The Residential Population Generator (RPGen): A tool for creating internally consistent populations and households.

Alexander East
ORAU/EPA
7/27/2021

Outline

1. How RPGen works
2. Introduction of Homeowners Case Study
3. Conclusion

Description of RPGen

What is RPGen?

Written entirely in R, RPGen is a module that appends national surveys to create matched households and individuals for models of human exposure.

Image: https://www.rstudio.com/products/rstudio/

User Input

The user may use a .txt input file or enter the commands directly into the console.

```
 HEM population generator module
Name for this run: userguide
Number of simulated persons: 1000
Minimum age in this run (years): 0
Maximum age in this run (years): 99
Include males (y/n): y
Include females (y/n): y
Include non-Hispanic (y/n): y
Include Mexican-American (y/n): y
Include Other Hispanic (y/n): y
Include White persons (y/n): y
Include African-Americans (y/n): y
Include Native Americans (y/n): y
Include Asian Americans (y/n): y
Include Pacific Islanders (y/n): y
Include multiple/other races (y/n): y
Initial random number seed (1-2147483646): 12345
Enter list of region codes:1234
Enter list of state FIPS codes:
Saved run settings
```

Surveys Appended

Three surveys are combined to create the population.

Survey	Source	Data Collection Year	Number of Records	Release Period
Residential Energy Consumption Survey (RECS)	US Energy Information Association	2015	5,686	4 years
American Housing Survey (AHS)	US Census Bureau (AHS)	2017	57,972	2 years
Public Use Microdata Survey (PUMS)	US Census Bureau (PUMS)	2014-2018	15,094,428	Annual

Combination of Survey Data

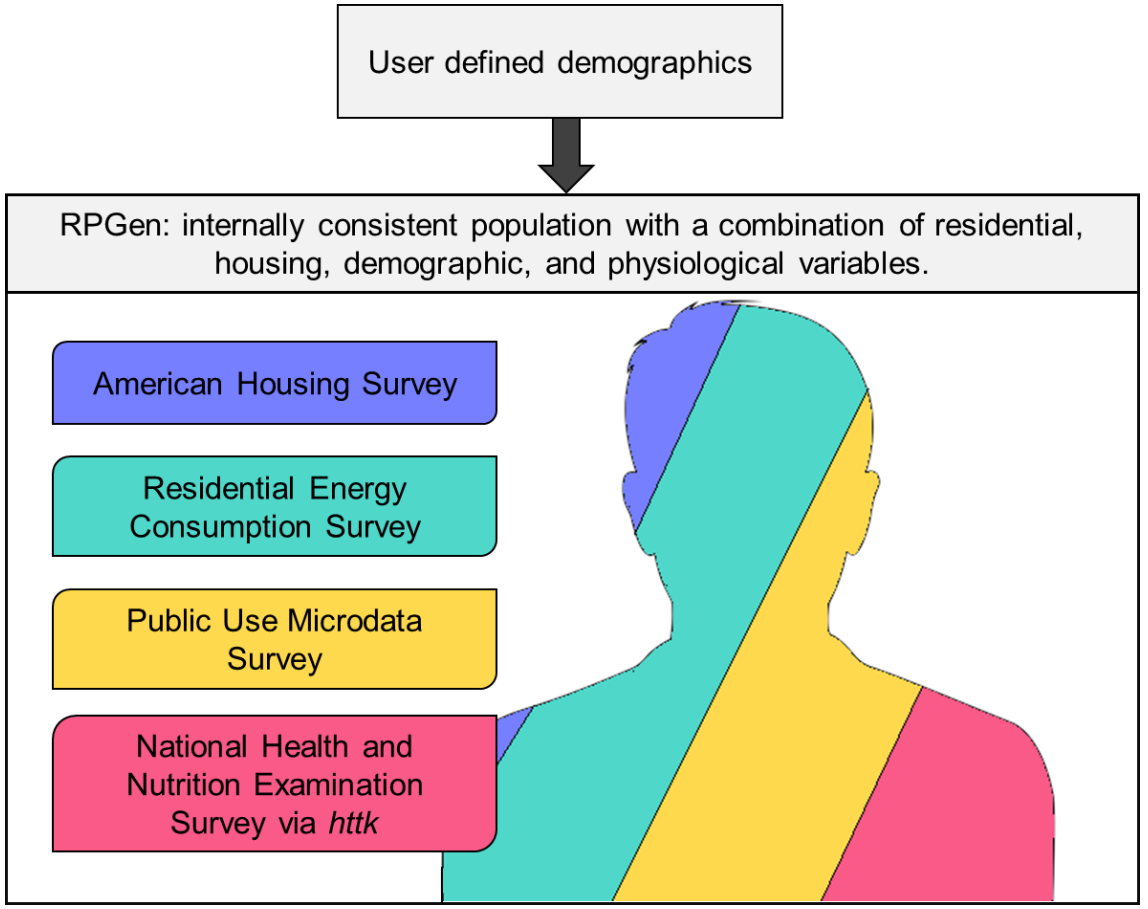
Data inputs are combined using similar variables to create *pools*. Each household in RECS, AHS, and PUMS are assigned a pool between 1-288 and are randomly paired to matching *pools*.

2 (location type) * 4 (region) * 3 (house type) * 4 (family category) * 3 (income category) = 288

288 Possible *Pool*s generated using similar variables in each input dataset to RPGen.				
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	

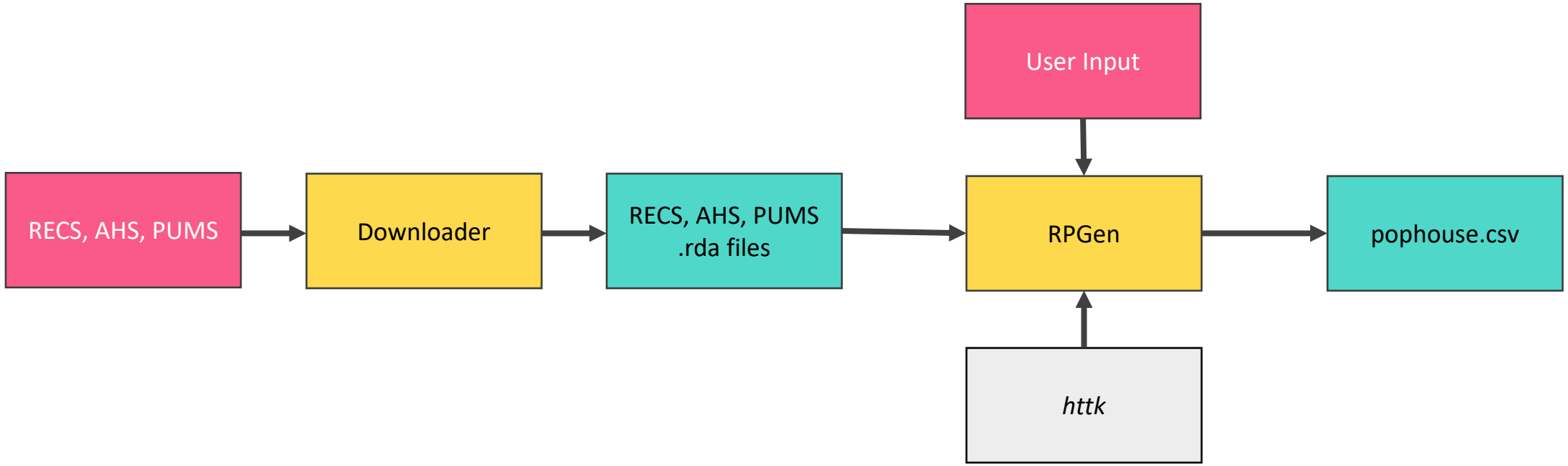
Addition of Physiological Variables

After matching households (RECS/AHS) and people (PUMS), physiological characteristics are added using *httk* (Ring et al. 2017).



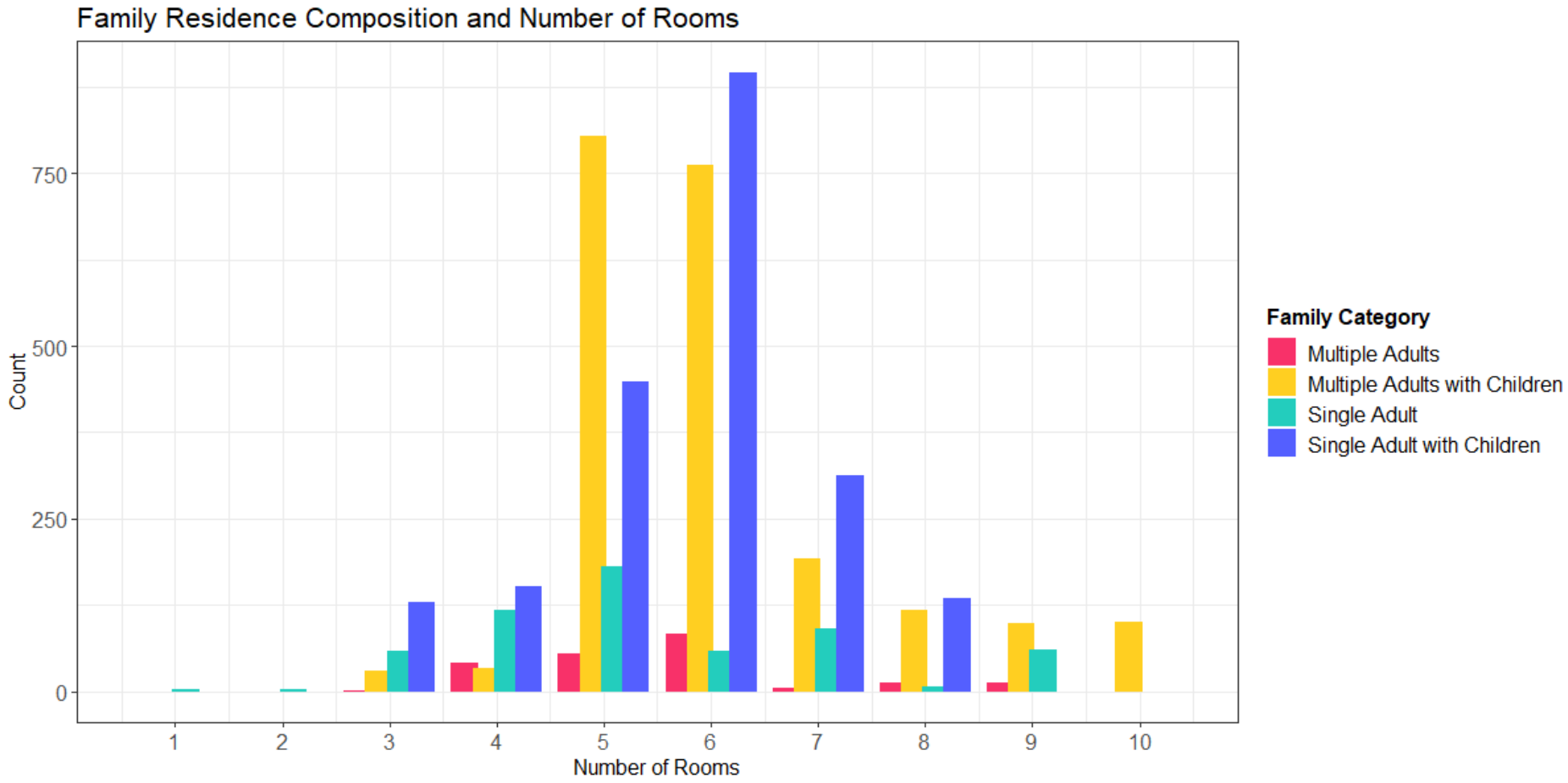
R Process Flow

RPGen has a ‘data downloader’ designed to pull the most recent data when ran. httk is an R package.



Consistent Households

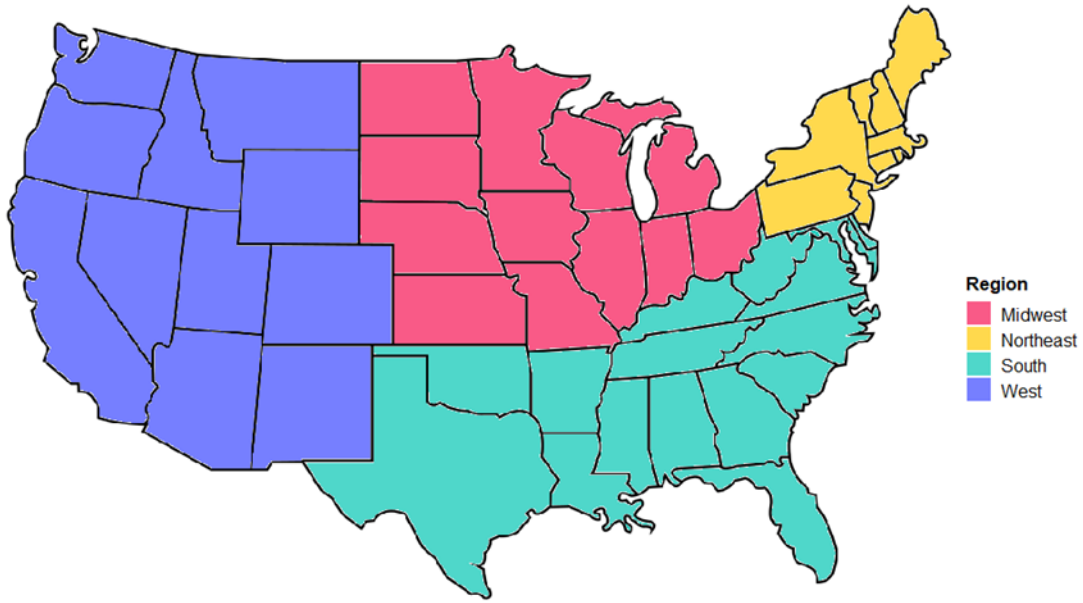
- National Sample Run of 5,000 Individuals (All genders, races, ages)
- Households are internally consistent.



Work in Progress

- Households do not exist for all 288 pools, meaning some scenarios are not included (rural, high-income apartments)
- Households are less granular than people as they are defined only by urban/rural and regional categories.

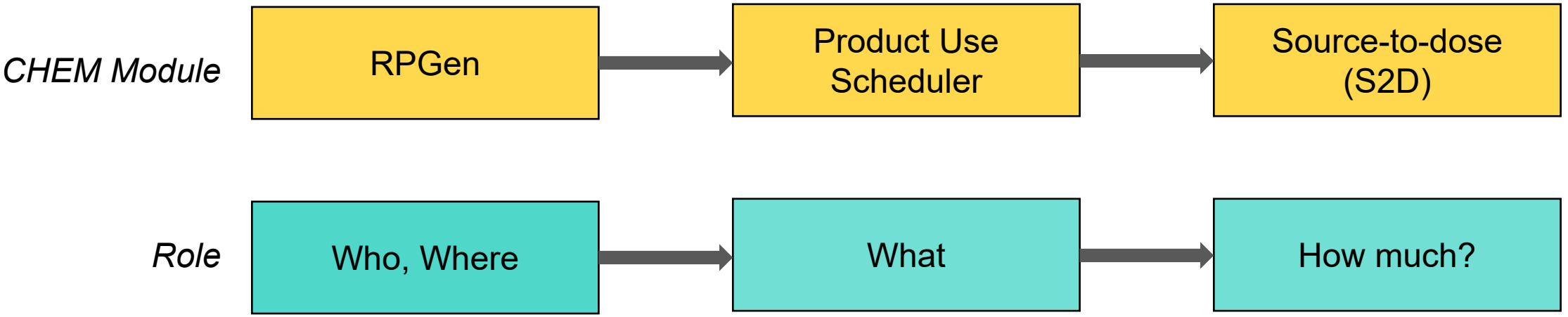
285	Urban West Other House type Adults High Income
286	Urban West Other House type Adults with Kids Low Income
287	Urban West Other House type Adults with Kids Middle Income
288	Urban West Other House type Adults with Kids High Income



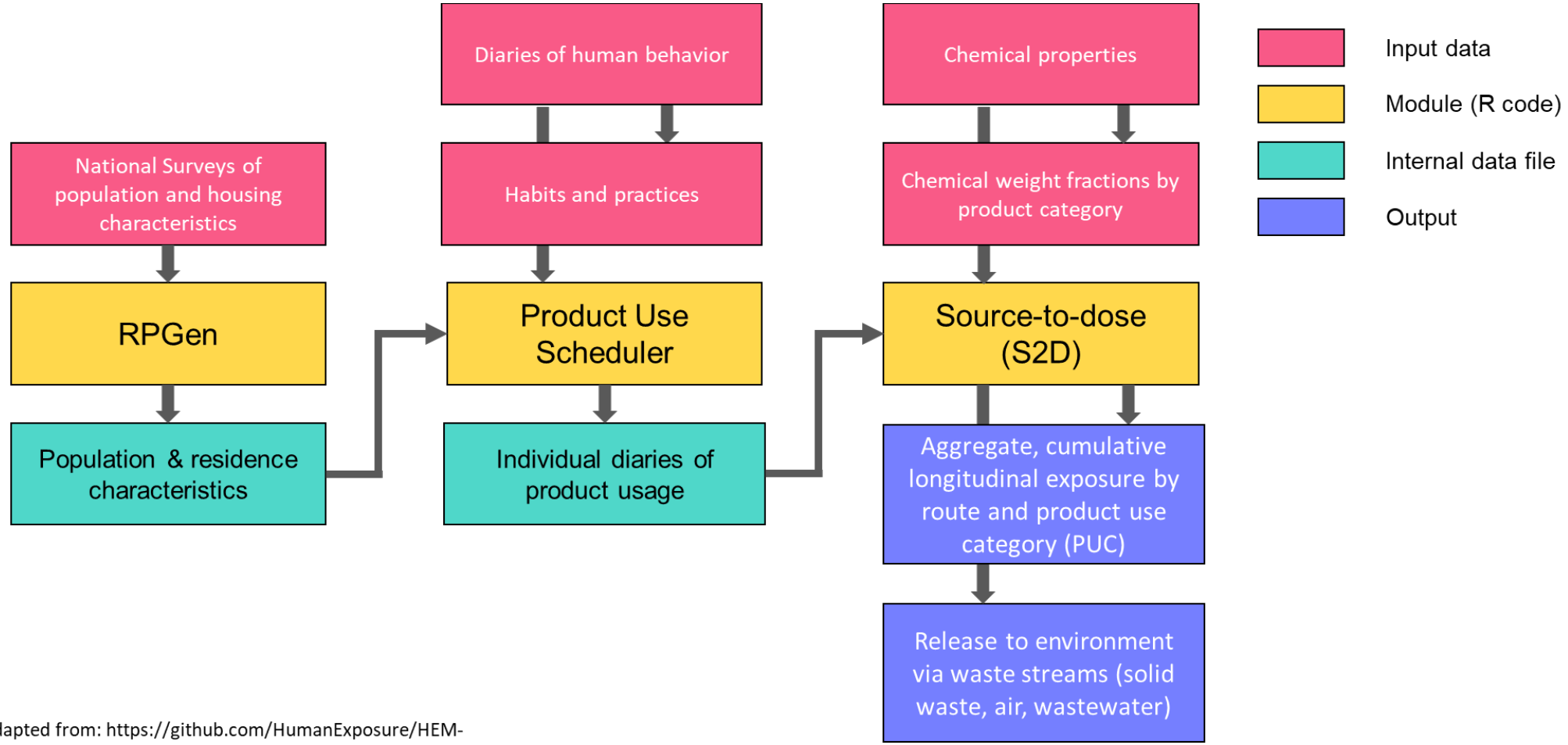
RPGen Case Study

RPGen in CHEM

- The Combined Human Exposure Model (CHEM) is a three-part R Model which estimates intake, dose, and down-the-drain for exposure to human products.



RPGen in CHEM



adapted from: https://github.com/HumanExposure/HEM-Documentation/blob/master/HEM_UserGuide_2018August28.pdf

Case Study: Homeowners

- RECS carries a variable called *kownrent*, which determines if a residence is owned or rented.
- This is an RPGen household characteristic that influences product use.

- In the Product Use Scheduler (PUS) these are called *ever/never* rules.
- *ever/never* rules include swimming pools, washing machines, septic tanks, or presence of a car, printer, or front yard.

Summary of Renters Input File

- RPGen was ran for all states, ages, ethnicities, and genders for 1,000 individuals.
- 659 live in an owned household
- 341 live in a rented house or stay without rent

gender	reth	pool	compid	race	ethnicity	age	pwgtp	ages	genders	commute	cars	state	
Male	Non-Hispanic White	151	12500200	W	N	72	96	'697247'	FMF	20	1	25	
Male	Non-Hispanic White	57	10607313	W	N	55	49	'5824552124'	FFMFM		0	6	
Female	Non-Hispanic White	9	10501000	W	N	44	29	'395044'	FFF		2	5	
Female	Non-Hispanic Black	190	13401904	B	N	40	25	'7716093940'	FFFMF		2	34	
Female	Non-Hispanic White	175	12102300	W	N	21	30	'2152'	FF		1	21	
Male	Non-Hispanic Black	192	14101305	B	N	32	21	'020065327234'	MFFMMF	60	3	41	

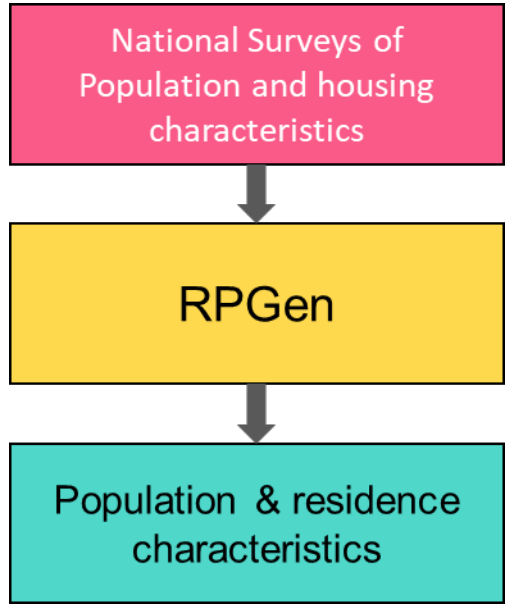
A subset of columns and rows from the RPGen output

Conclusion

Other Applications

- CHEM does a one-year longitudinal estimate of dose, exposure, and down-the-drain from consumer products.

- RPGen is a transferrable module available for different exposure models and applications.



Conclusion

- RPGen is software written in R that combines national surveys of the US population and the httk R package to create populations with residential, demographic, and physiological variables for use in models of exposure.
- The RPGen case study aims to quantify the exposure incurred by maintenance of an owned household.
- RPGen has application beyond human exposure to consumer products and can be implemented to EJ workflows.

Acknowledgements

EPA
Paul Price, Kathie Dionisio, Kristin Isaacs, Jeff Hollister, Dan Dawson, Elaine Hubal, Rogelio Tornero-Velez, Katherine Phillips, Alex Fisher, Dan Vallero

ICF
Graham Glen, Heidi Hubbard, Jessica Levasseur, Jeanne Luh