Developing a Predictive Model for Chemical Excretion in Urine

Presented by Zachary Stanfield

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

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

The views expressed in this presentation are those of the authors and do not necessarily reflect the views or policies of the U.S. EPA

This must be Slide 2 of your presentation or placed in the lower right portion of your poster.

Explanation:

Conflict of Interest: Presenters must declare funding sources for their work and disclose whether or not there are any competing financial interests in relation to the work described. If no such conflict exists, the statement will simply read that the authors have nothing to disclose. For the purposes of this statement, conflicts of interest are defined as those of a financial nature that, through their potential influence on behavior or content, or form perception of such potential influences, could undermine the objectivity, integrity or perceived value of a publication.

Employment: Recent (while engaged in the research project), present or anticipated employment by any organization that may gain or lose financially through this presentation. This includes positions on an advisory board, board of directors, or other type of management relationship.

Personal Financial Interests: Stocks or shares in companies that may gain or lose financially through publication; consultation fees or other forms of remuneration from organizations that may gain or lose financially; patents or paten applications whose value may be affected by publication.

Funding: Research support (including salaries, equipment, supplies, reimbursement for attending synopsia, and other expenses) by organizations that may gain or lose financially through this publication.

Funding Source(s): All authors are funded by the U.S. Environmental Protection Agency through its Office of Research and Development.

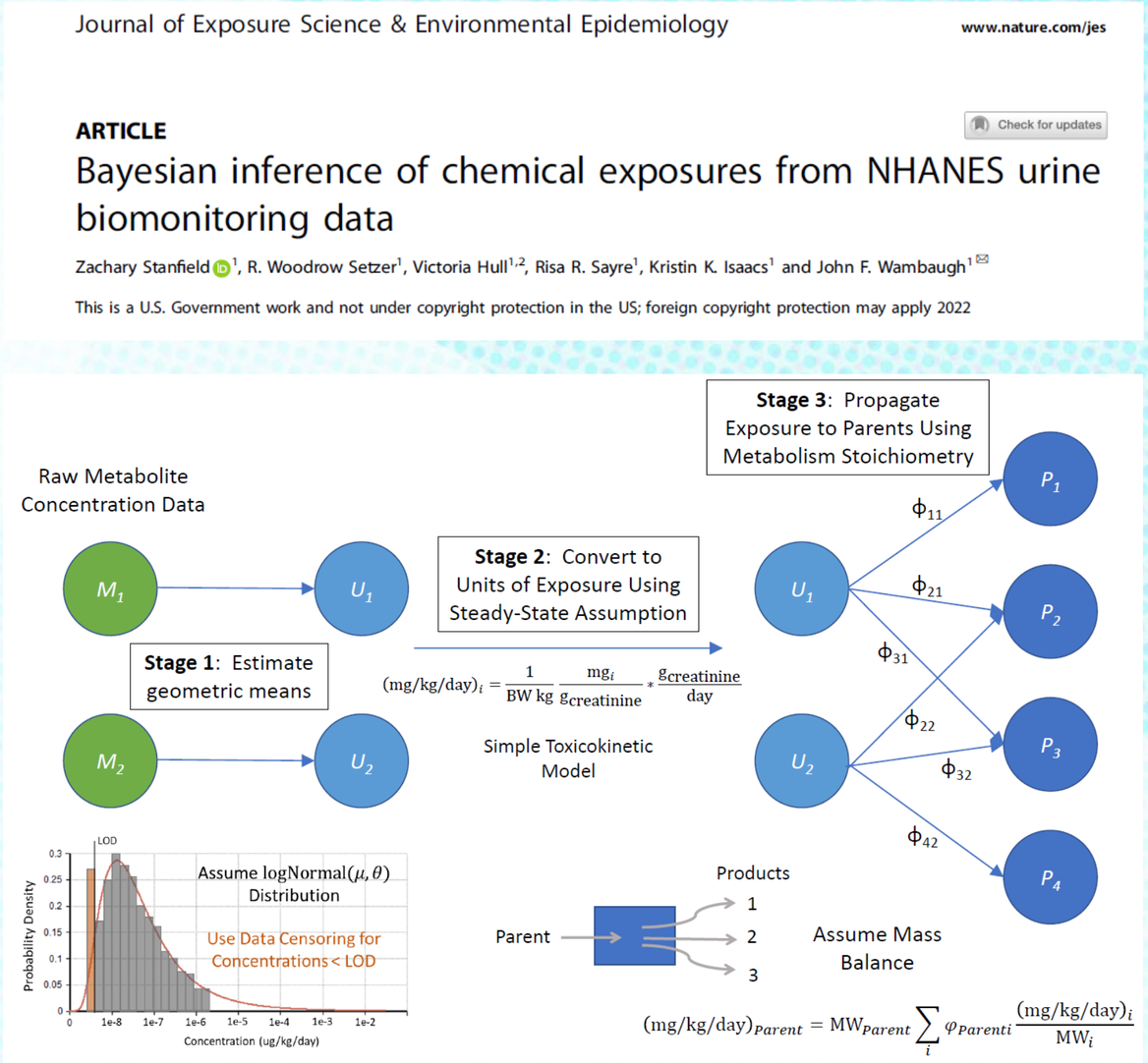
Conflict of Interest: The authors have nothing to disclose.

Employment: All authors are employees of the U.S. Environmental Protection Agency.

Personal Financial Interests: There are no personal financial interests.

Motivation

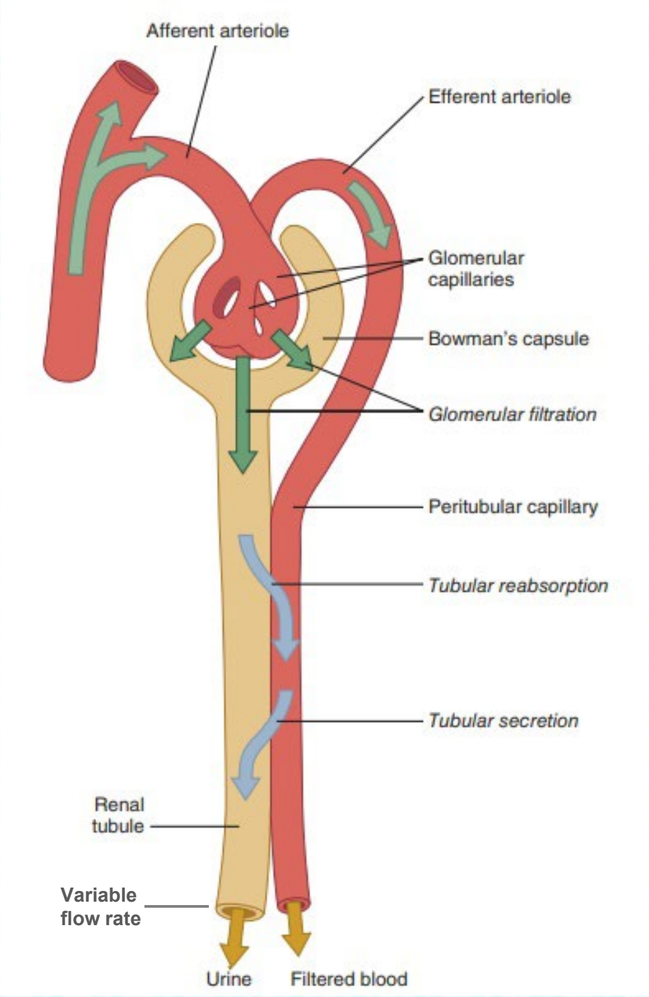
- Recently developed the R package “bayesmarker” (Stanfield et al, 2022; https://github.com/USEPA/CompTox-HumanExposure-bayesmarker)
- Employs Bayesian methodology to infer parent chemical exposures from biomonitoring data (concentrations of urinary metabolites)
- Urinary biomonitoring data plays an important role in exposure science (GAO, 2009; NRC 2006; NRC, 2007; NRC, 2009)
- Some research uses include:
  - Exposure reconstruction
  - PK/PBPK model evaluation
- During exposure calculation, we adjust all metabolite concentrations based on urine creatinine
- There is no general guidance on how to process urinary metabolite concentrations (Middleton et al, 2016; Aylward et al, 2012)
  - Many approaches apply creatinine correction for all chemicals by default (assumes same kidney elimination route)
  - However, the elimination pathway is more complex than this for some substances



#ISES2022LIBSON

Understanding Urinary Excretion

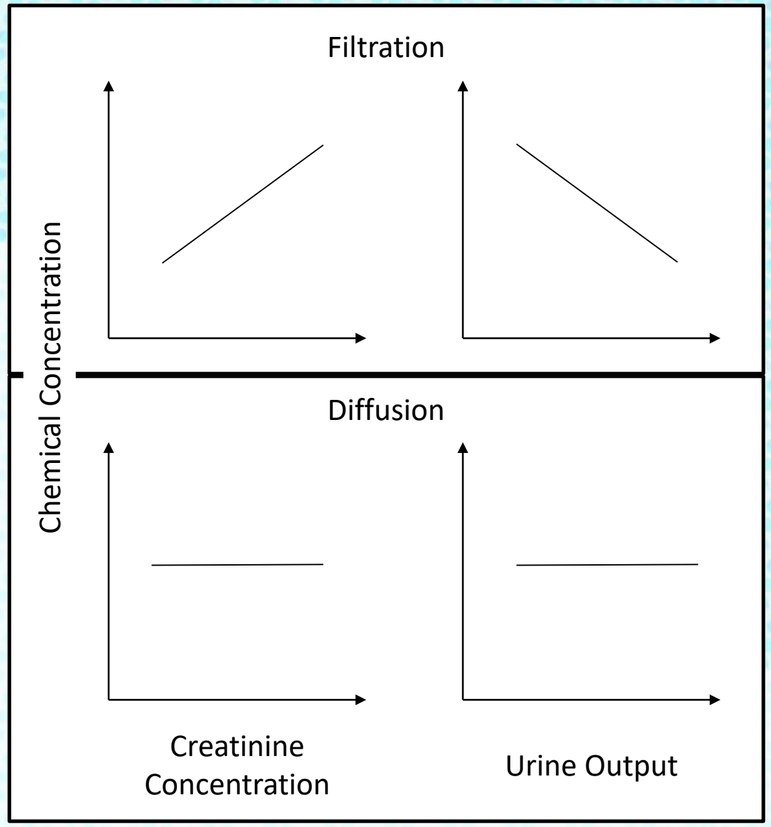
- There are 3 main steps in the process of urine elimination (Boeniger et al, 1993)
  1. Glomerular filtration: substances are filtered in the glomerulus capsule into the renal tubule
  2. Active transport: substances are actively transported between the blood and the filtrate in the renal tubule (sometimes called secretion when leaving the blood)
  3. Passive transport: substances diffuse between the blood and the filtrate in the renal tubule (sometimes called reabsorption when leaving the filtrate)



https://www.pharmacy180.com/article/tubular-secretion-3673/

Understanding Urinary Excretion (Cont.)

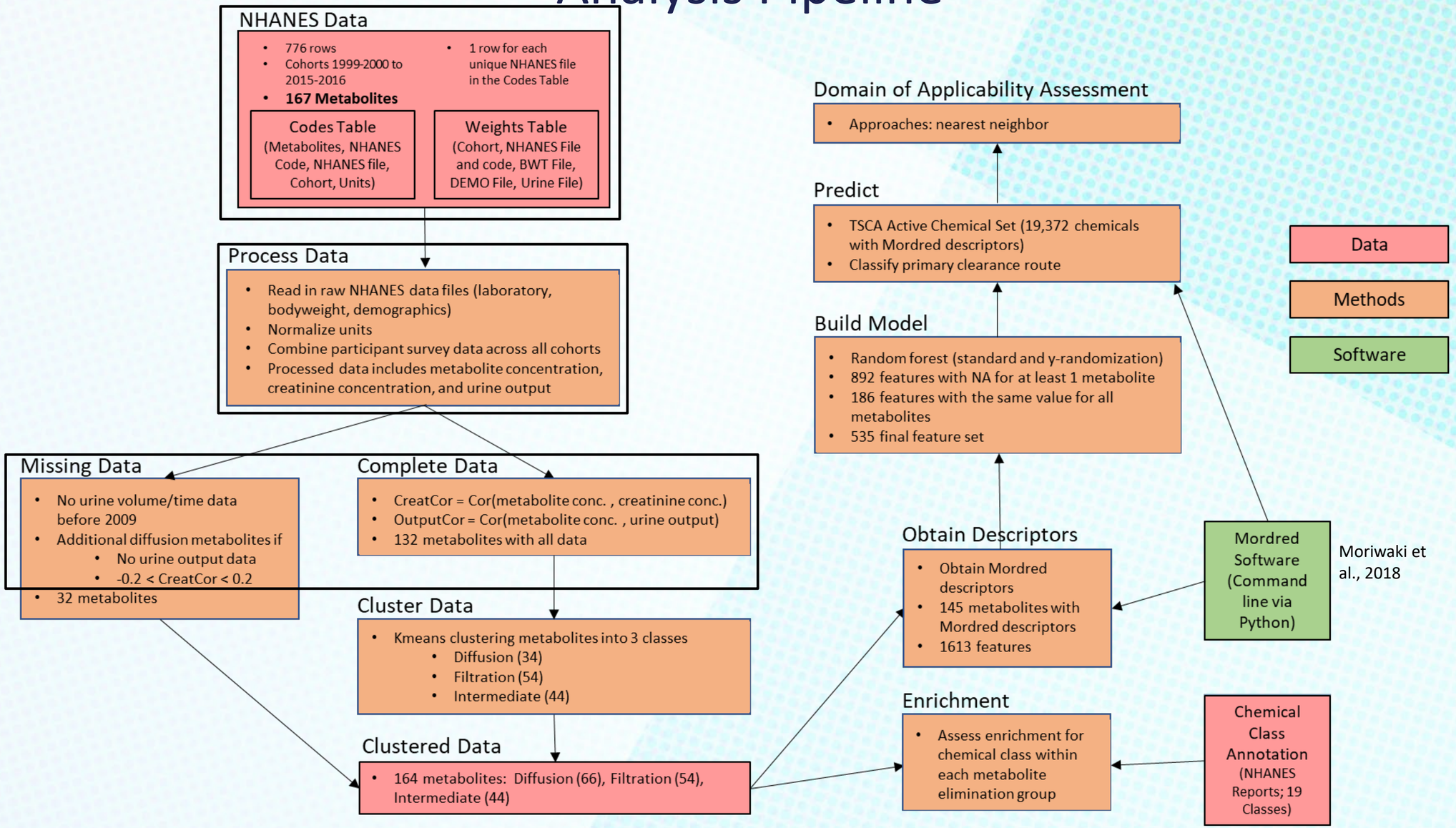
- There are 3 main steps in the process of urine elimination (Boeniger et al, 1993)
  1. Glomerular filtration: substances are filtered in the glomerulus capsule into the renal tubule
  2. Active transport: substances are actively transported between the blood and the filtrate in the renal tubule (sometimes called secretion when leaving the blood)
  3. Passive transport: substances diffuse between the blood and the filtrate in the renal tubule (sometimes called reabsorption when leaving the filtrate)
- Due to our knowledge about renal elimination, we hypothesize that:
  1. Substances undergoing glomerular filtration should follow creatinine concentration (directly) and urine output (indirectly)
  2. Substances undergoing passive diffusion should be independent of creatinine concentration and urine output
- A number of studies (Boeniger et al, 1993; Watanabe et al, 2019) have suggested that some portion of substances belong to an “intermediate” class, which isn’t as well defined as the filtration and diffusion groups



Project Goals

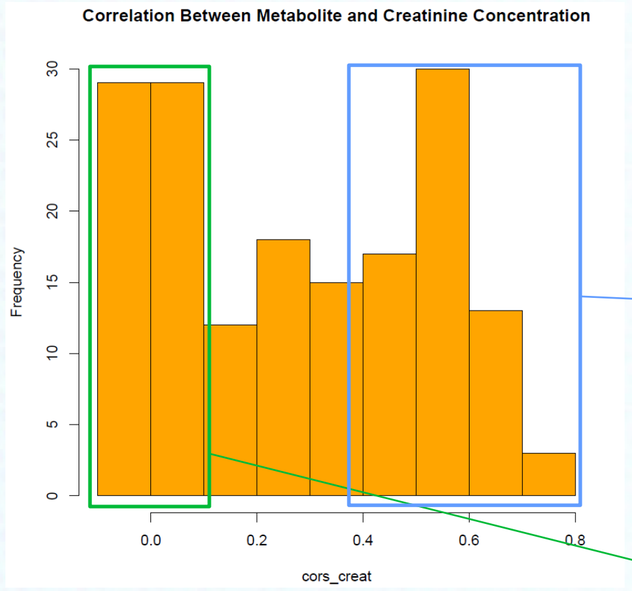
- Test hypotheses of relationships between chemical concentration, creatinine concentration, and urine output for different kidney elimination routes
  - Apply to the CDC's National Health and Nutrition Examination Survey (NHANES)
    - https://www.cdc.gov/nchs/nhanes/index.htm
    - Combine data from the NHANES continuous survey (9 two-year cohorts spanning 1999-2016)
  - Cluster metabolites based on correlation between these 3 measures
  - Characterize these clusters by some criteria
- Build a high-throughput model to predict the primary kidney route of elimination for thousands of chemicals
  - Use clustering of NHANES chemicals as the training data in a classification model
  - Use structural or molecular features that are easily obtained or calculated for most chemicals
- Apply model to a large set of chemicals of interest to provide recommendations on how to handle chemicals for various approaches in exposure reconstruction and/or PBPK model evaluation
  - Assess applicability of model to this large chemical set

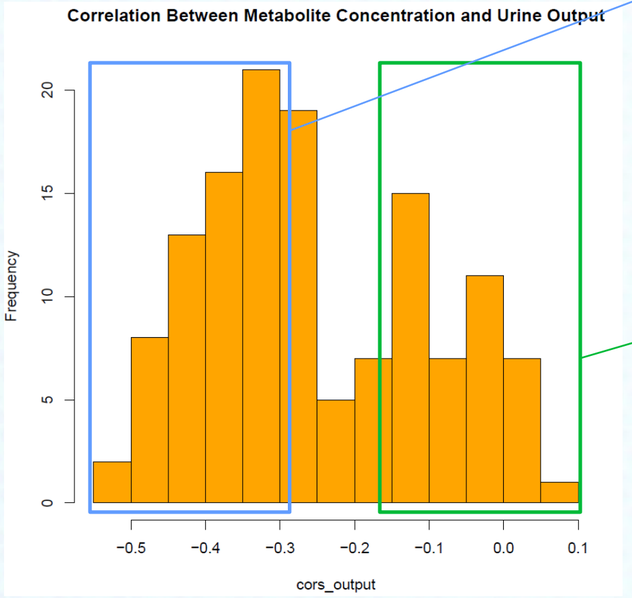
Analysis Pipeline



#ISES2022LIBSON

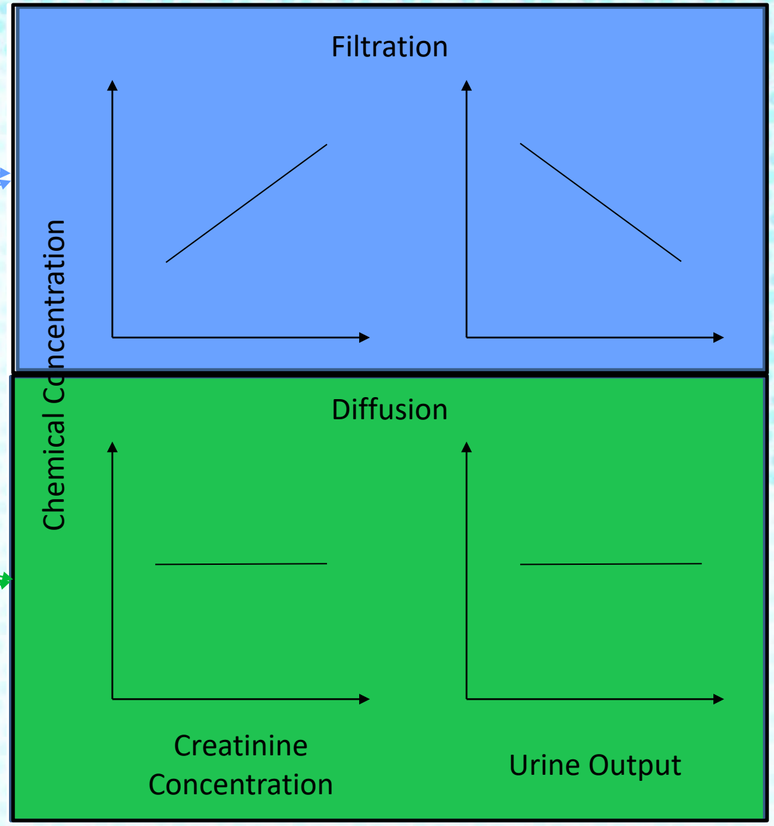
Observing Correlations



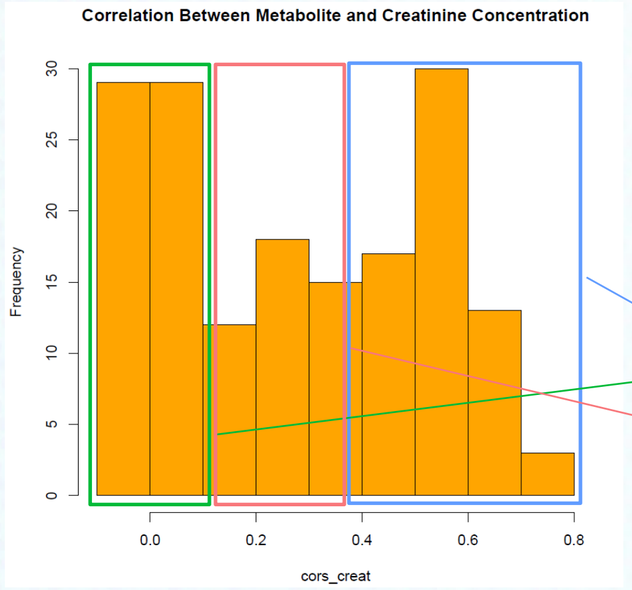


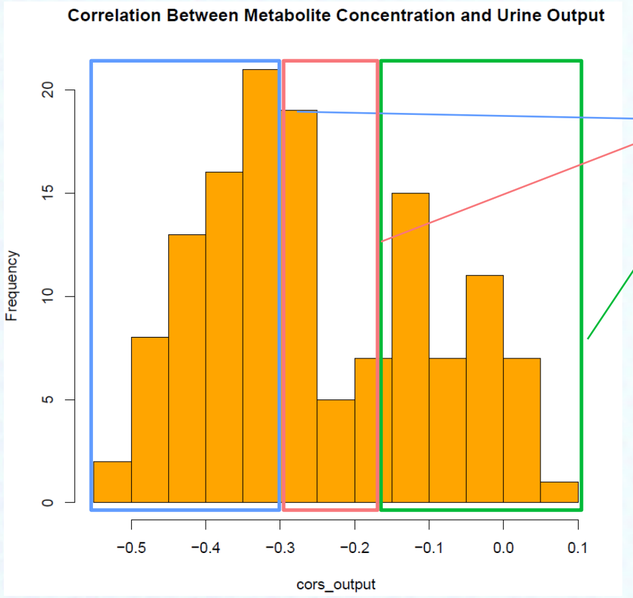
Filtration: follow creatinine concentration (directly) and urine output (indirectly)

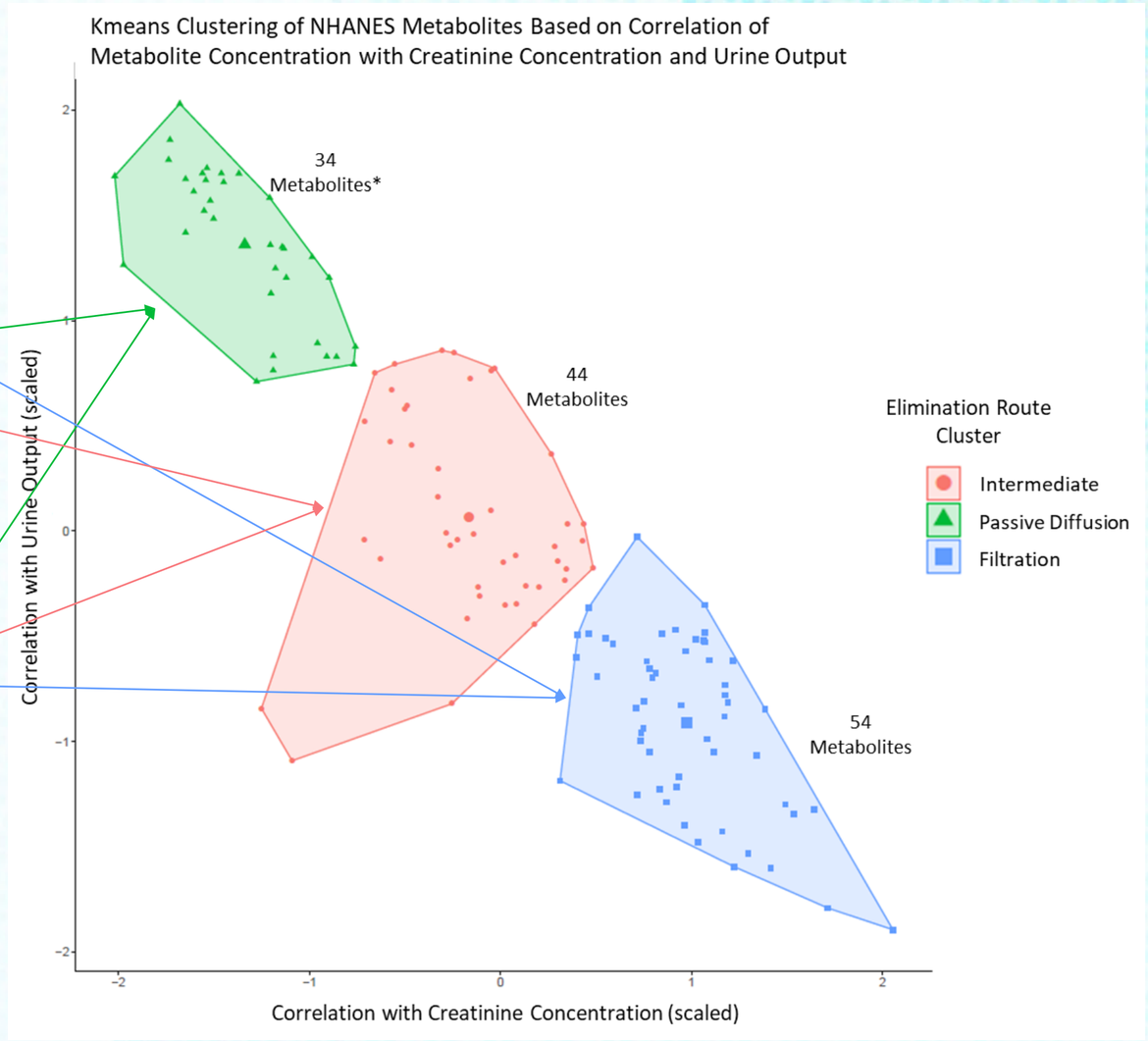
Passive diffusion: independent of creatinine concentration and urine output



Clustering Metabolites by Kidney Elimination Route



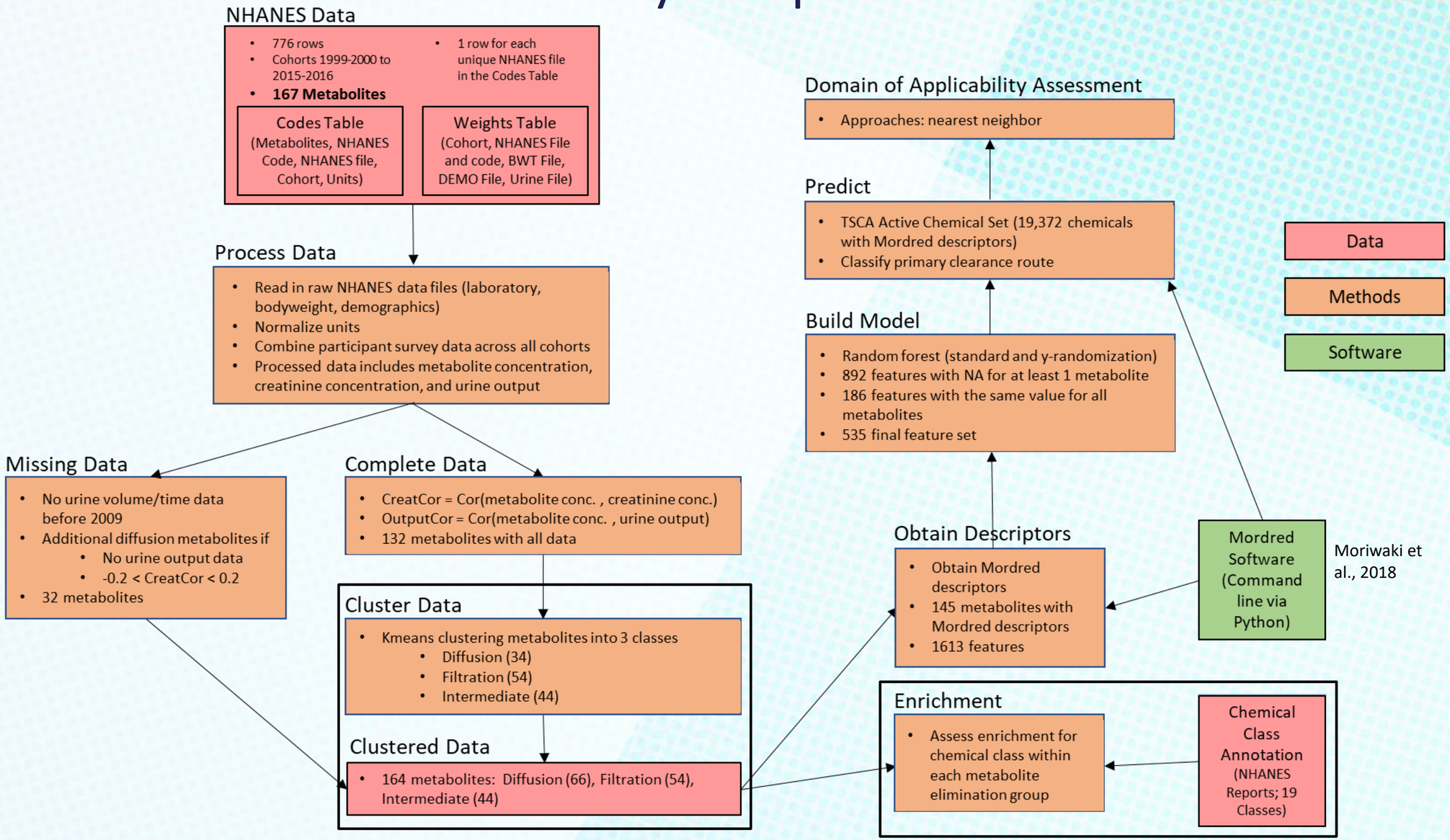




*32 additional metabolites added to the diffusion group. Missing urine output data but had very small correlation with creatinine concentration.

#ISES2022LIBSON

Analysis Pipeline



#ISES2022LIBSON

Chemical Class Enrichment in Kidney Elimination Route Clusters

Enrichment determined by the hypergeometric test using the phyper() function of the stats R package (R Core Team, 2013).

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	4.702e-08
Fungicides	66	3	3	0.0634

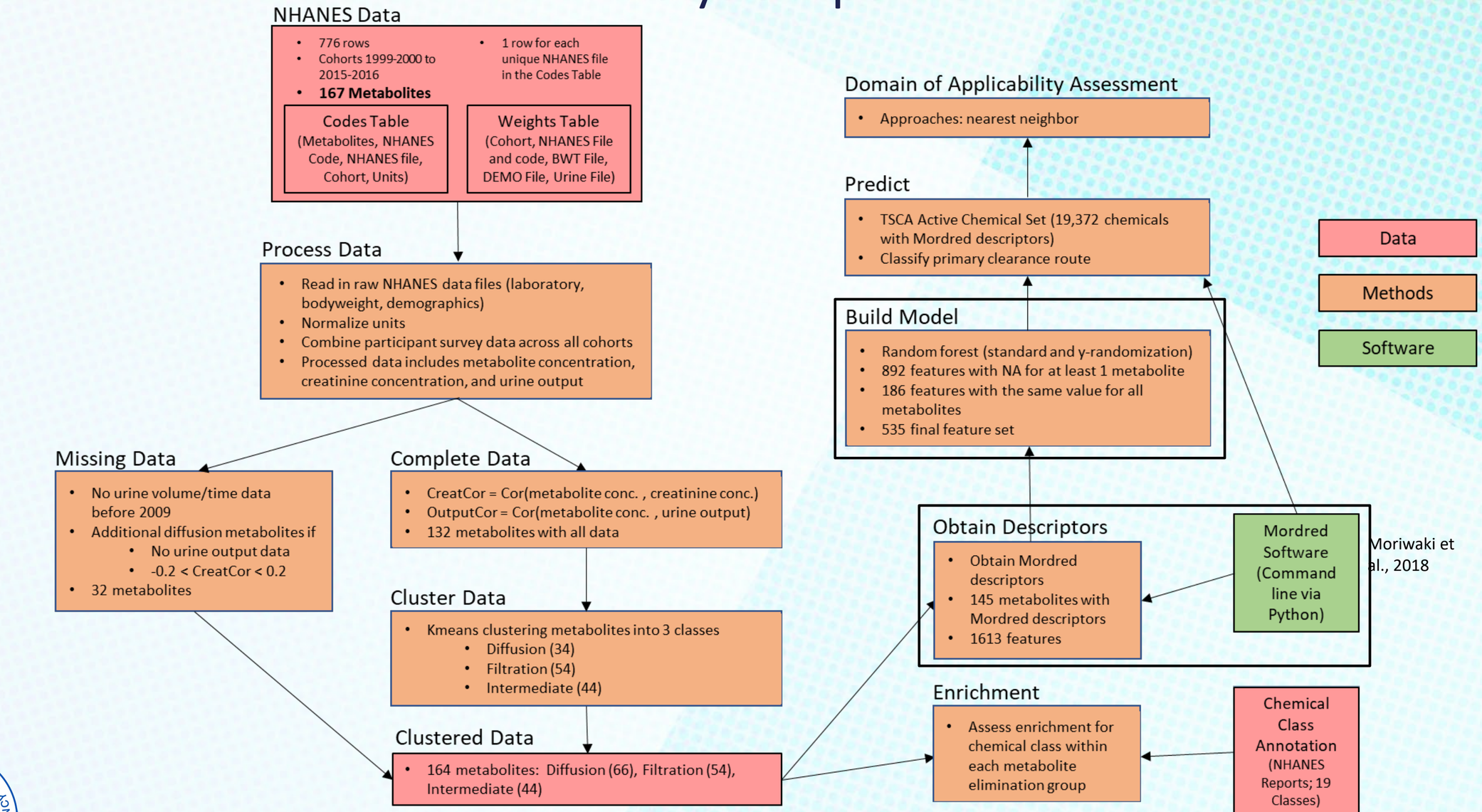
Intermediate

Chemical Class	# in Route	# in Class Total	# in Class in Route	P-value
Personal Care/Consumer Product	44	13	8	0.0066
Phytoestrogens	44	6	6	0.0003
Organochlorine 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
Volatile Organic Compounds	54	22	13	0.0063
Polycyclic Aromatic Hydrocarbons	54	10	9	0.0002
Metals	54	20	11	0.0258

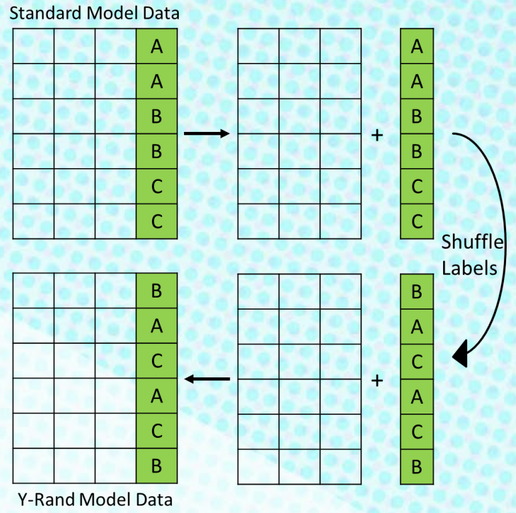
Analysis Pipeline



#ISES2022LIBSON

Build Model

- Trained a Random Forest model using the randomForest R package (RColorBrewer & Liaw, 2018)
- Mordred feature filtering:
  - Started with 1613 features
  - Dropped all features having a NA for any training chemical (892 descriptors)
  - Dropped all features having the same value for all training chemicals (186 descriptors)
  - Resulted in a final model of 535 features
- Balancing classes:
  - Sampled training instances from each route cluster where the sample size was set to the number of chemicals in the smallest cluster



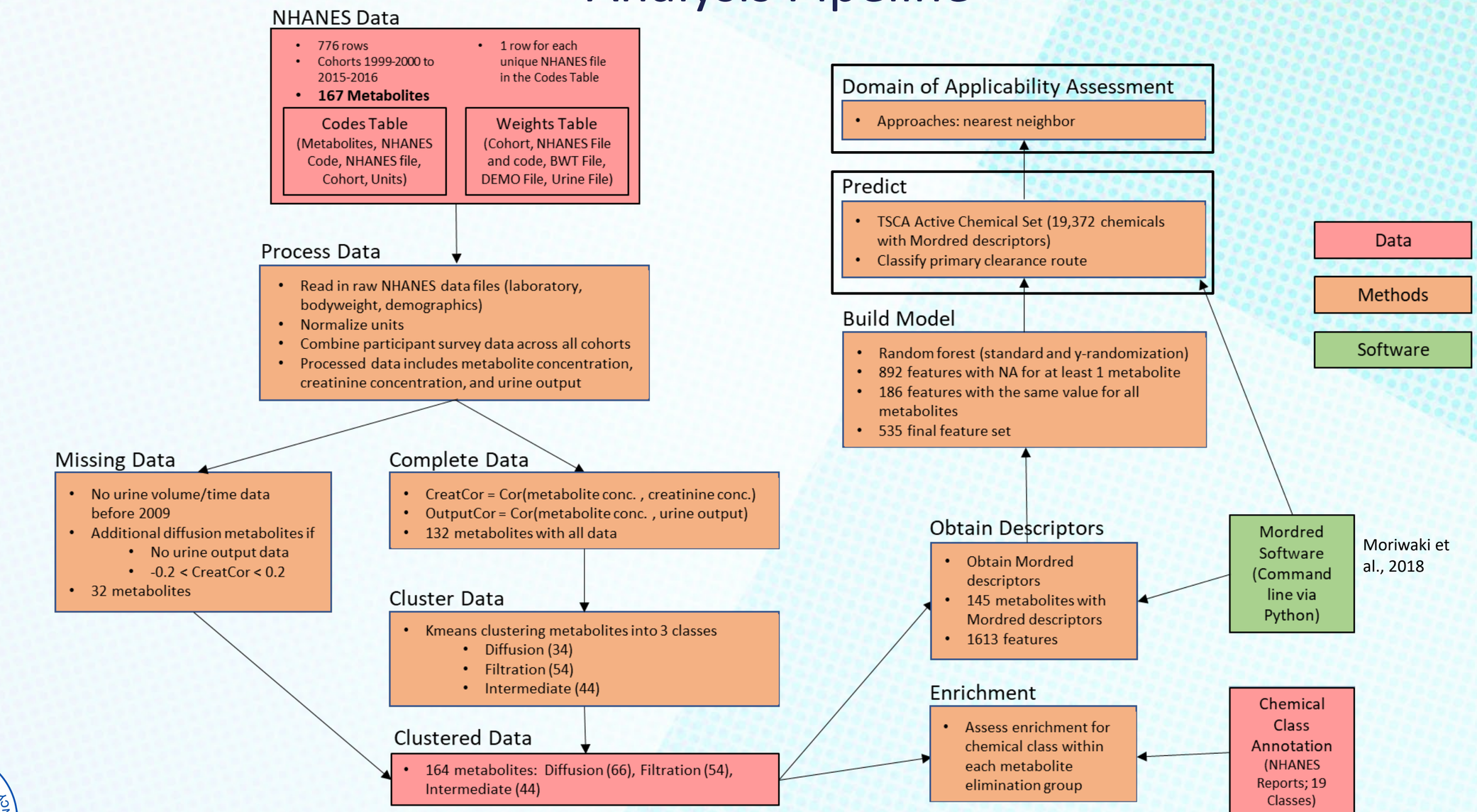
Standard Model Overall OOB Error = 34.48%

	Diffusion	Filtration	Intermediate	Class Error
Diffusion	41	5	11	0.2807
Filtration	2	34	13	0.3061
Intermediate	9	10	20	0.4872

Y-Randomized Model Overall OOB Error = 75.86%

	Diffusion	Filtration	Intermediate	Class Error
Diffusion	20	27	10	0.6491
Filtration	27	10	12	0.7959
Intermediate	22	12	5	0.8718

Analysis Pipeline



High-Throughput Predictions

- List of Toxic Substances Control Act (TSCA) Active Chemicals obtained from the EPA's CompTox Chemicals Dashboard (https://comptox.epa.gov/dashboard)
- TSCA_ACTIVE_NCTI_0221 list with 33599 chemicals
- After obtaining SMILES and calculating Mordred descriptors, 19372 chemicals available for prediction
- Assess domain of applicability in the context of these predictions
  - Use a nearest neighbor approach with similarity threshold as first described in Tropsha & Golbraikh, 2007

Route	# Predicted (% of Total)	# In Domain (% of Route)	# Out Domain (% of Route)
Diffusion	13986 (72.20%)	10422 (74.52%)	3564 (25.48%)
Filtration	1682 (8.68%)	1339 (79.61%)	343 (20.39%)
Intermediate	3364 (19.91%)	2214 (60.43%)	1450 (39.57%)

Future Work

- Additional analyses:
  - Collect more NHANES data (most recent cohorts will have more metabolites all with complete urine metrics)
  - Incorporate additional features into the predictive model
- Model evaluation in the form of two case studies using chemicals clustered/predicted to undergo passive diffusion as primary route of urinary elimination:
  1. PBPK evaluation: run a PBTK model with GFR correction turned off and see if agreement with *in vivo* assay data improves
  2. Exposure reconstruction: estimate exposure intake rates based on urine metabolite concentrations; compare exposure ranges with GFR correction turned on and off

Future Work: Case Study 1

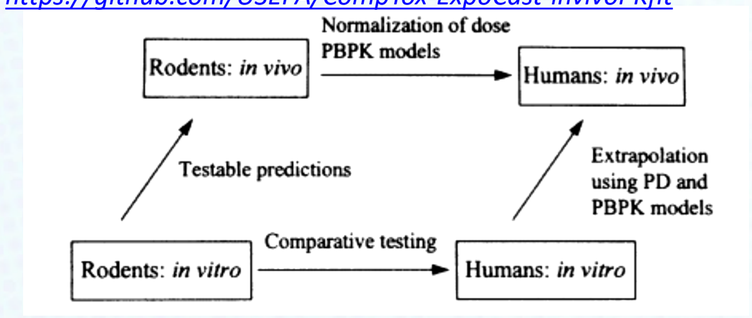
Journal of Statistical Software
Information on Mission | Information for Authors | Style Guide | Volumes | About
Home / Archives / Vol. 79 (2017) / Issue 4

httk: R Package for High-Throughput Toxicokinetics

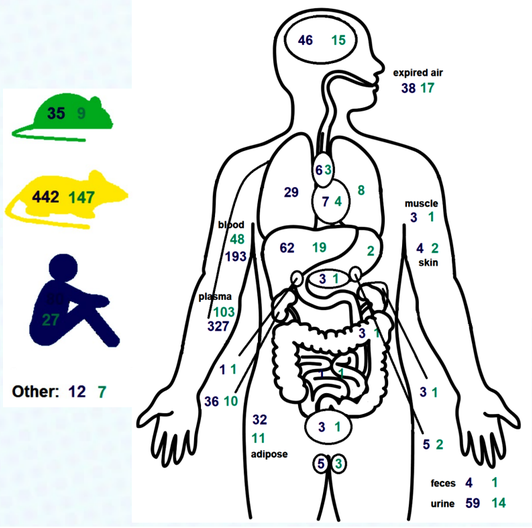
Robert G. Pearce, R. Woodrow Setzer, Cory L. Strope, Nisha S. Sipes, John F. Wambaugh

https://CRAN.R-project.org/package=httk

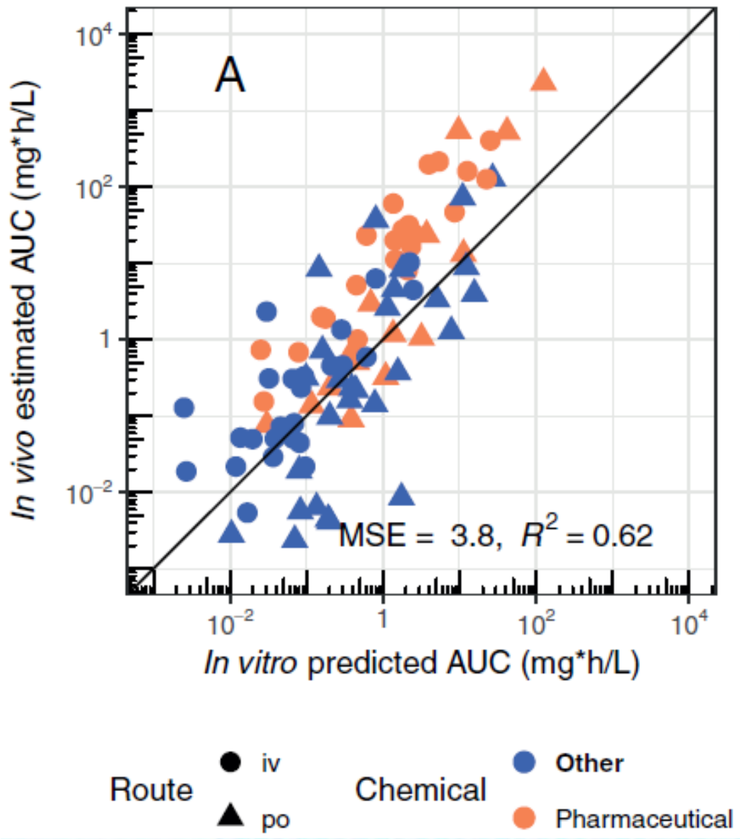
In Vitro-In Vivo Extrapolation (IVIVE)
https://github.com/USEPA/CompTox-ExpoCast-invivoPKfit



Concentration vs Time Database (CvTdb)
http://github.com/USEPA/CompTox-PK-CvTdb

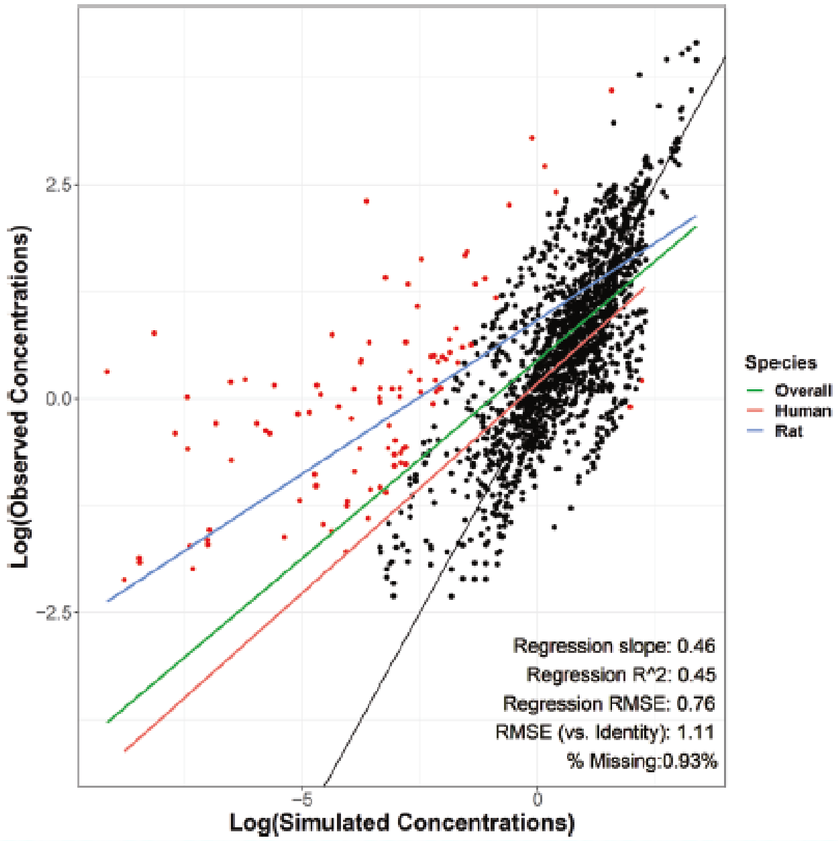


Comparison to *in vivo* rat data for 45 chemicals using httk and IVIVE methods



Wambaugh et al, 2018

Comparison to *in vivo* human and rat data for 41 volatile organic compounds under 142 exposure scenarios using a generic inhalation model (httk) and CvTdb



Linakis et al, 2018

#ISES2022LIBSON

Future Work: Case Study 2

- Revisit our original motivation, the bayesmarker package for calculating parent chemical intake rates
- Analysis plan:
  - Rerun all metabolites clustered into the passive diffusion primary kidney elimination group through bayesmarker without creatinine correction
  - Compare exposure estimates for parent chemicals with and without correction

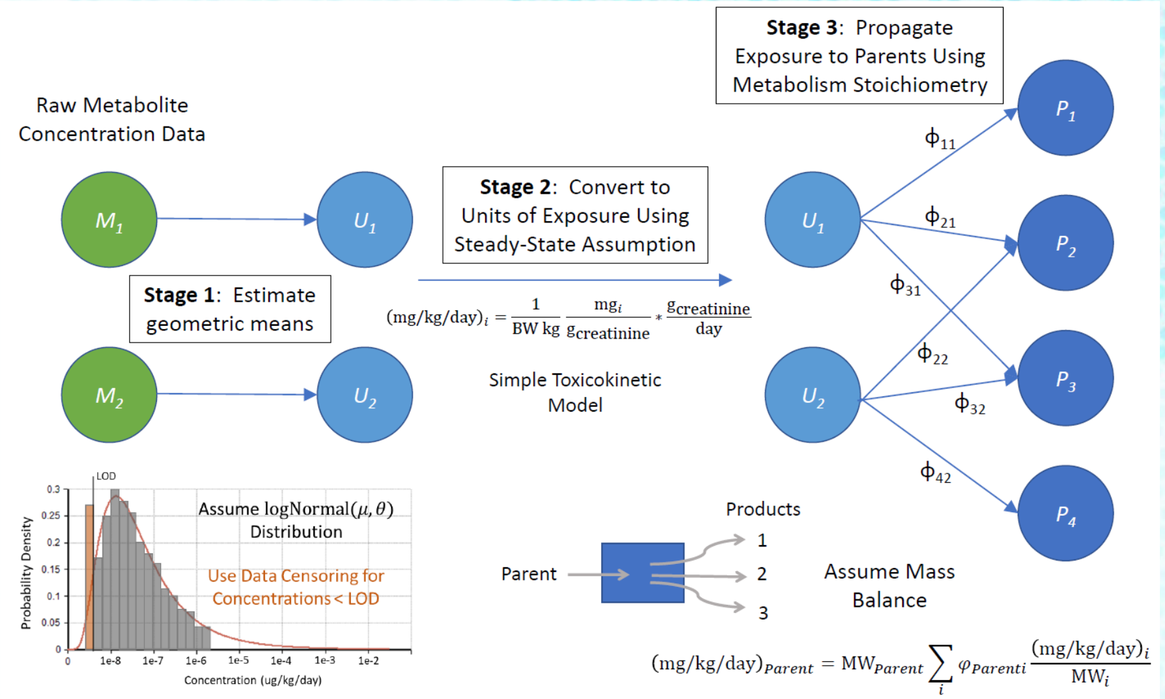
Journal of Exposure Science & Environmental Epidemiology www.nature.com/jes

ARTICLE

Check for updates

Bayesian inference of chemical exposures from NHANES urine biomonitoring data

Zachary Stanfield[1], R. Woodrow Setzer[1], Victoria Hull[1,2], Risa R. Sayre[1], Kristin K. Isaacs[1] and John F. Wambaugh[1]



#ISES2022LIBSON

Thank You!

References

- Aylward, L. L., Kirman, C. R., Adgate, J. L., McKenzie, L. M., & Hays, S. M. (2012). Interpreting variability in population biomonitoring data: role of elimination kinetics. *Journal of exposure science & environmental epidemiology*, *22*(4), 398-408.
- Boeniger, M. F., Lowry, L. K., & Rosenberg, J. (1993). Interpretation of urine results used to assess chemical exposure with emphasis on creatinine adjustments: a review. *American Industrial Hygiene Association journal*, *54*(10), 615-627.
- GAO (U.S. Government Accountability Office). Biomonitoring: EPA Needs to Coordinate its Research Strategy and Clarify its Authority to Obtain Biomonitoring Data. Washington, DC:GAO. 2009.
- Middleton, D. R., Watts, M. J., Lark, R. M., Milne, C. J., & Polya, D. A. (2016). Assessing urinary flow rate, creatinine, osmolality and other hydration adjustment methods for urinary biomonitoring using NHANES arsenic, iodine, lead and cadmium data. *Environmental Health*, *15*(1), 1-13.
- NRC (National Research Council). Human Biomonitoring for Environmental Chemicals. Washington, DC:National Academies Press. 2006.
- NRC (National Research Council). Toxicity Testing in the 21St Century: A Vision and a Strategy. Washington, DC:National Academies Press. 2007.
- NRC (National Research Council). Science and Decisions: Advancing Risk Assessment. Washington, DC:National Academies Press. 2009.
- Linakis, M. W., Sayre, R. R., Pearce, R. G., Sfeir, M. A., Sipes, N. S., Pangburn, H. A., ... & Wambaugh, J. F. (2020). Development and evaluation of a high throughput inhalation model for organic chemicals. *Journal of exposure science & environmental epidemiology*, *30*(5), 866-877.
- Moriwaki, H., Tian, Y. S., Kawashita, N., & Takagi, T. (2018). Mordred: a molecular descriptor calculator. *Journal of cheminformatics*, *10*(1), 1-14.
- 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.
- R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.
- RColorBrewer, S., & Liaw, M. A. (2018). Package 'randomforest'. *University of California, Berkeley: Berkeley, CA, USA*.
- Sayre, R. R., Wambaugh, J. F., & Grulke, C. M. (2020). Database of pharmacokinetic time-series data and parameters for 144 environmental chemicals. *Scientific data*, *7*(1), 1-10.
- Stanfield, Z., Setzer, R. W., Hull, V., Sayre, R. R., Isaacs, K. K., & Wambaugh, J. F. (2022). Bayesian inference of chemical exposures from NHANES urine biomonitoring data. *Journal of Exposure Science & Environmental Epidemiology*, 1-14.
- Tropsha, A., & Golbraikh, A. (2007). Predictive QSAR modeling workflow, model applicability domains, and virtual screening. *Current pharmaceutical design*, *13*(34), 3494-3504.
- Wambaugh, J. F., Hughes, M. F., Ring, C. L., MacMillan, D. K., Ford, J., Fennell, T. R., ... & Thomas, R. S. (2018). Evaluating in vitro-in vivo extrapolation of toxicokinetics. *Toxicological Sciences*, *163*(1), 152-169.
- Watanabe, R., Ohashi, R., Esaki, T., Kawashima, H., Natsume-Kitatani, Y., Nagao, C., & Mizuguchi, K. (2019). Development of an in silico prediction system of human renal excretion and clearance from chemical structure information incorporating fraction unbound in plasma as a descriptor. *Scientific reports*, *9*(1), 1-11.