Mining Human Biomonitoring Data to Identify Prevalent Chemical Mixtures

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

[1]National Center for Computational Toxicology, US Environmental Protection Agency, Research Triangle Park, NC, USA

[2]Oak Ridge Institute for Science and Education, Oak Ridge, Tennessee, USA

[3]National Exposure Research Laboratory, US Environmental Protection Agency, Research Triangle Park, NC, USA

Through food, water, air, and consumer products, humans are exposed to tens of thousands of environmental chemicals, and most of these have not been evaluated to determine their potential toxicities. In recent years, high-throughput screening (HTS) methods have been developed that begin to address the need for more efficient toxicity assessment, but testing still tends to focus on individual chemicals. Meanwhile, the exposures that people actually experience involve mixtures of chemicals. Unfortunately, the number of mixtures that can be formed from the thousands of environmental chemicals is huge, and testing all of them would be impossible, even using HTS techniques. We therefore need tools for identifying those mixtures that are most relevant. By applying frequent itemset mining (FIM), a technique traditionally used for finding patterns in consumer purchasing behavior, to data from the 2009-2010 cycle of the National Health and Nutrition Examination Survey NHANES, we were able to identify 90 prevalent chemical combinations consisting of 4 or fewer chemicals that occur in at least 30% of the US population, as well as 3 combinations consisting of 24 or more chemicals that occur in a small but non-negligible proportion of the US population. We have thus demonstrated a method for reducing a large set of possible chemical combinations to a much smaller set of prevalent chemical combinations. *The views expressed herein are those of the authors and do not necessarily reflect the views or policies of the US EPA.*