Developing a Predictive Model for Chemical Excretion in Urine

Zachary Stanfield, R. Woodrow Setzer, Jon R. Sobus, John F. Wambaugh

Center for Computational Toxicology and Exposure, Office of Research and Development , U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, United States

Abstract: 3818
Poster: P307

Zachary Stanfield | Stanfield.zachary@epa.gov | 919-541-5780

Background

- Most pharmacokinetic models adjust urine biomarker levels using creatinine correction, which assumes clearance by glomerular filtration.
- Two potential limitations of this approach are: 1) since creatinine excretion is not constant, creatinine concentration (conc.) is not exactly (inversely) proportional to urine output, and 2) urinary conc. of chemicals excreted (at least partially) by passive diffusion is independent of urine output and should not be corrected.
- Models for predicting primary route of urine clearance for xenobiotics are not yet known to exits.
- There are 3 known elimination process for chemicals in urine: glomerular filtration, passive diffusion, and active transport. However, we currently cannot model active transport and thus ignore it here.
- Glomerular filtration and diffusion/transport across the proximal tubules are competing mechanisms for elimination from the kidney into the urine. These are the 2 major routes we consider here.
- Kidney physiology leads us to expect 1) the conc. of chemicals undergoing glomerular filtration follows creatinine conc. directly and urine output inversely, and 2) the conc. of chemicals undergoing passive diffusion should be independent of creatinine conc. and urine output.
- No broadly accepted guidance on how to treat urine data currently exists, meaning a predictive model for primary route of urine clearance may serve as a first step toward developing that guidance.

Methods

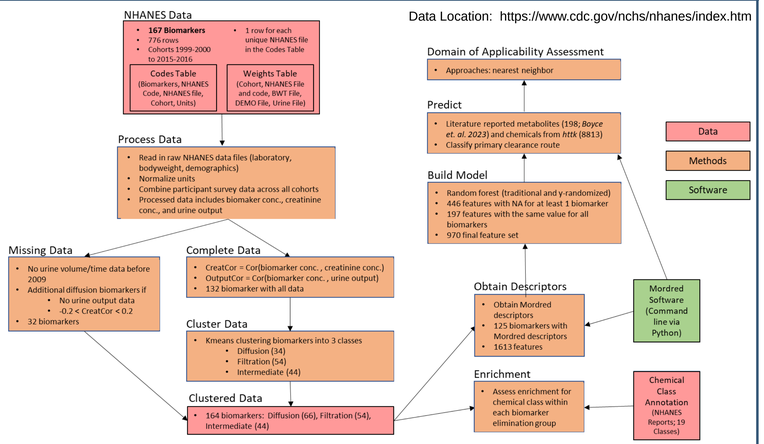

Figure 1. Analysis pipeline. NHANES urinary biomarker data was processed to obtain two correlations: 1) biomarker conc. vs creatinine conc. and 2) biomarker conc. vs urine output. Data from all available cohorts were combined. 164 biomarkers (21 metals and 143 organics) were clustered by these 2 correlations into 3 elimination route groups (glomerular filtration, passive diffusion, and intermediate). Cluster enrichment analysis was performed using chemical classes. A random forest model was built on molecular descriptors (Mordred; *Moriwaki et al. 2018*). The model was then applied to known urinary metabolites from the literature and the *httk* R package.

Results

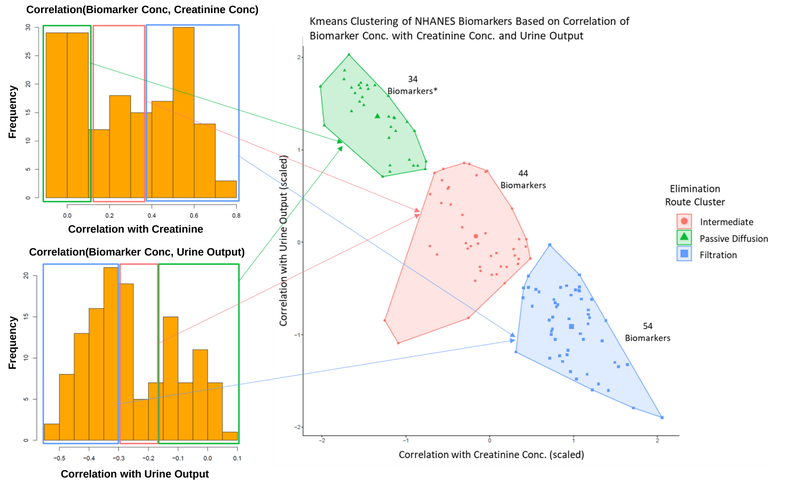

Figure 2. Clustering NHANES biomarkers into primary urine elimination groups. Elimination clusters were calculated using k-means with k = 3. *Chemicals without urine flow data (prior to the 2009-2010 cohort) that exhibited little or no correlation (-0.2 ≤ cor. ≤ 0.2) between urine biomarker conc. and creatinine conc. were subsequently added to the passive diffusion cluster.

Passive Diffusion

Chemical Class	# in Route	# in Class Total	# in Class in Route	P-value
Herbicides	66	7	6	0.0174
Sulfonyl Urea Herbicides	66	17	17	4.70e-08
Fungicides	66	3	3	0.0634

Intermediate

Chemical Class	# in Route	# in Class Total	# in Class in Route	P-value
Personal Care/ Consumer Product	44	13	8	0.0066
Phyto-estrogens	44	6	6	0.0003
Organo-chlorine Pesticides	44	2	2	0.0708

Filtration

Chemical Class	# in Route	# in Class Total	# in Class in Route	P-value
Phthalates	54	16	10	0.0106
VOCs	54	22	13	0.0063
PAHs	54	10	9	0.0002
Metals	54	20	11	0.0258

Figure 3. Cluster enrichment analysis. Chemical class information for the NHANES biomarkers was obtained from https://www.cdc.gov/exposurereport/pdf/Report_Chemical_List-508.pdf. Enrichment was calculated using the hypergeometric test via the phyper() function of the *stats* R package (*R Core Team, 2022*). Chemical classes exhibiting moderate overrepresentation (p-value < 0.1) or higher in each cluster are shown.

Results

Training Model Performance

Model	Predicted Elim. Route	Diffusion	Filtration	Intermediate	Class Error
Traditional (32% OOB Error)	Diffusion	37	4	10	0.2745
	Filtration	4	26	8	0.3158
	Intermediate	6	8	22	0.3889
Y-random (75% OOB Error)	Diffusion	16	15	20	0.6863
	Filtration	11	13	14	0.6579
	Intermediate	17	18	1	0.9722

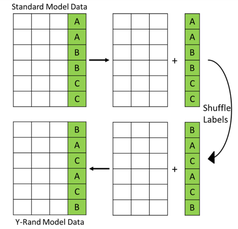

Figure 4. Confusion matrix result of a random forest model using Mordred features on the NHANES training data (labels determined by clustering). Y-randomization, depicted right, represents the null model.

Classification of Primary Urine Elimination Route for Known Urinary Metabolites

Chemical Set	Scenario	Diffusion	Filtration	Intermediate
Literature Reported *(Boyce et. al. 2023)*	# Predicted (% of Total)	99 (50.00%)	23 (11.62%)	76 (38.38%)
	# In Domain (% of Route)	35 (35.35%)	22 (95.65%)	67 (88.16%)
Chemicals in the *httk* R package *(Pearce et al. 2017)*	# Predicted (% of Total)	5769 (65.46%)	1065 (12.08%)	1979 (22.46%)
	# In Domain (% of Route)	2152 (37.30%)	806 (75.68%)	1795 (90.70%)

Table 1. Classification of selected chemical validation sets. Includes 198 literature-derived chemicals and 8813 data-present chemicals from *httk*. A nearest-neighbor applicability domain assessment was used.

Discussion & Future Work

Discussion

- NHANES biomarkers were successfully clustered into 3 distinct, similarly-sized groups based on two chosen correlations and were supported by existing knowledge of urinary kidney elimination pathways.
- A high-throughput machine learning model was built using Mordred molecular descriptors, which performed relatively well on the training data, clearly separating the two primary routes of elimination.
- NHANES provides both uncorrected and creatinine-corrected biomarker conc. Across the large majority of NHANES biomarkers, mean creatinine conc. are > 1 mg/L, and corrected biomarker conc. are divided by this value. Therefore, creatinine correction will likely lead to underestimated daily intake rates for compounds that don't undergo glomerular filtration.
- Note that if a chemical is classified to the intermediate route, it doesn't mean that glomerular filtration isn't happening, rather that filtration isn't the full story. Determining the appropriate correction for this class is a challenge and may be best achieved through a partial correction of biomarker conc.

Future Work

- Use the *httk* R package to calculate plasma conc. in two differing kidney scenarios: 1) GFR is only mechanism of renal elimination, 2) A second route of elimination representing the tubules is also present and is much larger than GFR. All chemicals in *httk* are currently modeled as scenario 1. We will examine whether using the model predictions to place chemicals into the appropriate group will decrease the difference between predicted and observed plasma concentrations.

ORCID: **https://orcid.org/0000-0002-7579-7873**

Innovative Research for a Sustainable Future