

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



Zachary Stanfield and Kristin Isaacs

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

The views expressed in this presentation are those of the authors and do not necessarily reflect the views or policies of the U.S. EPA

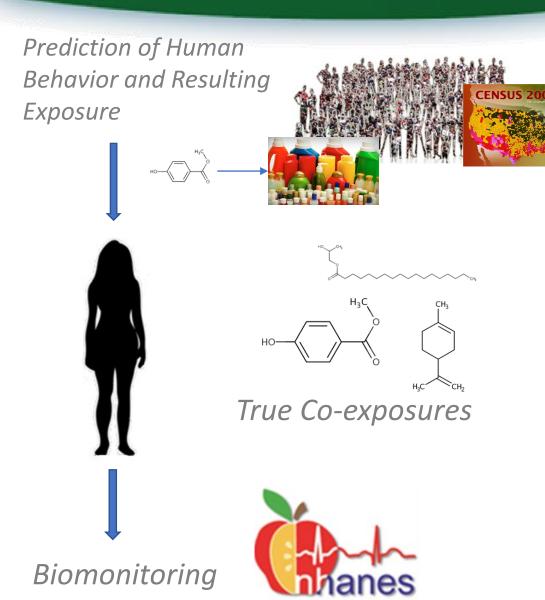
Introduction



- 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





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





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

0.3770

0.3434 ♀

0.3432

0.3432

0.3430

0.3409

0.3342

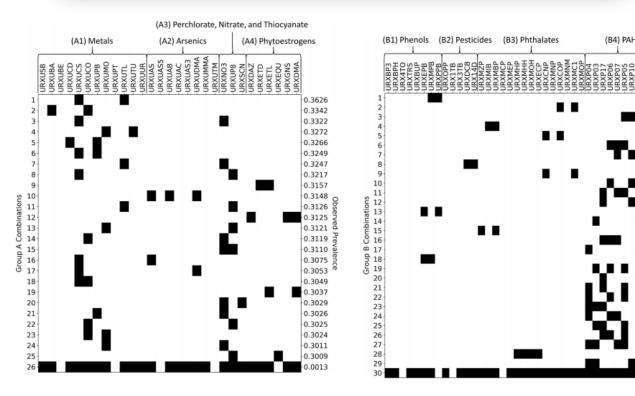
0.3309

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

Dustin F. Kapraun, John F. Wambaugh, Caroline L. Ring, 1-2 Rogelio Tornero-Velez, 3 and R. Woodrow Setzer1

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

³National 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



Transaction Dataset
Chemicals introduced to Household
Hypothesized Co-Exposure

{Chemical 1, Chemical 2, Chemical 3, Chemical X, Chemical Z.....}

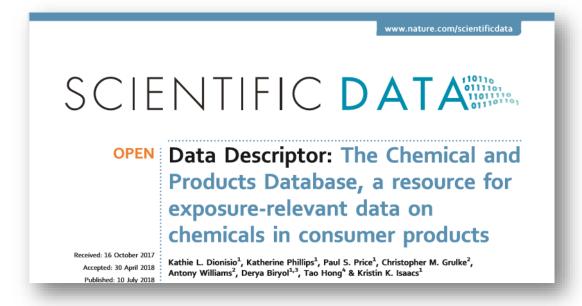
Frequent Itemset Mining



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



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

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)





Consumer Product Purchasing Study

United States
Environmental Protection
Agency

- 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.

Data Integration



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		

Data Integration



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% purchases

Monthly Transaction Matrix

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

Data Integration



Product-Chemical Data

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% purchases

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

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

Frequent Itemset Mining



•	Ite	m	ς	ρ.	t
	ונכ		3	L	L

- 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. $\sigma(\{DTXSID1, DTXSID4, DTXSID5\}) = 2/5$
- 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

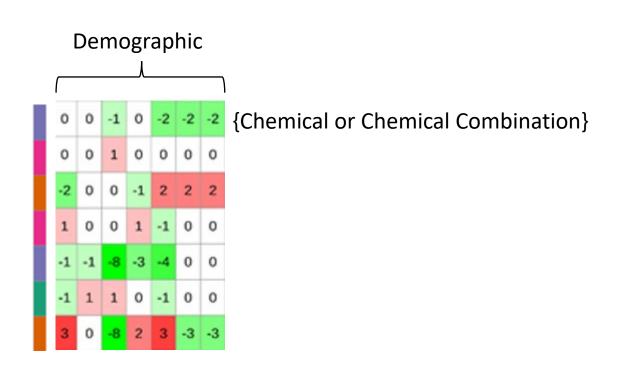
Apply to transaction data to identify prevalent combinations

Frequent Itemset Mining Analyses

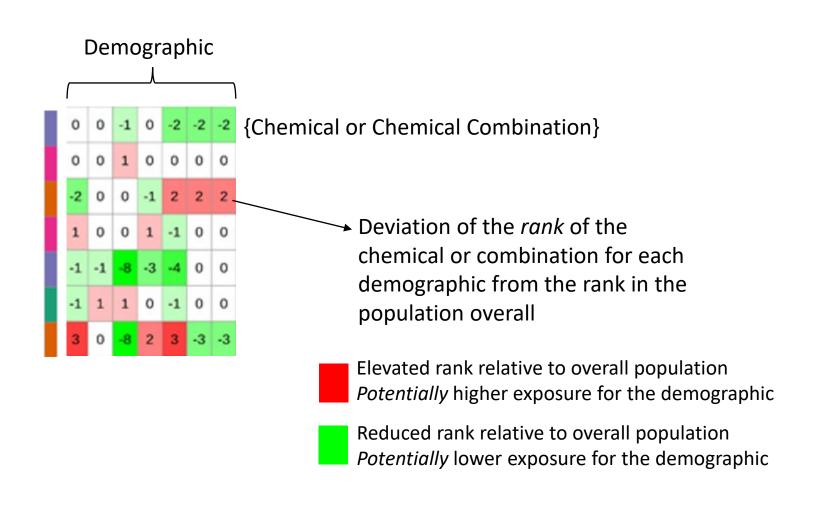


- Performed using the *ECLAT* (Equivalence Class Clustering and bottom-up Lattice Traversal) function of the *Arules* R package
- Performed identification of prevalent individual chemicals and combinations
 - For TSCA chemicals and EACs, based on a threshold prevalence for chemical group that provided 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

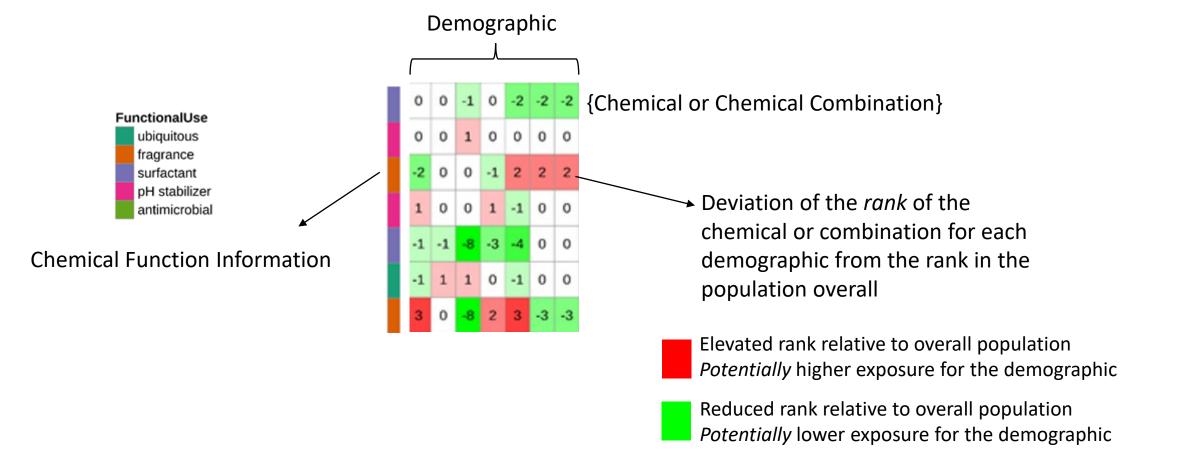




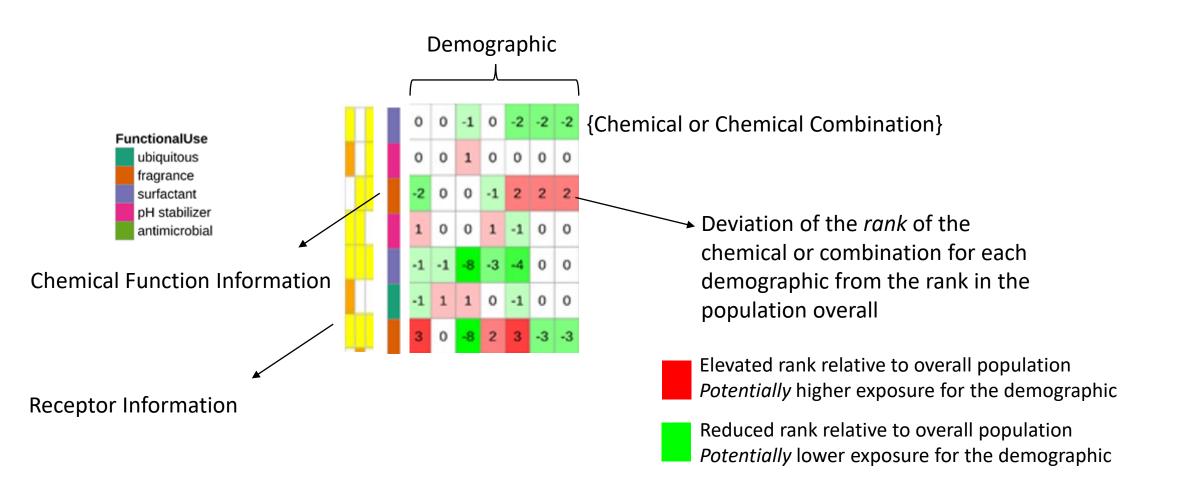


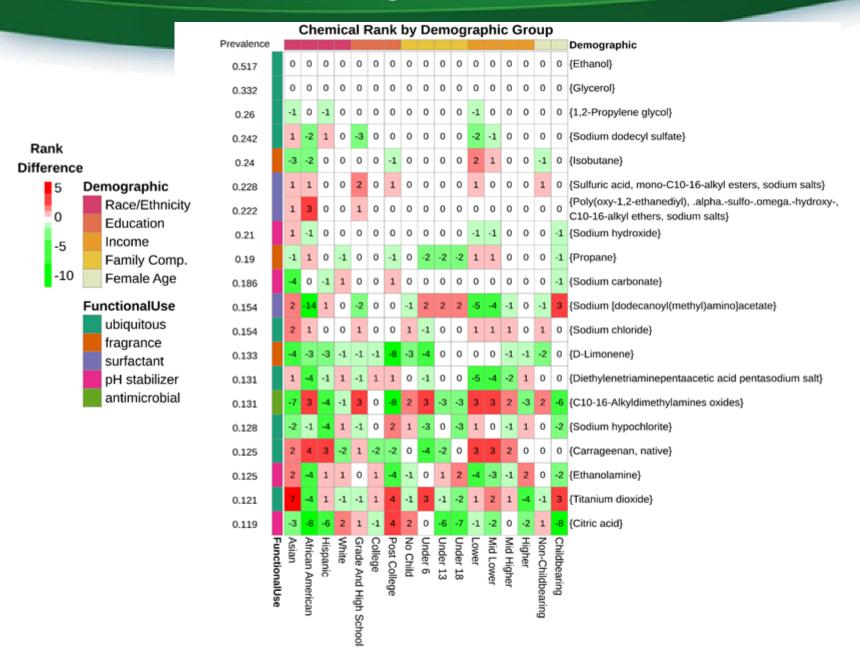








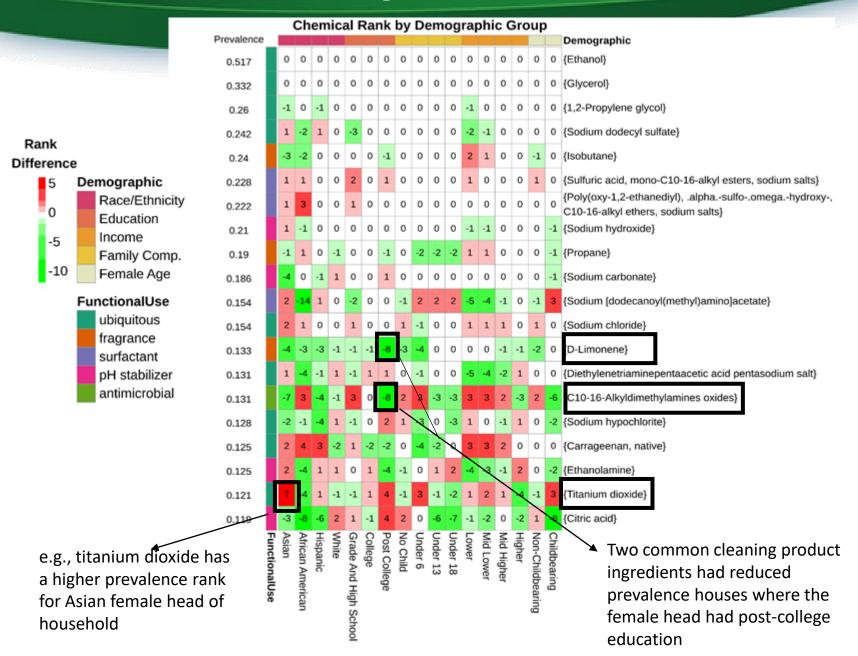






Group 1 (Broad TSCA Inventory)

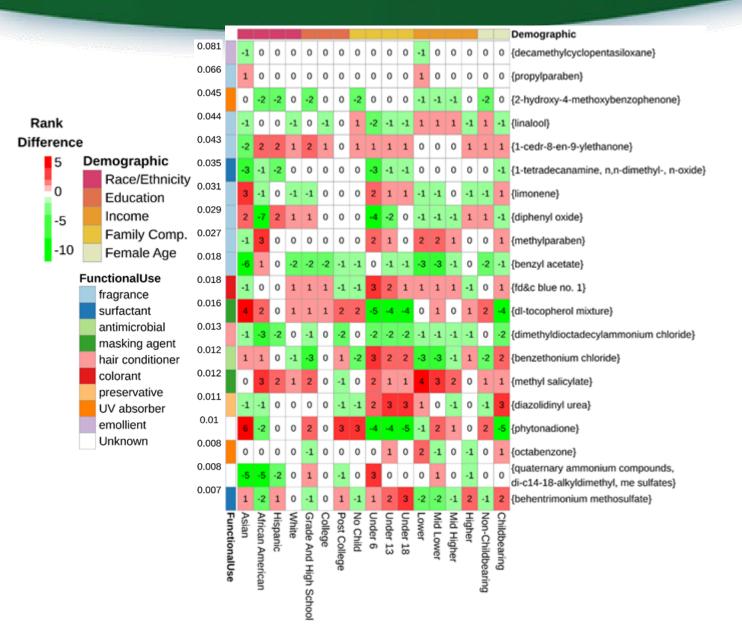
- 20 overall most prevalent individual chemicals
- Top 5 chemicals are what were termed "ubiquitous function" chemicals perform a variety of functions in products





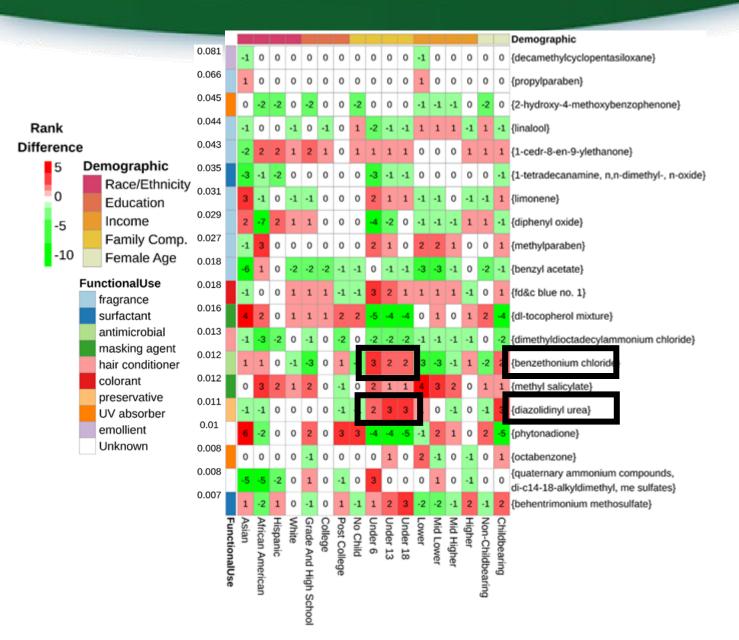
Group 1 (Broad TSCA Inventory)

- 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





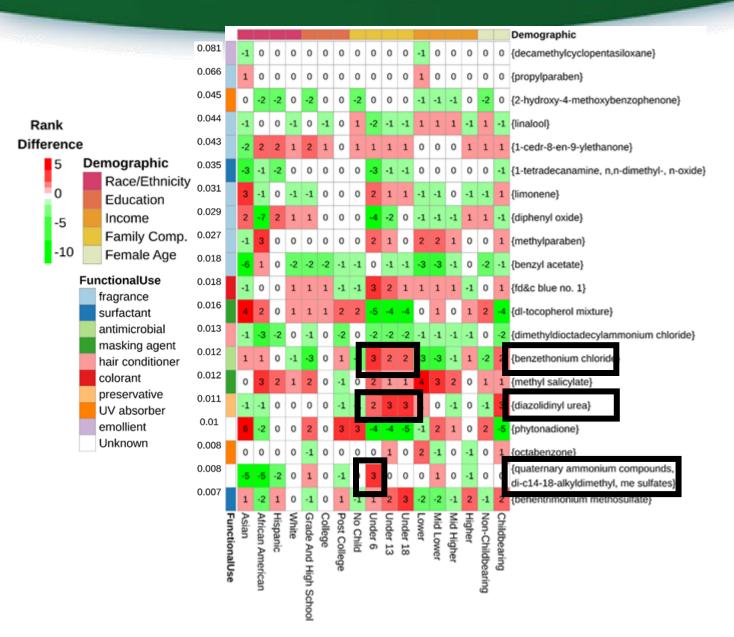
- Many of the most prevalent EAC chemicals were fragrances (or categorized as such due to presence in fragrance formulations)
- Many of these chemicals were present in a variety of personal care products





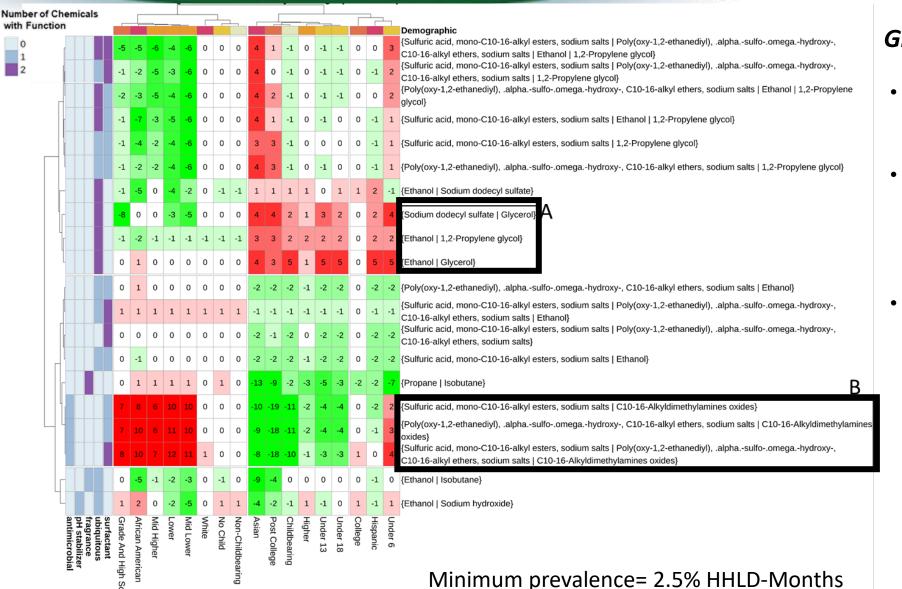
- Many of the most prevalent EAC chemicals were fragrances (or categorized as such due to presence in fragrance formulations)
- Many of these chemicals were present in a variety of personal care products
- Benzethonium chloride and diazolidinyl urea, which were ranked 2 or 3 places higher in households with children, are commonly used as topical antimicrobial agents in baby wipes, bubble baths, cosmetics, and skin care products





- 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

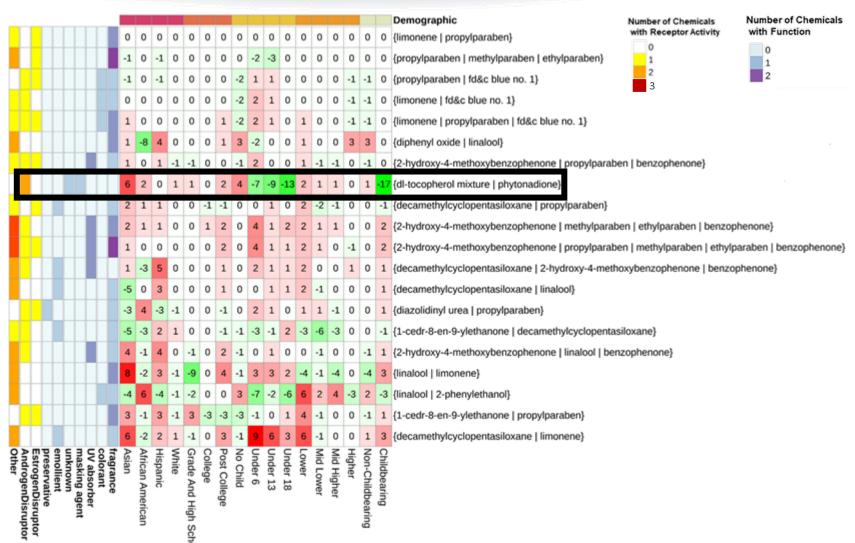




Group 1 (Broad TSCA Inventory)

- 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
- 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

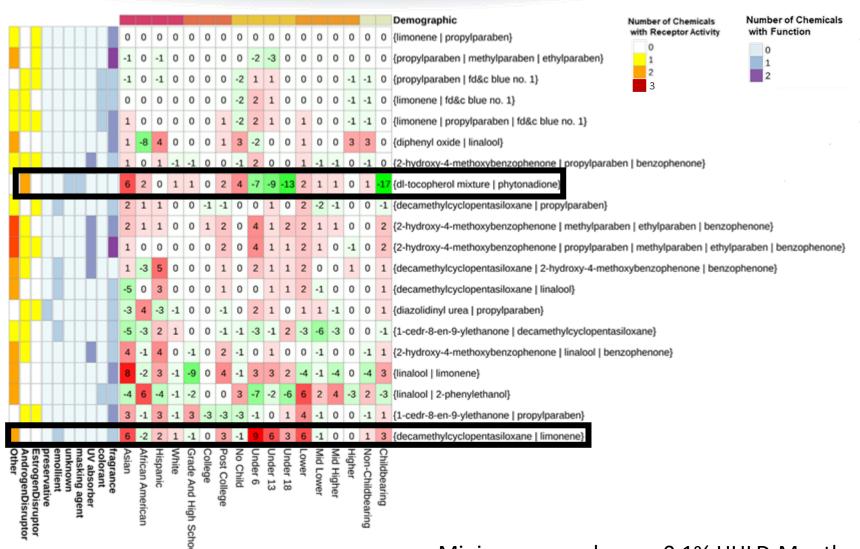




Endocrine Active Chemicals

• One itemset {dl-tocopherol mixture | phytonadione}, contained two chemicals that targeted the same receptor (AR).



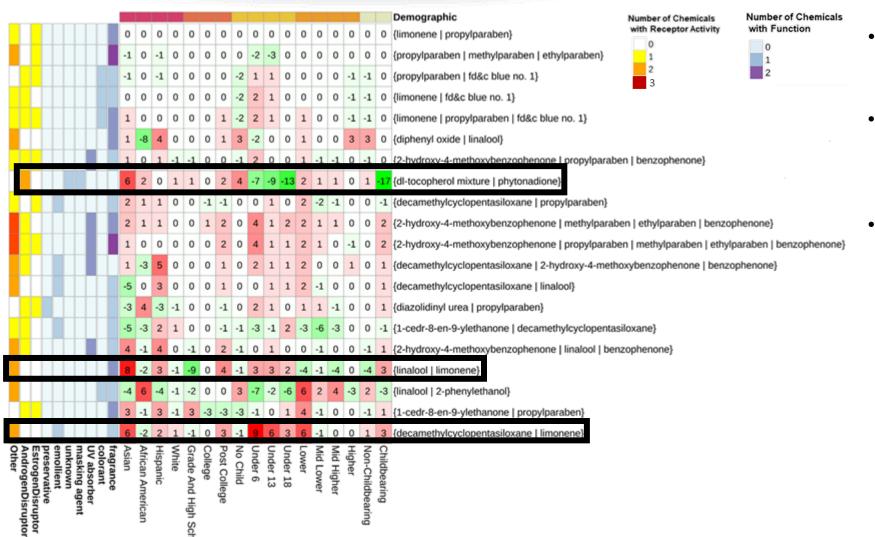


Endocrine Active Chemicals

- One itemset {dl-tocopherol mixture | phytonadione}, contained two chemicals that targeted the same receptor (AR).
- The highest positive rank departure for households with children occurred for the itemset {decamethylcyclopentasiloxane | limonene}.

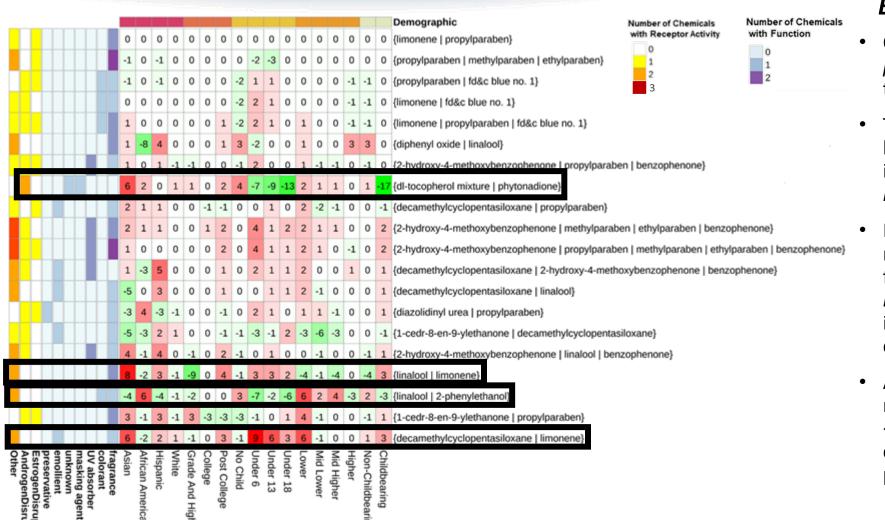
Minimum prevalence= 0.1% HHLD-Months





- One itemset {dl-tocopherol mixture | phytonadione}, contained two chemicals that targeted the same receptor (AR).
- The highest positive rank departure for households with children occurred for the itemset {decamethylcyclopentasiloxane | limonene}.
- Households with a female head of Asian race had 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.





- One itemset {dl-tocopherol mixture | phytonadione}, contained two chemicals that targeted the same receptor (AR).
- The highest positive rank departure for households with children occurred for the itemset {decamethylcyclopentasiloxane | limonene}.
- Households with a female head of Asian race had 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.

Testing Chemical Mixtures



- Use existing ToxCast data and amenability to screening as pilot information for mixture design
- Construct chemical mixtures (each with 3 chemicals) and screen each constituent
- General hypothesis: the point-ofdeparture for mixture bioactivity can be predicted from the concentration response data for the chemical constituents using models of concentration addition, independent action, or a joint model

Useful publication and approach that explored similar analyses of nuclear receptor and cell stress data

Environmental Toxicology and Chemistry—Volume 39, Number 12—pp. 2552–2559, 2020

Received: 23 June 2020 | Revised: 23 July 2020 | Accepted: 30 August 2020

2552

Hazard/Risk Assessment

Exploring the Concepts of Concentration Addition and Independent Action Using a Linear Low-Effect Mixture Model

Beate Escher, a,b,* Georg Braun, and Christiane Zarflb

https://doi.org/10.1002/etc.4868

^a Helmholtz Centre for Environmental Research–UFZ, Leipzig, Germany

^bCenter for Applied Geoscience, Eberhard Karls University of Tübingen, Tübingen, Germany

Summary



- 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.



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 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 lan 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