



www.epa.gov

The Residential Population Generator (RPGen) - parametrization of residential, demographic, and physiological data to model intraindividual exposure, dose, and risk.

Alexander East¹, Paul Price², Daniel Dawson², Graham Glen³, Katherine Dionisio², Kristin Isaacs², Elaine Cohen Hubal⁴, and Daniel Vallero²

¹Oak Ridge Associated Universities, assigned to U.S. Environmental Protection Agency, Office of Research and Development, Center for Computational Toxicology and Exposure, Research Triangle Park, NC 27709

²U.S. Environmental Protection Agency, Office of Research and Development, Center for Computational Toxicology and Exposure, Research Triangle Park, NC 27709

³ICF, Inc., Durham, NC 27713

⁴U.S. Environmental Protection Agency, Office of Research and Development, Center for Environmental Measurement and Modeling, Research Triangle Park, NC 27709

Abstract: The Residential Population Generator (RPGen) creates a population for use in intraindividual probabilistic models of exposure and dose by appending datasets. To do this, two requirements should be considered: the set of inputs used for an individual must be internally consistent (reflect data that could come from a single individual); and the variation in the data sets across individuals should be representative of variation among individuals in the modeled population. RPGen meets these requirements by creating synthetic populations with demographic, residential, and physiological characteristics. A database of modeled individuals is created by linking data from US census and US housing surveys with physiological characteristics from the National Health and Nutrition Examination Survey data. The final modeled population data parameters include characteristics of the individual's community residence demographics, and physiology. RPGen output can be used to support user-developed chemical exposure models that estimate intraindividual exposure in a desired population. By creating profiles and characteristics that determine exposure, RPGen allows modelers to generate data-driven populations and identify those potentially vulnerable to chemical exposures.

Alexander W. East | East.Alexander@epa.gov | 919-541-4064

Background

EPA's Office of Research and Development is developing the Combined Human Exposure Model (CHEM) to characterize chemicals exposures that occur as a result of the use of consumer products and support lifecycle impact assessments (Csiszar et al., 2016). In CHEM, individual exposure is characterized by the residence, activities, and concentration of chemicals in products. These three elements are handled in the Residential Population Generator (RPGen), Product Use Scheduler (PUS) and Source-to-Dose (S2D) respectively. As the first module, RPGen creates internally consistent records with demographic, housing, residential, and physiological data. In total, RPGen contains 126 characteristics, with records varying by age, gender, income and region, which reflect interindividual variability in exposure-related parameters. In CHEM, these parameters inform patterns of use in the Product Use Scheduler, but RPGen is transferrable to other models, tools, and dashboards. RPGen output informs the exposome, providing physiological variables and residential characteristics that characterize activities undertaken. By creating a consistent, granular population, RPGen creates an informative component for models of exposure, dose, and risk.

Methods

No single survey contains all parameters required for characterizing individuals in RPGen. Therefore, three surveys are linked using similar variables: the Residential Energy Consumption Survey (RECS), the American Housing Survey (AHS), and the Public Use Microdata Survey (PUMS). Additionally, the R Package *httk* (Pearce et al. 2017; Ring et al., 2017) is called, which generates physiological traits based on age and ethnicity inputs from PUMS. Data is updated manually as needed in the RPGen GitHub repository. Each input in Table 1 provides a different component to the population: RECS describes the residence, parameters in AHS describe the household, PUMS variables contain demographics, and NHANES data informs the physiological variables. Examples of variables from RPGen records are visible in Tables 3-6.

Table 1. Data sets used in this version of the RPGen. Note: NHANES data derived from EPA's <i>httk</i> .			
Survey	Source	Date of Data Collection	Number of Records
Residential Energy Consumption Survey (RECS)	US Energy Information Association: https://www.eia.gov/consumption/residential/data/2015/	2015	5,686
American Housing Survey (AHS)	US Census Bureau: https://www.census.gov/programs-surveys/ahs/data.html	2017	57,972
Public Use Microdata Survey (PUMS)	US Census Bureau: https://www.census.gov/programs-surveys/acs/data/pums.html	2014-2018	15,094,428
National Health and Nutrition Examination Survey (NHANES)	US CDC: https://wwwn.cdc.gov/nchs/nhanes/Default.aspx	2007-2008, 2009-2010, 2011-2012	24,546

Records are linked using similar parameters in each dataset. For this reason, both the 'person' and 'housing' versions of PUMS are imported into RPGen. A *pool* variable is created, with 288 possibilities, matching households from RECS and AHS with individuals in PUMS (Table 2).

Table 2. Assembly of 288 Possible Household Bins (<i>pool</i>).				
Setting	Region	House Type	Family Category	Income Category
Urban	Northeast	Stand Alone	1 Adult, 0 Children	1
Rural	Midwest	Multi Structure	2+ Adults, 0 Children	2
	South	Other	1 Adult, 1+ Children	3
	West		2+ Adults, 1+ Children	

Location is determined by crossing an either *rural* or *urban* setting with one of four US regions. Houses are grouped as stand-alone, multi-structure (e.g. apartments and condominiums), or other (e.g. mobile homes). Family category depends on the number of adults and children in the household, as co-inhabitants inform interindividual exposure. Income category rank is calculated for each of the eight *location* outcomes, as purchasing power varies between *setting* and *region*. When running RPGen, the user selects the number of individuals, demographics, age ranges, states, and a seed value for replicability.

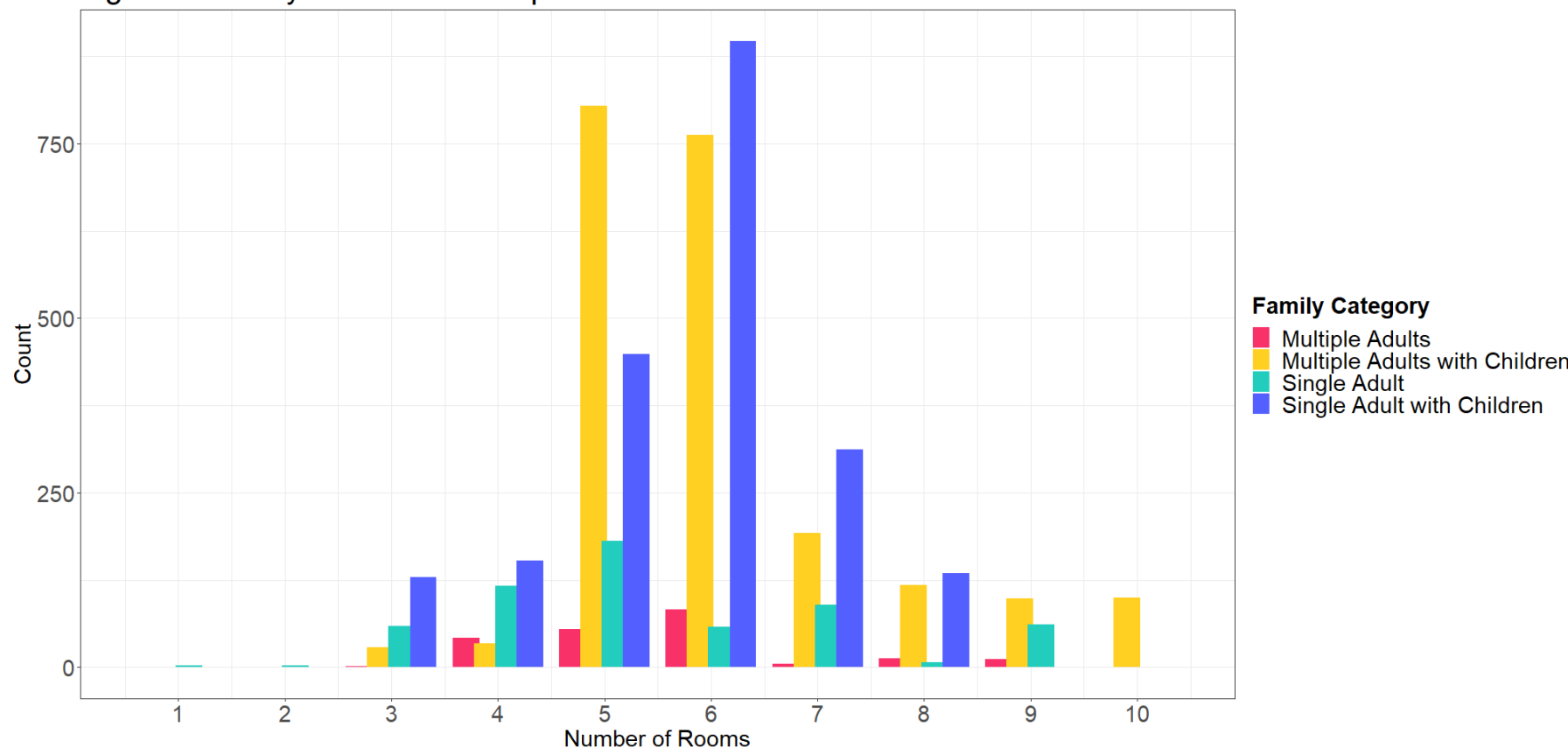
U.S. Environmental Protection Agency
Office of Research and Development

Although this work was reviewed by EPA and approved for presentation, it may not necessarily reflect official Agency policy.

RPGen Results for 5,000 Individuals

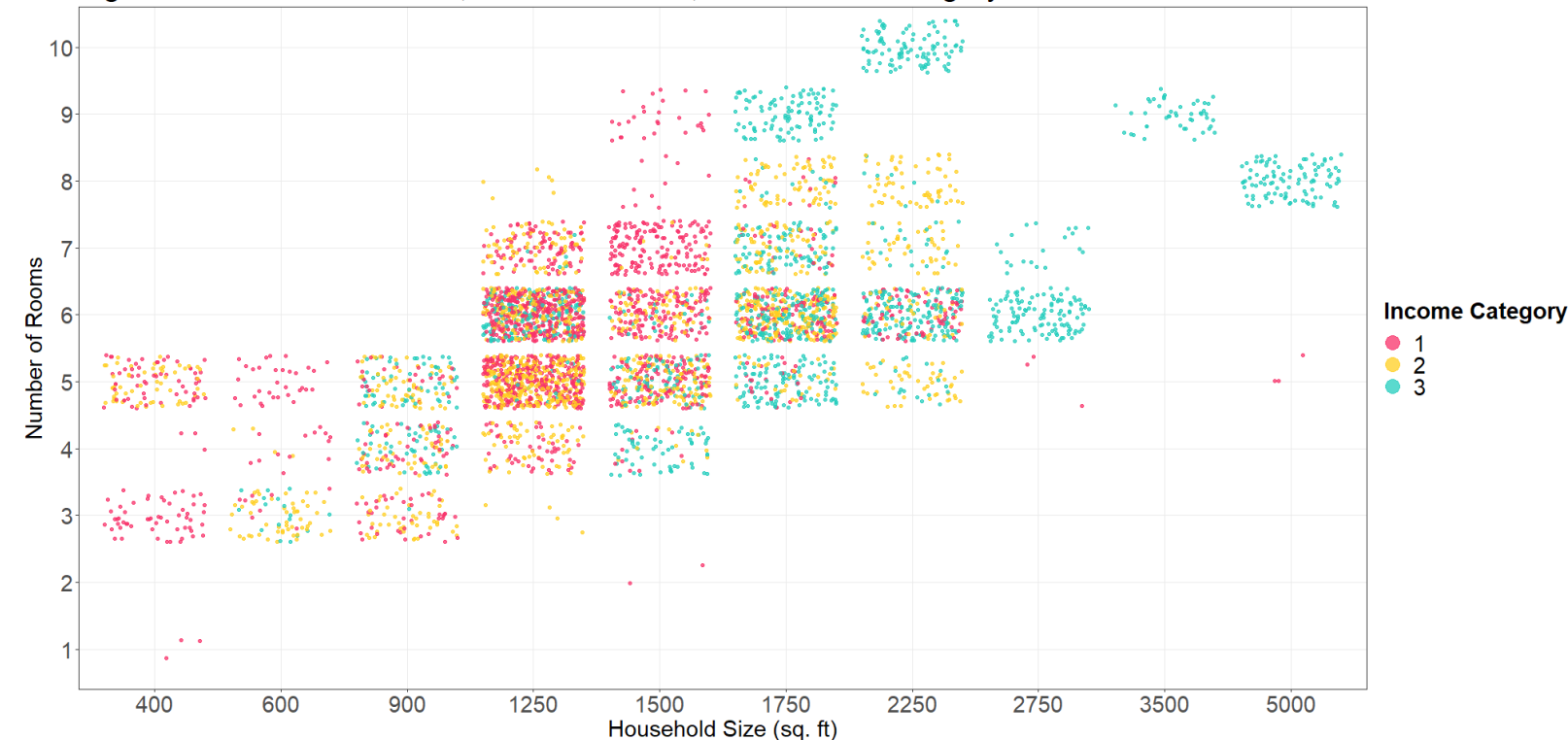
A nationwide sample run with all genders, ethnicities, and ages with 5,000 individuals was performed to illustrate the population generated by RPGen. Creation of a household is imperative in modeling interindividual exposure, or exposure between individuals. Therefore, family category serves as an input to *pool* as well as a descriptor of residences. As visualized in Figure 1, distributions of family category vary across the number of rooms.

Figure 1. Family Residence Composition and Number of Rooms



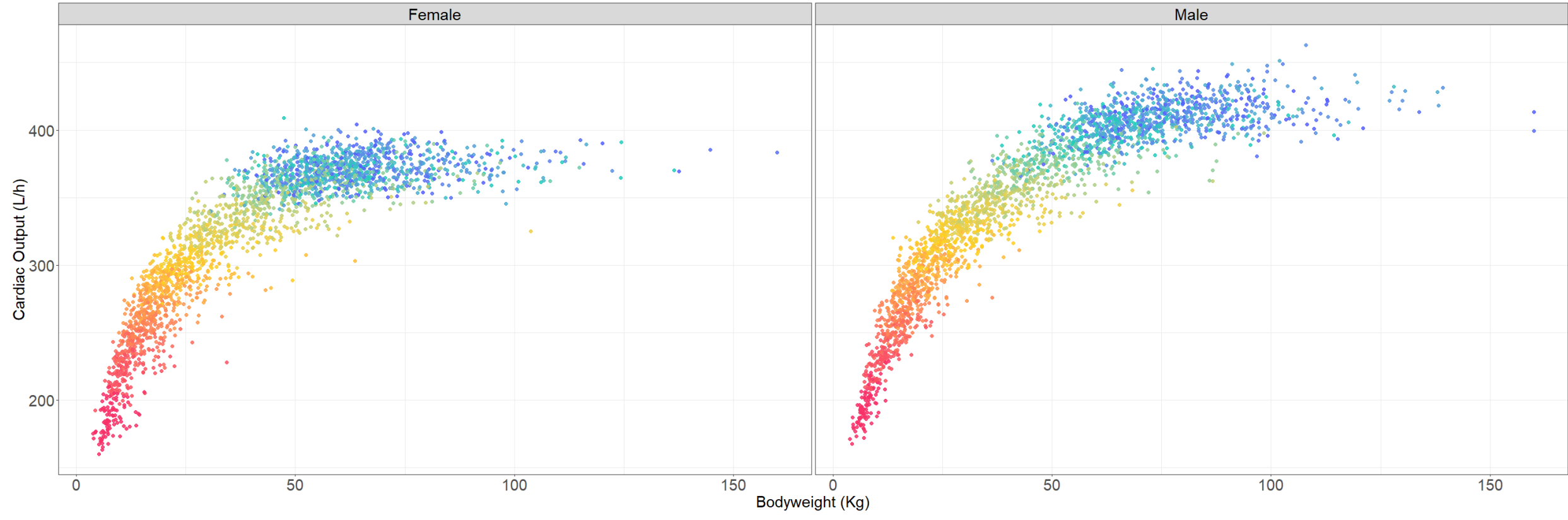
The largest households, with the greatest number of rooms, are occupied by those in the highest income category (Figure 2). Points in Figure 2 are jittered to show the density of each household size and number of room outcome, representing the likelihood of each occurrence. More 6 room 1250 square foot residences were generated than any other pairing.

Figure 2. Number of Rooms, Household Size, and Income Category Distribution in RPGen



As children have been identified as a particularly vulnerable subgroup (Egeghy et al., 2011), another sample run of 5,000 individuals was performed with ages restricted between 0 and 21. The distributions of bodyweight against cardiac output are different between males and females, and this is captured in RPGen (Figure 3). A random element is applied to physiological variables from *httk* to create the variation in Figure 3.

Figure 3. Cardiac Output and Bodyweight by Gender in Adolescents



RPGen Output for Four Individuals

Tables 3-6 present selected parameters from RPGen records for four example individuals of different ages and genders living in different types of homes, in different portions of the country, with different family categories. Data are color coded to reflect the survey that provides the data. PUMS—red, httk—yellow, AHS—teal, and RECS—purple.

Table 3. Sample of variables describing demographics (PUMS)							
Individual	Age (Years)	Age (Months)	State	Race	Gender	Household Income	Genders, Ages of co-inhabitants
Infant	0	7	Florida	White	Female	N/A	Male, 2, Female, 47, Male, 53
Teen	16	65	Mass.	Black	Female	N/A	Male, 15 Female, 45, Male, 45
Adult	47	564	Arkansas	White	Male	\$ 75,000	Male 70, Female, 41
Older Adult	78	909	Illinois	Asian	Male	\$ 2,000	Female, 70

Table 4. Sample of generated physiological variables (<i>httk</i>)						
Individual	Bodyweight (kg)	Height (cm)	Body Surface Area (cm ²)	Cardiac Output (L/h)	Liver Blood Flow (L/h)	Serum Creatinine (mg/dL)
Infant	3.97	65	2670	121	2.09	N/A
Teen	74.5	168	18800	248	54.1	0.64
Adult	78.1	178	19600	268	51.2	1.24
Older Adult	64.3	173	17600	227	48.8	1.14

Table 5. Sample of variables describing appliances (RECS)							
Individual	Dish Washer	Washing Machine	Dryer	Loads washed per week	Outdoor Grill	Ceiling Fans	Televisions
Infant	Yes	Yes	Yes	4	Yes	3	3
Teen	No	Yes	Yes	6	Yes	0	2
Adult	Yes	Yes	Yes	7	Yes	2	3
Older Adult	No	No	No	0	No	3	2

Table 6. Sample of variables describing residence (AHS, RECS, PUMS)							
Individual	Size of Unit (ft²)	Number of Rooms	Garage	Cellar	Stories	Windows	Cars
Infant	1250	7	N/A	N/A	N/A	3 to 5	1
Teen	1250	6	0	0	2	10 to 15	2
Adult	2750	7	1	1	1	10 to 15	3
Older Adult	600	5	N/A	N/A	N/A	6 to 9	2

Discussion

RPGen output can be combined with a wide range of user-developed chemical exposure scenarios to estimate intraindividual exposure in a desired population. However, as RPGen was designed as a module in CHEM, a model for consumer products, only breathing rates (inhalation) and skin surface area (dermal) are included. Therefore, parameters that describe the ingestion of liquids and diet constructions are not included in RPGen. Furthermore, households are matched in eight possible location bins across the contiguous US. Therefore, households lack the geographic resolution that are afforded to individuals, as each PUMS record has as assigned Public Use Microdata Area (PUMA). Nevertheless, the pooled RPGen profiles offer increased granularity when compared with point estimates of exposure factors or distributions of exposure factors. This is due to RPGen's ability to define the exposome for each unique individual as opposed to assuming a collection of probabilistic exposure outcomes. The benefits are twofold: first, chemical exposure predictions to are informed by the correlated physiological, demographic, and housing variables within RPGen. Grouping of these variables into profiles improve the exposure estimates of individuals. Secondary benefit comes from the same variables: age, ethnicity, income, and location allow for comparison with dose outcomes to determine high risk subpopulations and communities.

References

Csiszar, S. A., Meyer, D. E., Dionisio, K. L., Egeghy, P., Isaacs, K. K., Price, P. S., . . . Vallero, D. (2016). Conceptual framework to extend life cycle assessment using near-field human exposure modeling and high-throughput tools for chemicals. *Environmental science & technology*, 50(21), 11922-11934.

Egeghy, P. P., Cohen Hubal, E. A., Tulve, N. S., Melnyk, L. J., Morgan, M. K., Fortmann, R. C., & Sheldon, L. S. (2011). Review of pesticide urinary biomarker measurements from selected US EPA children's observational exposure studies. *International journal of environmental research and public health*, 8(5), 1727-1754.

Pearce, R. G., Setzer, R. W., Strope, C. L., Wambaugh, J. F., & Sipes, N. S. (2017). Httk: R package for high-throughput toxicokinetics. *Journal of statistical software*, 79(4), 1.

Ring, C. L., Pearce, R. G., Setzer, R. W., Wetmore, B. A., & Wambaugh, J. F. (2017). Identifying populations sensitive to environmental chemicals by simulating toxicokinetic variability. *Environment international*, 106, 105-118.