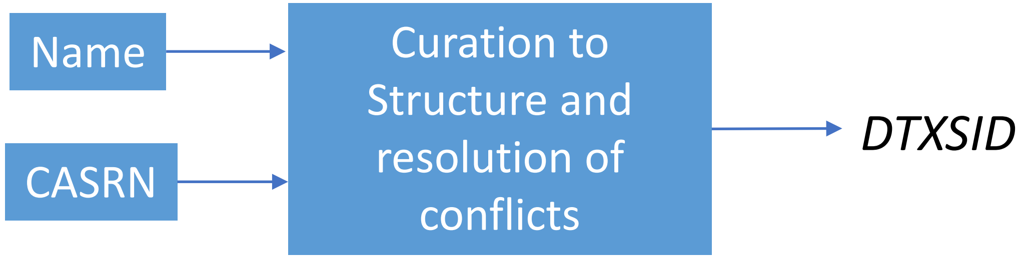
ELSEVIER

EPA's DSSTox database: History of development of a curated chemistry resource supporting computational toxicology research

Christopher M. Grulke[a], Antony J. Williams[a], Inthirany Thillanadarajah[b], Ann M. Richard[a,*]

[a] National Center for Computational Toxicology, Office of Research & Development, US Environmental Protection Agency, Mail Drop D143-02, Research Triangle Park, NC 27711, USA

[b] Senior Environmental Employment Program, US Environmental Protection Agency, Research Triangle Park, NC 27711, USA



Grulke et al. Comp Tox 12:100096 (2019).

Consumer Product Purchasing Study

- EPA initiated a collaboration with Nielsen in 2013
- Shared data from the National Consumer Panel (NCP)
- Formerly called “Homescan” project

Purchasing Data

- 60,000 U.S. households for 1 year (2012)
- Demographic information for each household
  - Income, number of household members, Nielsen market (metro area), county size, race, presence and age of children, age and occupation of female head of household
- All purchases for product categories of interest to Nielsen
  - 29 broad categories called "Groups" (e.g., Household Cleaners, Cosmetics, Fresheners and Deodorizers).
  - Date of purchase, UPC, brand, number of units, size
  - ~4.6 million individual product purchase records
- 133,966 unique product UPCs
- Recent publication: Tornero-Velez et. al (2020) examined product co-purchases which gave us some idea about chemical co-exposure from previous ingredient data; the ability to link individual purchases to specific chemicals is a major step forward.

Tornero-Velez et al., Risk Analysis, 125:8, 2017

Data Integration

United States Environmental Protection Agency

Product-Chemical Data

UPC	Chemicals
UPC1	DTXSID1, DTXSID2, DTXSID4
UPC2	DTXSID2, DTXSID3
UPC3	DTXSID1, DTXSID5, DTXSID6

Purchasing Data

Date	Household (HHLD)	UPC (12 digits)	Product Variables	Demographic Variables
2012-01-01	HHLD1	UPC3	...	...
2012-01-23	HHLD1	UPC1	...	...
2012-02-09	HHLD2	UPC2	...	...

Direct and Fuzzy Matching by UPC

Could match ~50.3% products

Monthly Transaction Matrix

HHLD-month	Chemicals
HHLD1-01	DTXSID1, DTXSID2, DTXSID4, DTXSID5, DTXSID6
HHLD2-02	DTXSID2, DTXSID3
⋮	⋮

Data Summary

Data	Count
Transactions	539,857
Households	53,525
Products	31,375
Chemicals	783

Chemical Lists

- Analysis of co-occurring chemicals was restricted to chemicals of regulatory or biological interest in order to avoid identification of prevalent chemical combinations containing common substances having little relevance to risk assessment (e.g., water)
- **Broad Chemical List:** Active public chemical inventory of the Toxic Substances Control Act (TSCA)
  - **649** chemicals in the consumer product transaction dataset
- **Case-Study: Potential Endocrine Active Chemicals (EACs)**

Source	Investigated Biological Action	Chemicals Predicted to be Active	Chemicals Mapped to Purchased Products
Collaborative Estrogen Receptor Activity Prediction Project (CERAPP)[1]	Estrogen Disruptors	1,142	10
Collaborative Modeling Project for Androgen Receptor Activity (COMPARA)[2]	Androgen Disruptors	16,112	42
Additional potential EACs from Literature Sources[3]	Multiple		17
Total (unique)			**65**

[1]*Mansouri, K. et al. 2016. Environmental Health Perspectives. 124:1023-1033.*
[2]*Mansouri, K. et al. 2020. Environmental health perspectives. 128:27002.*
[3]*Dodson et. al. Environmental health perspectives. 120:935-943.*

Frequent Itemset Mining

- Itemset
  - A collection of one or more items
    - Example: {DTXSID1, DTXSID4, DTXSID5}
  - k-itemset
    - An itemset that contains k items
- Relative support/prevalence (σ)
  - Fraction of transactions that contain an itemset
  - E.g. σ({DTXSID1, DTXSID4, DTXSID5}) = 2/5
- Frequent itemset
  - An itemset whose support is greater than or equal to a *minimum support* threshold

Transaction ID	Items
HHLD-01	DTXSID1, DTXSID4
HHLD-02	DTXSID4, DTXSID5, DTXSID3, DTXSID2
HHLD-03	DTXSID1, DTXSID5, DTXSID3, DTXSID6
HHLD-04	DTXSID4, DTXSID1, DTXSID5, DTXSID3
HHLD-05	DTXSID4, DTXSID1, DTXSID5, DTXSID6

Apply to transaction data to identify prevalent combinations

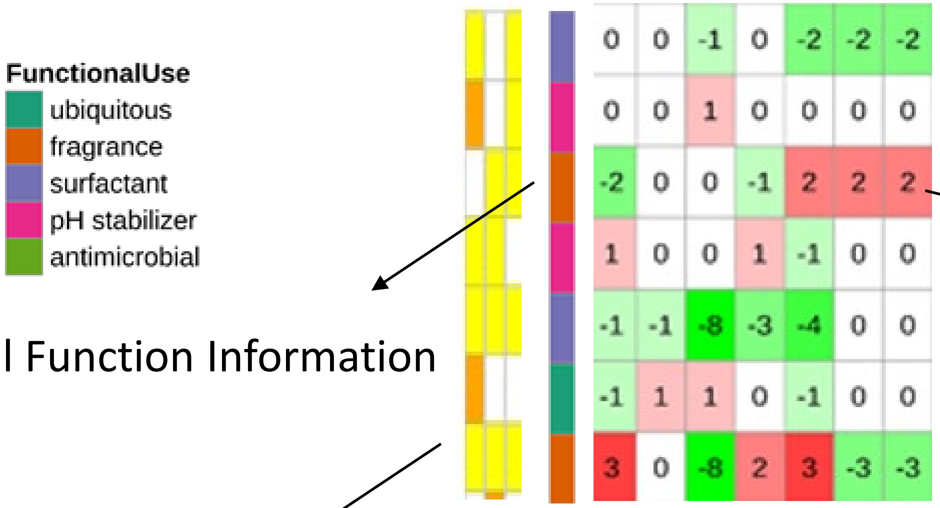
Frequent Itemset Mining Analyses

- Performed using the *ECLAT (*Equivalence Class Clustering and bottom-up Lattice Traversal) function of the *Arules* R package
- Performed identification of prevalent individual chemicals and combinations
  - For TSCA chemicals and EACs, based on a threshold prevalence for chemical group that provided a manageable number of itemsets
  - Within product groups
  - Within demographics, including:
    - Women of childbearing age
    - Different income ranges
    - Race of female head of household
    - Education level
    - Different family sizes/ages of children
- Interpreted chemicals within prevalent combinations by examining chemical functions
  - Harmonized functional uses defined by Phillips et al. (*2017).*
    - Dataset of 14,000+ reported chemical-function pairs

Phillips et al. Green Chem 19:1063-1074, 2017.

Results: Orientation

Demographic



{Chemical or Chemical Combination}

Chemical Function Information

Receptor Information

Deviation of the *rank* of the chemical or combination for each demographic from the rank in the population overall

Elevated rank relative to overall population
Potentially higher exposure for the demographic

Reduced rank relative to overall population
Potentially lower exposure for the demographic

Most Prevalent Single Chemicals

Chemical Rank by Demographic Group

Prevalence	Asian	African American	Hispanic	White	Grade And High School	College	Post College	No Child	Under 6	Under 13	Under 18	Lower	Mid Lower	Mid Higher	Higher	Non-Childbearing	Childbearing	Demographic
0.517	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	{Ethanol}
0.332	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	{Glycerol}
0.26	-1	0	-1	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	{1,2-Propylene glycol}
0.242	1	-2	1	0	-3	0	0	0	0	0	0	-2	-1	0	0	0	0	{Sodium dodecyl sulfate}
0.24	-3	-2	0	0	0	0	-1	0	0	0	0	2	1	0	0	-1	0	{Isobutane}
0.228	1	1	0	0	2	0	1	0	0	0	0	1	0	0	0	1	0	{Sulfuric acid, mono-C10-16-alkyl esters, sodium salts}
0.222	1	3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	{Poly(oxy-1,2-ethanediyl), .alpha.-sulfo-.omega.-hydroxy-, C10-16-alkyl ethers, sodium salts}
0.21	1	-1	0	0	0	0	0	0	0	0	0	-1	-1	0	0	0	-1	{Sodium hydroxide}
0.19	-1	1	0	-1	0	0	-1	0	-2	-2	-2	1	1	0	0	0	-1	{Propane}
0.186	-4	0	-1	1	0	0	1	0	0	0	0	0	0	0	0	0	-1	{Sodium carbonate}
0.154	2	-14	1	0	-2	0	0	-1	2	2	2	-5	-4	-1	0	-1	3	{Sodium [dodecanoyl(methyl)amino]acetate}
0.154	2	1	0	0	1	0	0	1	-1	0	0	1	1	1	0	1	0	{Sodium chloride}
0.133	-4	-3	-3	-1	-1	-1	-8	-3	-4	0	0	0	0	-1	-1	-2	0	{D-Limonene}
0.131	1	-4	-1	1	-1	1	1	0	-1	0	0	-5	-4	-2	1	0	0	{Diethylenetriaminepentaacetic acid pentasodium salt}
0.131	-7	3	-4	-1	3	0	-8	2	3	-3	-3	3	3	2	-3	2	-6	{C10-16-Alkyldimethylamines oxides}
0.128	-2	-1	-4	1	-1	0	2	1	-3	0	-3	1	0	-1	1	0	-2	{Sodium hypochlorite}
0.125	2	4	3	-2	1	-2	-2	0	-4	-2	0	3	3	2	0	0	0	{Carrageenan, native}
0.125	2	-4	1	1	0	1	-4	-1	0	1	2	-4	-3	-1	2	0	-2	{Ethanolamine}
0.121	7	-4	1	-1	-1	1	4	-1	3	-1	-2	1	2	1	-4	-1	3	{Titanium dioxide}
0.118	-3	-8	-6	2	1	-1	4	2	0	-6	-7	-1	-2	0	-2	1	-8	{Citric acid}

FunctionalUse

Rank Difference

5
0
-5
-10

Demographic
- Race/Ethnicity
- Education
- Income
- Family Comp.
- Female Age

FunctionalUse
- ubiquitous
- fragrance
- surfactant
- pH stabilizer
- antimicrobial

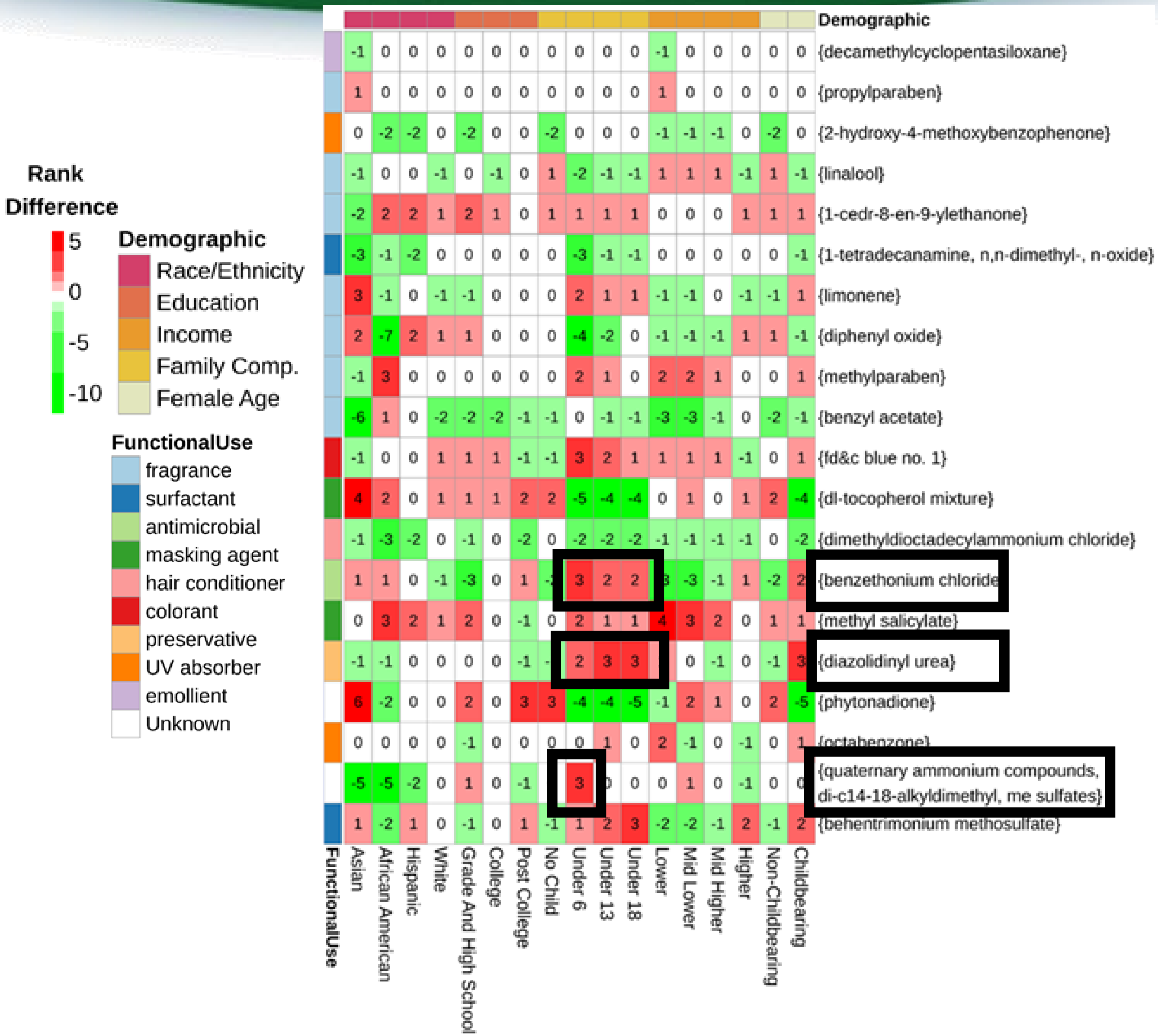
e.g., titanium dioxide has a higher prevalence rank for Asian female head of household

Two common cleaning product ingredients had reduced prevalence houses where the female head had post-college education

Group 1 (Broad TSCA Inventory)

- 20 overall most prevalent individual chemicals
- Top 5 chemicals are what were termed "ubiquitous function" chemicals - perform a variety of functions in products
- Differences by demographic can be observed

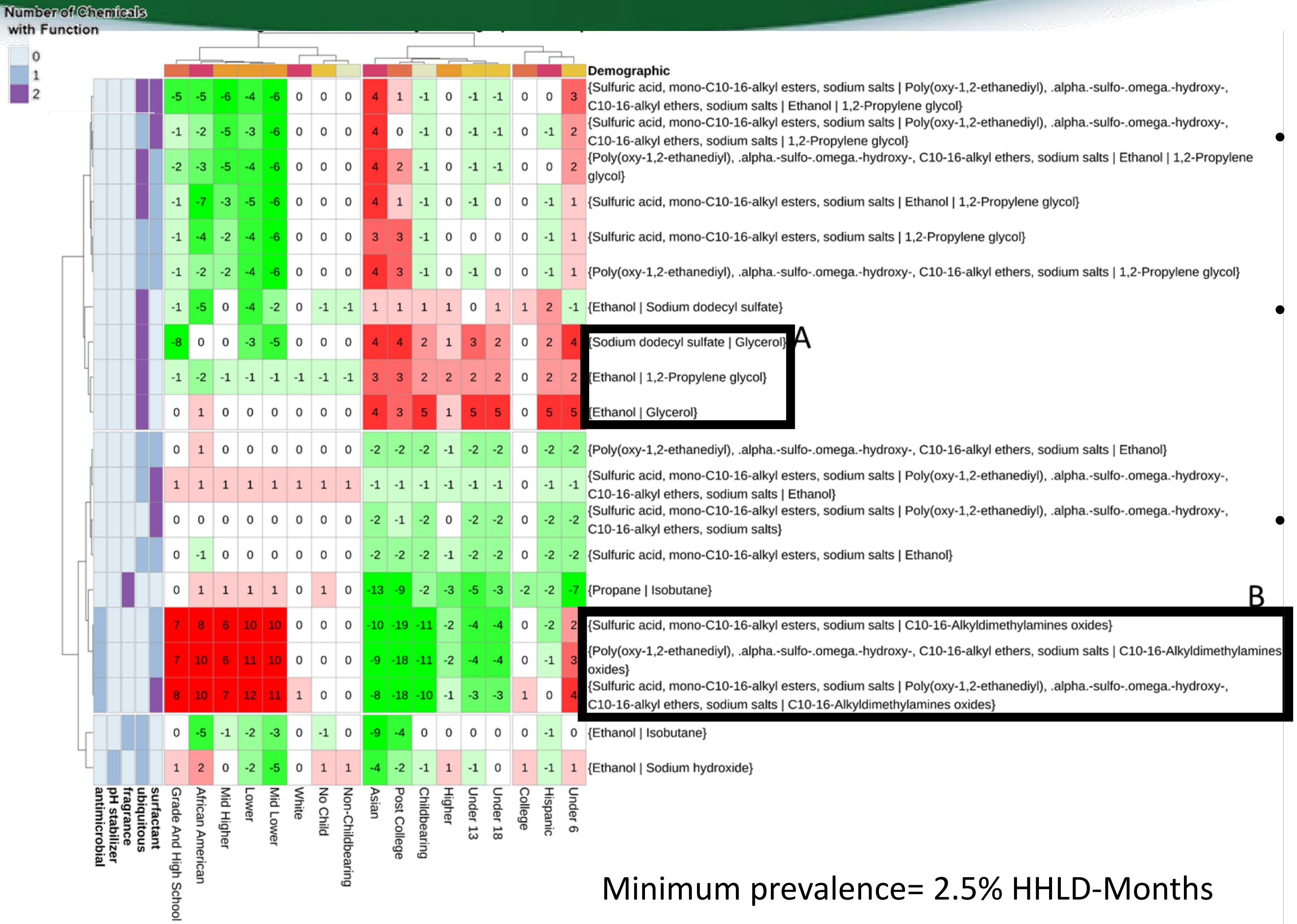
Most Prevalent Single Chemicals



Endocrine Active Chemicals

- Many of the most prevalent EAC chemicals were fragrances (or categorized as such due to presence in fragrance formulations)
- Many of these chemicals were present in a variety of personal care products
- *Benzethonium chloride* and *diazolidinyl urea*, which were ranked 2 or 3 places higher in households with children, are commonly used as topical antimicrobial agents in baby wipes, bubble baths, cosmetics, and skin care products
- Households with children under 6 have a higher ranking for *quaternary ammonium compounds, di-c14-18-alkyldimethyl, me sulfates*, which are commonly used in disinfectants and hand soaps

Prevalent Combinations



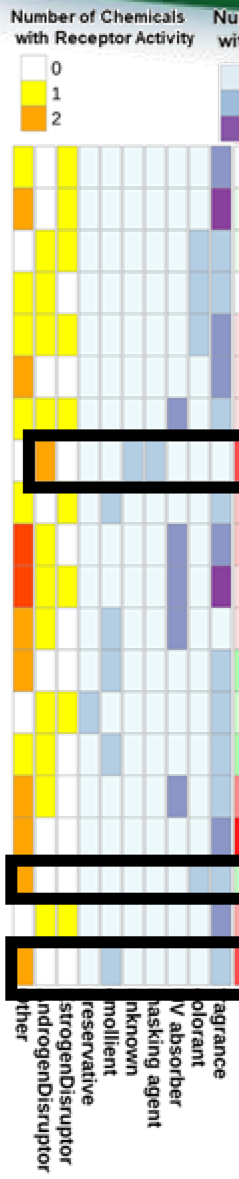
Minimum prevalence= 2.5% HHLD-Months

Group 1 (Broad TSCA Inventory)

- Here demographics and chemical sets are clustered to indicate the similarity of rankings of chemical combinations
- Set A: ubiquitous consumer product chemicals present in households with children, higher income, and more highly educated, representing generally high consumer product use
- Set B: elevated difference in rank in lower to middle income demographics and African American households, and reduced rank differences in Asian households and females with post-college education and females of childbearing age; these three sets contained antimicrobials and surfactants found in cleaning products

Prevalent Combinations

United States Environmental Protection Agency



sian	frican American	ispanic	White	rade And High School	ollege	ost College	o Child	nder 6	nder 13	nder 18	ower	id Lower	id Higher	igher	on-Childbearing	hildbearing	Demographic
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	{limonene | propylparaben}
-1	0	-1	0	0	0	0	0	-2	-3	0	0	0	0	0	0	0	{propylparaben | methylparaben | ethylparaben}
-1	0	-1	0	0	0	0	-2	1	1	0	0	0	0	-1	-1	0	{propylparaben | fd&c blue no. 1}
0	0	0	0	0	0	0	-2	2	1	0	0	0	0	-1	-1	0	{limonene | fd&c blue no. 1}
1	0	0	0	0	0	1	-2	2	1	0	1	0	0	-1	-1	0	{limonene | propylparaben | fd&c blue no. 1}
1	-8	4	0	0	0	1	3	-2	0	0	1	0	0	3	3	0	{diphenyl oxide | linalool}
1	0	1	-1	-1	0	0	-1	2	0	0	1	-1	-1	0	-1	0	{2-hydroxy-4-methoxybenzophenone | propylparaben | benzophenone}
6	2	0	1	1	0	2	4	-7	-9	-13	2	1	1	0	1	-17	{dl-tocopherol mixture | phytonadione}
2	1	1	0	0	-1	-1	0	0	1	0	2	-2	-1	0	0	-1	{decamethylcyclopentasiloxane | propylparaben}
2	1	1	0	0	1	2	0	4	1	2	2	1	1	0	0	2	{2-hydroxy-4-methoxybenzophenone | methylparaben | ethylparaben | benzophenone}
1	0	0	0	0	0	2	0	4	1	1	2	1	0	-1	0	2	{2-hydroxy-4-methoxybenzophenone | propylparaben | methylparaben | ethylparaben | benzophenone}
1	-3	5	0	0	0	1	0	2	1	1	2	0	0	1	0	1	{decamethylcyclopentasiloxane | 2-hydroxy-4-methoxybenzophenone | benzophenone}
-5	0	3	0	0	0	1	0	0	1	1	2	-1	0	0	0	1	{decamethylcyclopentasiloxane | linalool}
-3	4	-3	-1	0	0	-1	0	2	1	0	1	1	-1	0	0	1	{diazolidinyl urea | propylparaben}
-5	-3	2	1	0	0	-1	-1	-3	-1	2	-3	-6	-3	0	0	-1	{1-cedr-8-en-9-ylethanone | decamethylcyclopentasiloxane}
4	-1	4	0	-1	0	2	-1	0	1	0	0	-1	0	0	-1	1	{2-hydroxy-4-methoxybenzophenone | linalool | benzophenone}
8	-2	3	-1	-9	0	4	-1	3	3	2	-4	-1	-4	0	-4	3	{linalool | limonene}
-4	6	-4	-1	-2	0	0	3	-7	-2	-6	6	2	4	-3	2	-3	{linalool | 2-phenylethanol}
3	-1	3	-1	3	-3	-3	-3	-1	0	1	4	-1	0	0	-1	1	{1-cedr-8-en-9-ylethanone | propylparaben}
6	-2	2	1	-1	0	3	-1	9	6	3	6	-1	0	0	1	3	{decamethylcyclopentasiloxane | limonene}

Minimum prevalence= 0.1% HHLD-Months

Endocrine Active Chemicals

- One itemset {*dl-tocopherol mixture* | *phytonadione*}, contained two chemicals that targeted the same receptor (AR).
- The highest positive rank departure for households with children occurs for the itemset {*decamethylcyclopentasiloxane* | *limonene*}.
- Households with a female head of Asian race have the highest positive rank departure for the combination of *limonene* and *linalool*, the latter of which is used as a scent in many perfumed hygiene products and cleaning agents.
- African American households had a positive rank departure of 6 for the combination {*linalool* | *2-phenylethanol*}; the second chemical is a floral fragrance primarily present in air fresheners.

Discussion

- Collectively across all products and by product group, results indicated that households with children, households headed by women of color, and lower income households exhibited divergence from the general population in the chemical combinations they encounter most frequently.
    - This may be due to a need for different types of personal care products designed specifically for given races or ethnicities, brand or regional preferences, or simply the need for a wider variety of products in households with multiple children.
    - These patterns may reflect differential experiences and thus differential exposures among demographics.
- Lists of most prevalent combinations (overall and for various demographics) can be evaluated for feasibility for testing in *in vitro* assay systems, and further prioritized based on single-chemical activity or exposure-related factors.
- New non-targeted analysis (NTA) studies of biological media such as blood or urine can complement and evaluate predictions of co-exposures associated with consumer products.
    - Such studies also have the potential for identifying mixtures containing metabolites of consumer product chemicals.

Summary

United States Environmental Protection Agency

- Humans are exposed to thousands of chemicals from the products they purchase and use within the household.
- Assessing every possible set of chemicals for toxicity is an impossible task but also an unnecessary one as the number of chemical mixtures that are prevalent and occur in real-world scenarios is drastically less.
- We have presented here a novel approach that applies FIM on a dataset describing the chemicals entering households through purchased consumer products to identify a manageable number of chemical combinations that regularly occur in homes across the US.
- These identified combinations can inform the prioritization of chemical combinations for toxicity testing.

https://orcid.org/0000-0001-9547-1654
Isaacs.kristin@epa.gov

United States Environmental Protection Agency

Project Team

Cody Addington
Timothy Buckley
Kathie Dionisio
Kristin Isaacs
David Lyons
Katherine Phillips
Zachary Stanfield
Rogelio Tornero-Velez

ExpoCast Project (Exposure Forecasting)

CCTE
Linda Adams
Miyuki Breen*
Alex Chao*
Dan Dawson*
Mike Devito
Kathie Dionisio
Christopher Ecklund
Marina Evans
Peter Egeghy
Michael-Rock Goldsmith
Chris Grulke
Mike Hughes
Kristin Isaacs
Richard Judson
Jen Korol-Bexell*
Anna Kreutz*
Charles Lowe*
Seth Newton
Katherine Phillips
Paul Price
Tom Purucker
Ann Richard
Caroline Ring
Marci Smeltz*
Jon Sobus
Risa Sayre*
Mark Sfeir*
Mark Strynar
Zach Stanfield*
Rusty Thomas
Mike Tornero-Velez
Elin Ulrich
Dan Vallero
Taylor Wall*
John Wambaugh
Barbara Wetmore
Antony Williams

CEMM
Xiaoyu Liu

CPHEA
Jane Ellen Simmons
Jeff Minucci

CESER
David Meyer
Gerardo Ruiz-Mercado
Wes Ingwersen

***Trainees**

Collaborators

Arnot Research and Consulting
Jon Arnot
Johnny Westgate
Institut National de l'Environnement et des Risques (INERIS)
Frederic Bois
Integrated Laboratory Systems
Kamel Mansouri
National Toxicology Program
Steve Ferguson
Nisha Sipes
Ramboll
Harvey Clewell
Silent Spring Institute
Robin Dodson
Southwest Research Institute
Alice Yau
Kristin Favela
Summit Toxicology
Lesa Aylward
Technical University of Denmark
Peter Fantke
Tox Strategies
Miyoung Yoon
Unilever
Beate Nicol
Cecilie Rendal
Ian Sorrell
United States Air Force
Heather Pangburn
Matt Linakis
University of California, Davis
Deborah Bennett
University of Michigan
Olivier Jolliet
University of Texas, Arlington
Hyeong-Moo Shin