

## Mix Masters: Using a New Tool to Identify Commonly Occurring Chemical Mixtures

Carrie Arnold

<https://doi.org/10.1289/EHP2325>

Modern industry and manufacturing rely on a roster of more than 84,000 chemicals,<sup>1</sup> many of which have received minimal study.<sup>2</sup> Through programs such as ToxCast™ and Tox21, researchers are working hard to create safety profiles for as many of these compounds as possible. But while the primary paradigm in toxicology remains studying exposures to individual agents, few chemical exposures occur in isolation, and it can be difficult to predict how chemical mixtures might affect health. In a new report in *Environmental Health Perspectives*, researchers describe a new tool called frequent itemset mining (FIM) to identify the mixtures to which people are most commonly exposed.<sup>3</sup>

There is a “pretty big data gap” between what we know happens in real life and what researchers study in the lab, says first author Dustin Kapraun, an applied mathematician at the U.S. Environmental Protection Agency (EPA) National Center for Environmental Assessment. A group of just 20 chemicals would yield more than 1 million possible combinations, Kapraun says. However, chemicals do not occur in random mixtures; some mixtures are much more likely to occur than others, and knowledge of those that occur most often may be the key to narrowing the research gap.<sup>4</sup>

To narrow down the list of possibilities, Kapraun and colleagues used FIM to analyze data from the 2009–2010 round of the National Health and Nutrition Examination Survey (NHANES). FIM was developed initially by marketing researchers to identify items that are frequently purchased together. These data allowed sellers to present products—for instance, on store shelves or in catalogs—in a way that capitalizes on typical shopping habits.<sup>5</sup>

In the current study, Kapraun and colleagues used FIM to identify chemical combinations that frequently occur together in humans as measured in NHANES samples of blood and urine. Each chemical detected in NHANES samples was equivalent to an “item” in FIM lingo, with each combination of chemicals occurring together in the same sample defined as an itemset.

The FIM algorithms identified 90 chemical combinations found in a minimum of 30% of the respondents. The most common chemical combinations included pairs of metals (such as lead and cadmium, and thallium and cesium), a trio of phytoestrogens associated with soy consumption (genistein, daidzein, and *O*-desmethylangolensin, a metabolite of daidzein), polyaromatic hydrocarbon metabolites, parabens, and caffeine. The researchers also identified supercombinations



NHANES data, although limited in some respects, do allow researchers to identify sets of chemicals that often occur together. This is important given that people are typically exposed not to one chemical at a time, but to mixtures—for example, in the multiple personal care products that may be used daily. Image: © Susan M. Booker.

comprising at least 20 chemicals of concern. These supercombinations are relatively rare, but they are worrisome because of the large number of compounds they represent.

Kapraun notes that the study's ability to define itemsets is limited by the NHANES data themselves. NHANES is not intended to measure all possible chemicals to which a person is exposed. Another consideration is that chemicals and metabolites in any individual's blood and urine at the time of sampling will not reflect previous exposures that are no longer be detectable. NHANES also does not measure all chemicals in all participants, nor does it collect urine samples for children under 6 years of age or blood samples for children under 12 years of age. Finally, NHANES does not link specific exposures with negative health outcomes, information that would be useful in helping toxicologists identify the particular chemical combinations worthy of more intensive study.

Nevertheless, this study provides a good basis for future endeavors, according to Michelle Embry, an environmental health scientist at the nonprofit Health and Environmental Science Institute, which addresses global health and environmental problems. "The number of possible chemical combinations is a mind-bending problem," she says. "We currently have a window to prioritize future needs and begin to understand how risk assessments for mixture exposure may be different from those of individual chemicals." Embry was not involved with the study.

To truly understand the impact of chemical combinations, researchers also will have to factor in the amounts of the chemicals

to which individuals were exposed and the sequence of the exposures, says Margaret MacDonell, an environmental health scientist at the Argonne National Laboratory, who also was not involved in the study. Scientists can also integrate these results with data from research on toxicant action mechanisms, the microbiome, and diet. MacDonell says, "Improvements in technology have made this the perfect time to study chemical mixtures."

---

**Carrie Arnold** is a freelance science writer living in Virginia. Her work has appeared in *Scientific American*, *Discover*, *New Scientist*, *Smithsonian*, and more.

## References

1. Institute of Medicine, Roundtable on Environmental Health Sciences, Research, and Medicine, Board on Population Health and Public Health Practice. 2014. The challenge: chemicals in today's society. In: *Identifying and Reducing Environmental Health Risks of Chemicals in Our Society: Workshop Summary*. Washington, DC:National Academies Press.
2. Judson R, Richard A, Dix DJ, Houck K, Martin M, Kavlock R, et al. 2009. The toxicity data landscape for environmental chemicals. *Environ Health Perspect* 117(5):685–695, PMID: [19479008](https://pubmed.ncbi.nlm.nih.gov/19479008/), <https://doi.org/10.1289/ehp.0800168>.
3. Kapraun DF, Wambaugh JF, Ring CL, Tornero-Velez R, Setzer RW. 2017. A method for identifying prevalent chemical combinations in the U.S. population. *Environ Health Perspect* 125(8):087017, PMID: [28858827](https://pubmed.ncbi.nlm.nih.gov/28858827/), <https://doi.org/10.1289/EHP1265>.
4. Tornero-Velez R, Egeghy PP, Cohen Hubal EA. 2012. Biogeographical analysis of chemical co-occurrence data to identify priorities for mixtures research. *Risk Anal* 32(2):224–236, PMID: [21801190](https://pubmed.ncbi.nlm.nih.gov/21801190/), <https://doi.org/10.1111/j.1539-6924.2011.01658.x>.
5. Borgelt C. 2012. Frequent item set mining. *Wiley Interdiscip Rev Data Min Knowl Discov* 2(6):437–456, <https://doi.org/10.1002/widm.1074>.