Winning the Race Against Competing Risks: Optimizing Drinking Water Disinfection to Minimize Opportunistic Pathogen & DBP Risk

Multiple Project Directors:

Jade Mitchell, PhD, Michigan State University

Mark H. Weir, PhD, The Ohio State University

Co-PIs: Tiong Aw (Tulane); Zaid Chowdhury (Garver); Erin Dreelin (MSU); Charles N. Haas (Drexel); Tanju Karanfil (Clemson); Ursula Lauper (NYSDOH); Alexis Mraz (TCNJ); Patrick Palmer (NYSDOH); Joan Rose (MSU); Pauline Wanjugi (NYSDOH)

Post-Doctoral Research Associates/PhD students: Peng Dai (Clemson), Habib Uzun (Clemson), Sawaira Adil (Clemson), Melodie Nasr (Tulane), Fernanda Mac-Allister Cedraz (Tulane), Carly Gomez (MSU), and Ryan Julien (MSU)

Project email: BAE.drinkingwater@msu.edu

EPA Grant R84060401-0

Project Goal

Our project goal is to better understand and predict occurrence of disinfection byproducts, opportunistic pathogens, and the associated health risk tradeoffs posed by them in DWDS across the continental U.S.

Objectives

Objective 1: Develop a strategic sampling program based on health data and system characteristics.

Objective 2: Elucidate the locations and conditions leading to the occurrence and co-occurrence of DBPs and OPs in actual DWDS through field sampling, modeling, and analysis.

Objective 3: Characterize risks associated with DBP and OP occurrence and concentrations, as well as risk tradeoffs between them for sub-populations.

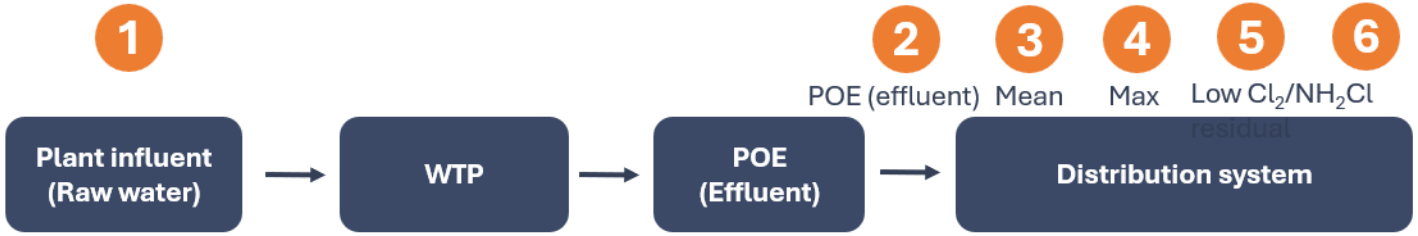
Utility Partners: Selection Criteria

Criteria:

- SDWIS violations (TCR and DBP rule)
- Regional diversity
- Sampling locations where DBPs/OPs are likely to occur

Contacted:

- ~ 40 utilities from SDWIS
- ~ 50 utilities approached by WRF study on opportunistic pathogens
- ~ 15 utilities that previously participated in another WRF project on unregulated DBPs
- ~ 50 utilities that were identified by other project participants



Utility Partners

25 utility partners across 9 EPA regions

Disinfection	EPA Regions								
	1	2	3	4	5	6	7	8	9
Cl2 Utility (15)	A	B	H		K	P			X
		C			M				Y
		D			N				
		E			O				
		F							
		G							
NH2Cl Utility (10)				I	L	Q	V	W	
				J		R			
						S			
						T			
						U			

Ongoing Work: Drinking Water Distribution System Sampling

Water Quality and DBPs (Clemson University)

Parameter	Unit	Measurement Method	Sample Bottle
NDMA	µg/L	EPA Method 521	500 mL (with ascorbic acid)
THMs	µg/L	EPA method 551	500 mL +125 mL (with ascorbic acid and sulfuric acid)
I-THMs	µg/L	EPA method 551	
HANs	µg/L	EPA method 551	
I-HAN	µg/L	EPA method 551	
HALs	µg/L	EPA method 551	
HAAs	µg/L	EPA method 552.2	
I-HAAs	µg/L	N/A	
TOCl, TOBr, TOI	µg/L	N/A	
pH	-	SM* 4500-H+	125 mL
Dissolved organic carbon (DOC)	mg/L	SM* 5310B	
UV	abs	SM* 5910	
Chlorine/chloramine	mg/L	SM* 4500-Cl F	
Dissolved Nitrogen (DN)	mg/L	SM 5310 High-Temperature Combustion	
I^-	µg/L	N/A	
Cl^-, Br^-, SO_4^{2-}, NO_3^-, NO_2^-	µg/L	EPA method 300	

1 cooler/plant (6 location samples)

Bottle 125mL	**× 5**
Bottle 125mL (quenching reagent)	**× 5**
Bottle 500mL (quenching reagent)	**× 12**

Bacteriological (Tulane University)

Parameter	Unit	Measurement Method	Sample Bottle
Legionella species	CFU	ISO 11731, Standard Methods for the Examination of Water and Wastewater	1 L with sodium thiosulfate
Pseudomonas aeruginosa	CFU	Standard Methods for the Examination of Water and Wastewater	1 L with sodium thiosulfate
Legionella pneumophila	MPN	IDEXX Legiolert	100 ml with sodium thiosulfate
Pseudomonas aeruginosa	MPN	IDEXX Pseudalert	100 ml with sodium thiosulfate

1 cooler can hold 6 location samples (1 plant)

Bottle 500ml (for raw water) × 1

Bottle 1L x 6

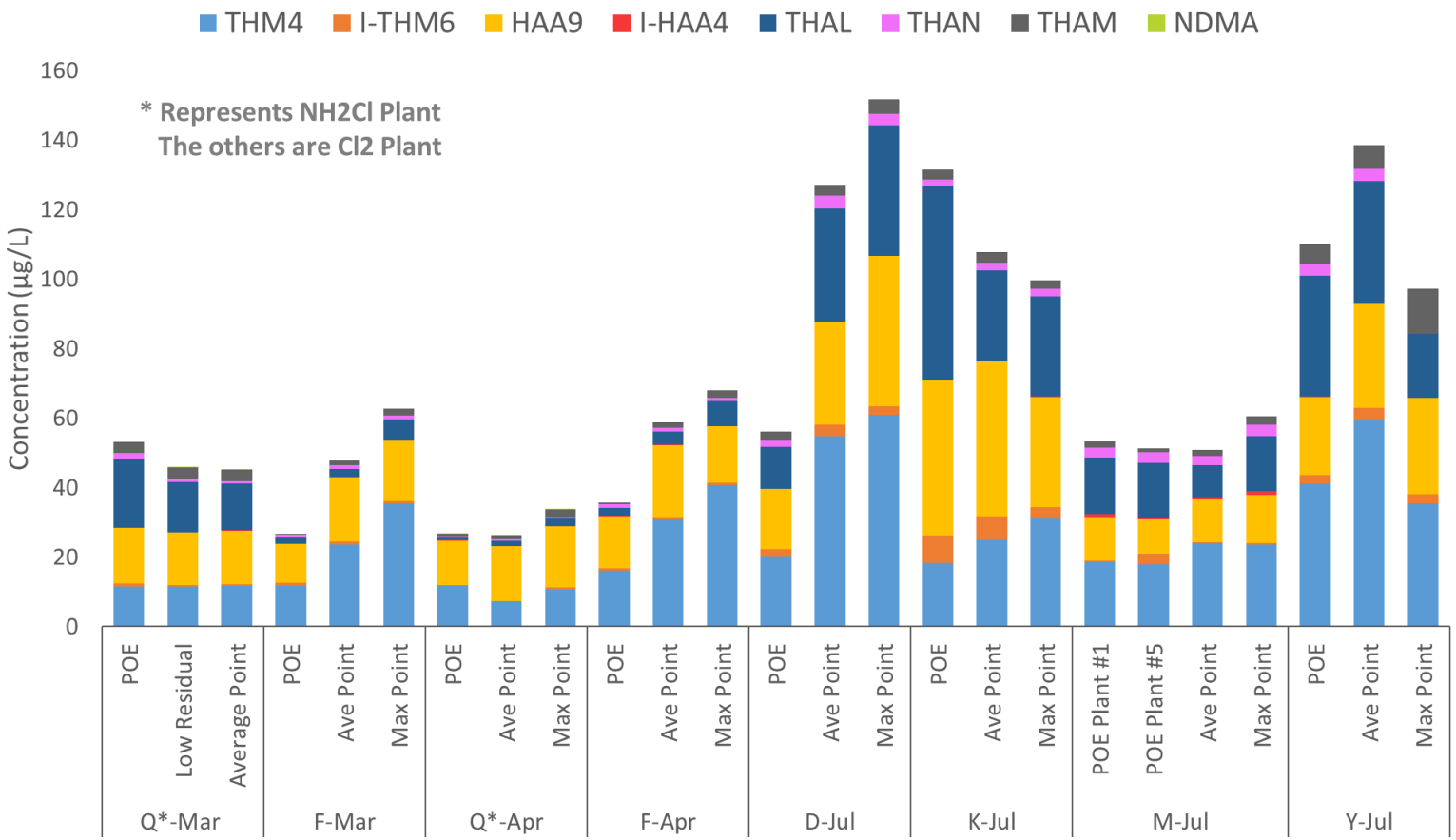
Bottle 100ml (for IDEXX) x 8

Progress: OPs

- Number of samples collected: 75 (11 source water, 12 finished water, 52 water samples within distribution systems) – as of July 31, 2024
- *Legionella pneumophila* was detected in
  - 1 finished water (concentration 2.2 MPN/100ml)
  - 3 samples within distribution systems (concentration ranged from 2.3 to 85.4 MPN/100ml)
- *Pseudomonas aeruginosa* was detected in
  - 7 source water samples (concentration ranged from 3.1 to >200.5 MPN/100ml)
  - 1 sample within distribution systems (concentration 3.1 MPN/100ml)

Progress: DBPs

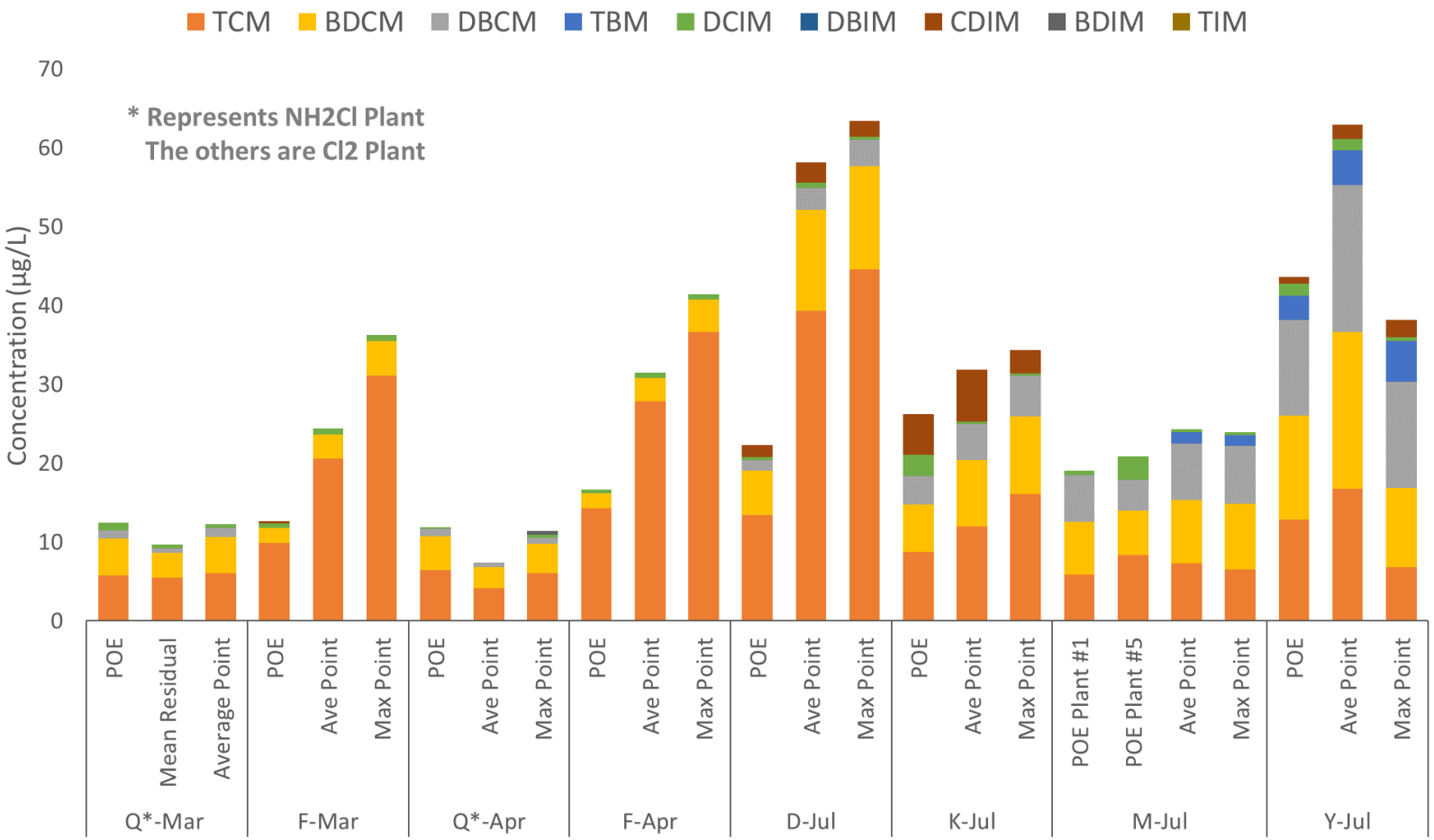
DBP concentration and water age



- DBP formation increased with water age in 5/8 sampling events
- THM4 and HAA5
  - Cl2 plants: THM and HAA increased
  - NH2Cl plants: No significant change
- HAL, HAM
  - Slightly increased
- HAN
  - No significant change
- NDMA
  - NH2Cls plant only
- Overall DBP formation
  - Cl2 plants > NH2Cl plants

Progress: Occurrence of THMs

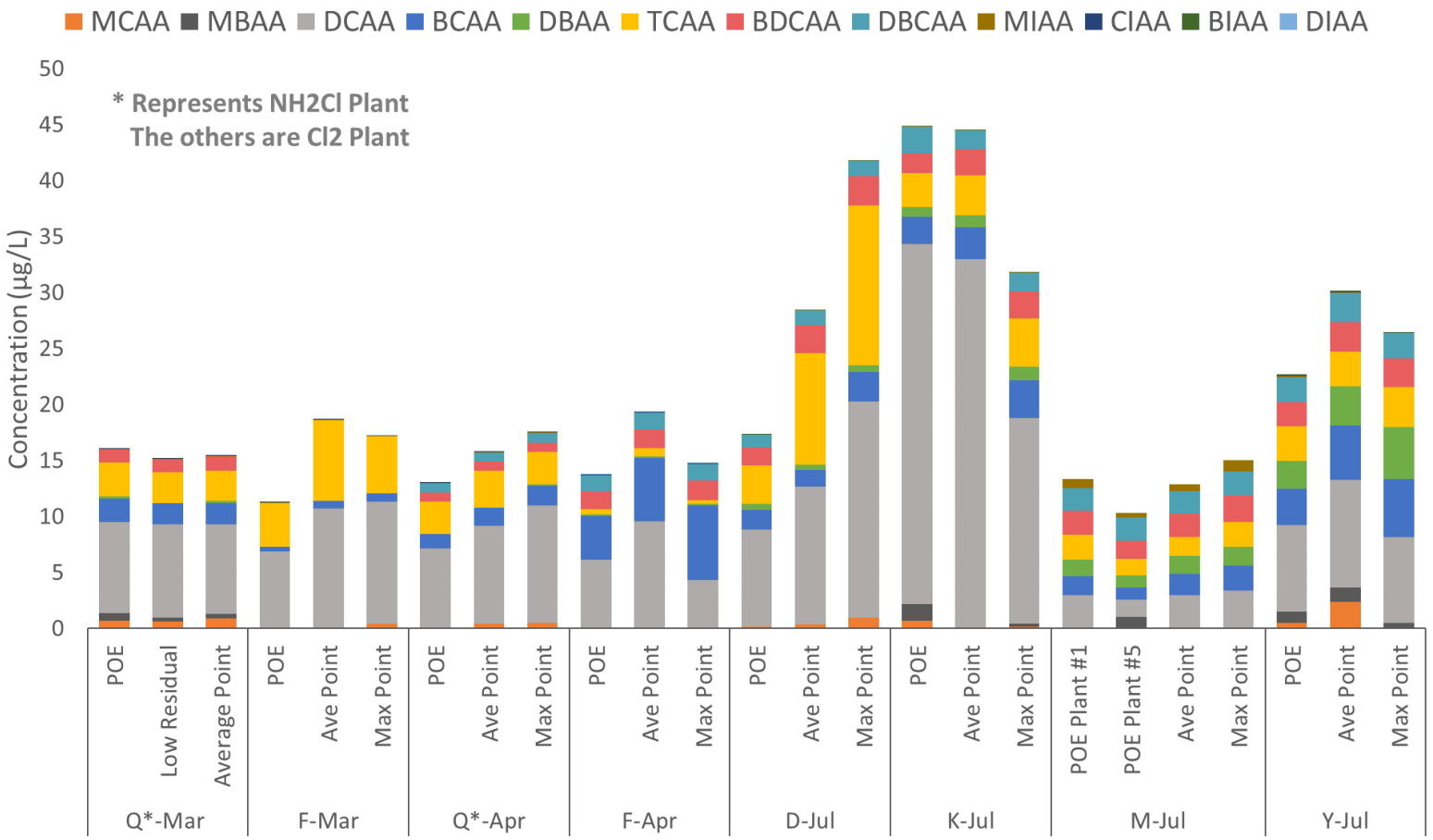
THM occurrence and water age



- Overall THM formation increased with water age
- Dominant species: TCM, BDCM, DBCM

Progress: Occurrence of HAAs

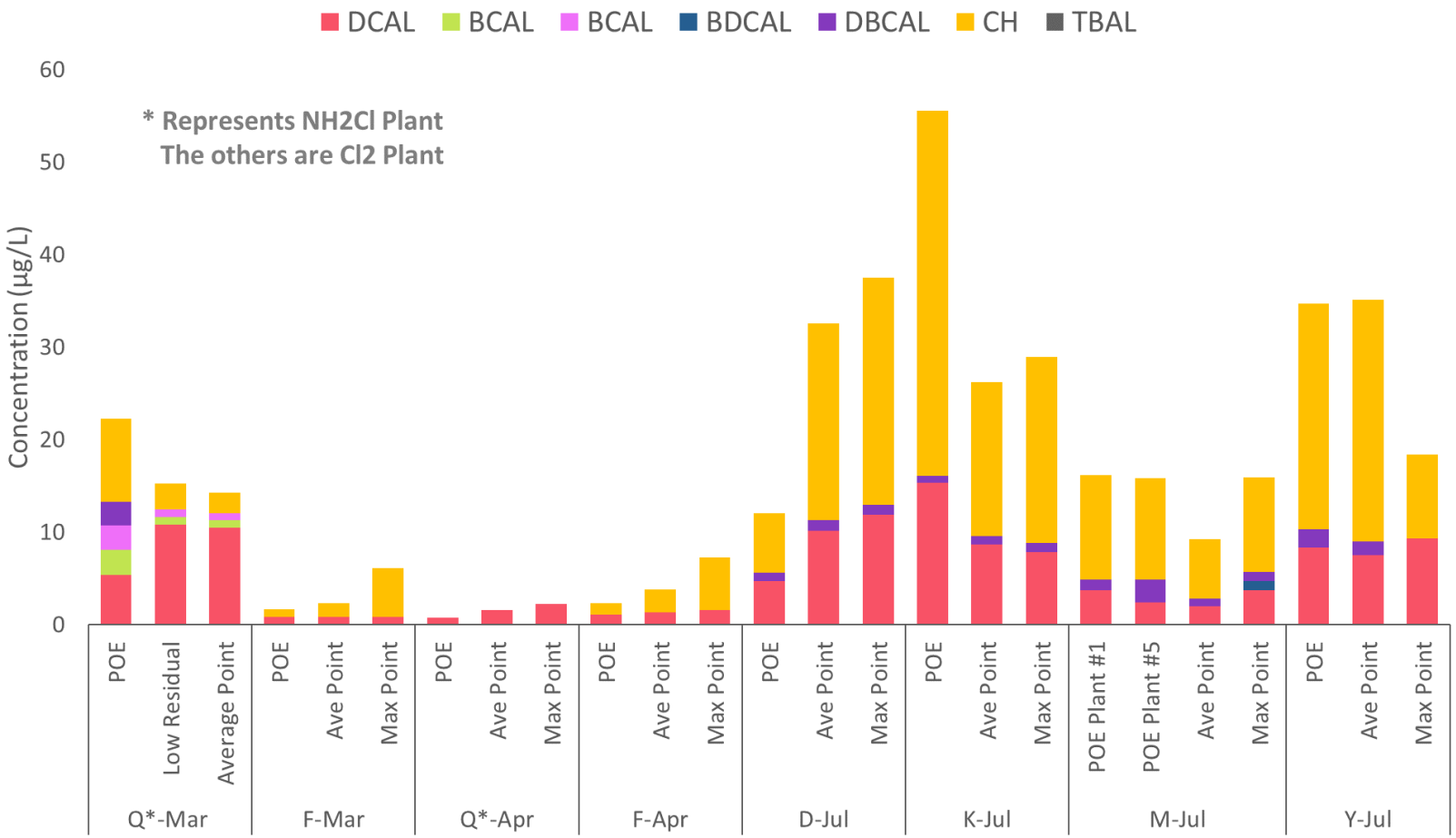
HAA occurrence and water age



- Overall HAA formation was stable or slightly increased with water age
- Dominant species: DCAA, TCAA, BCAA

Progress: Occurrence of HALs

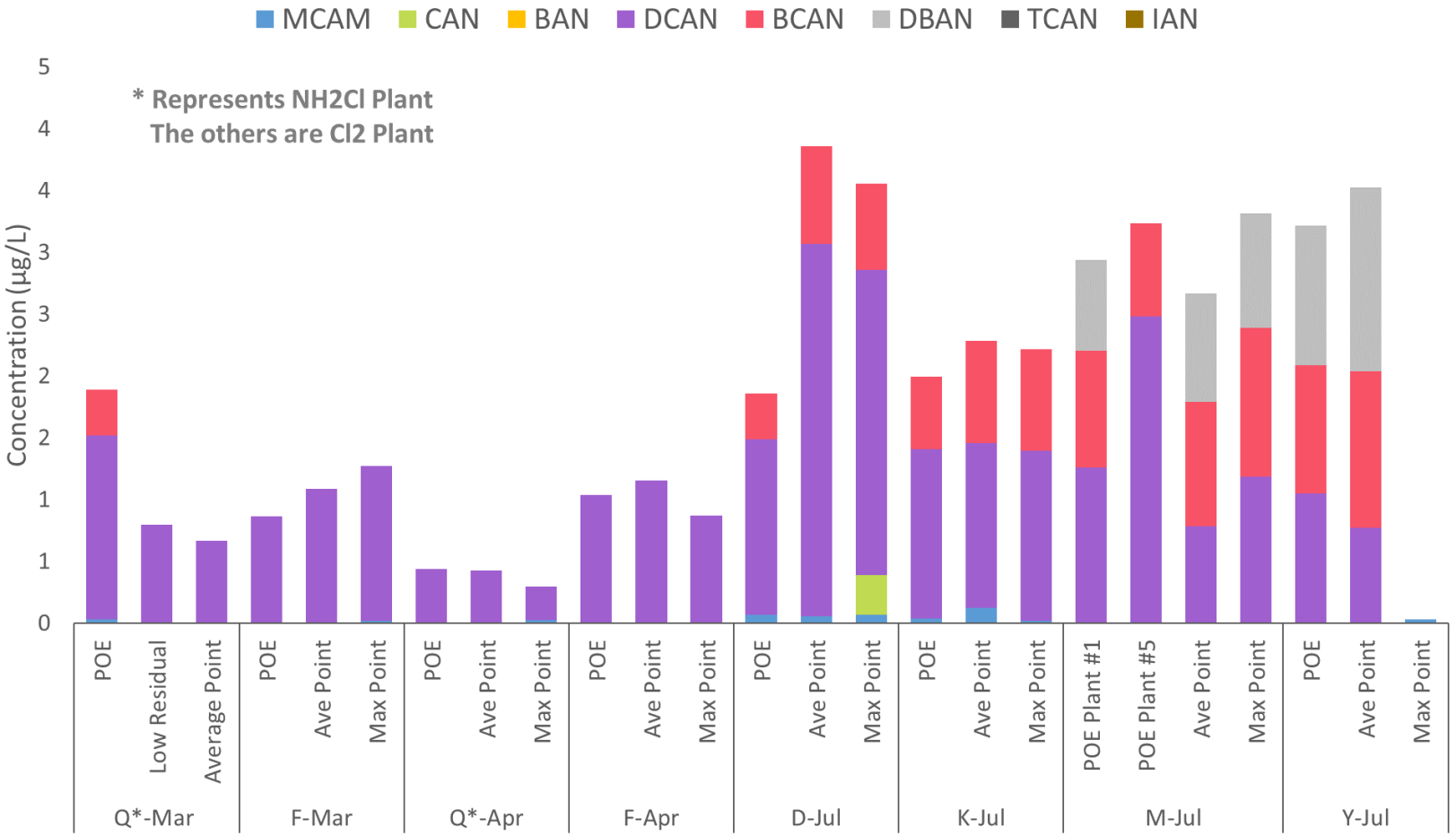
HAL occurrence and water age



- Overall HAL occurrence varied with water age
- Dominant species: DCAL, CH (TCAL)

Progress: Occurrence of HANs

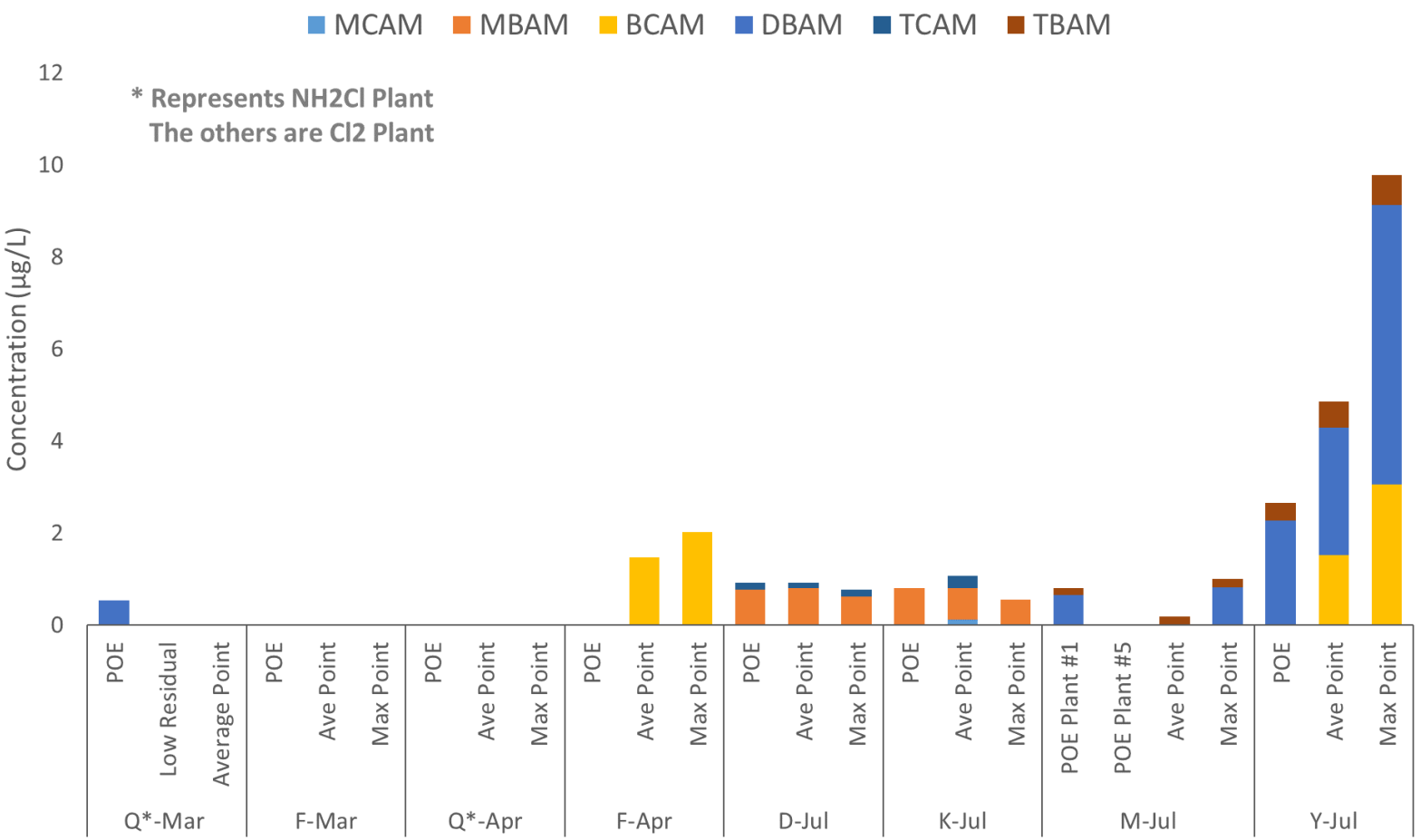
HAN occurrence and water age



- Overall HAN occurrence varied with water age
- Dominated species: DCAN, BCAN, (DBAN)

Progress: Occurrence of HAMs

HAM occurrence and water age



- The occurrence trend of HAM increased with water age.
- Dominated species: BAM, BCAM, DBAM

Ongoing Work: Data Analysis and Risk Assessment

Data Dictionary

Utility characteristics

Variable Name	Description	Data Type	Units
plant_id	Alpha code denoting plant sample was collected from	String	NA
plant_capacity	Daily treatment capacity of plant	String	MGD
population_served	Number of people served by plant	String	NA
treatment_train	Each step of treatment train, in order, separated by semicolon (e.g., "Raw Water; Coagulation; Flocculation; GAC; DS")	String	NA
disinfectant_type	Type of residual disinfectant used	Factor	{Cl2; NH2Cl}
disinfectant_dosage	Target dosage of residual disinfectant	Numeric	mg/L
PAC_used	Whether Poly Aluminum Chloride is used in treatment	Binary	{Y, N}
PAC_type	Type of Poly Aluminum Chloride used	Factor	
PAC_brand	Brand of Poly Aluminum Chloride used	String	
PAC_dose	Target dose of Poly Aluminum Chloride	Numeric	mg/L
coagulant_type	Type of coagulant used in treatment	Factor	
coagulant_dose	Dose of coagulant used in treatment	Numeric	mg/L
polymer_type	Type of polymer used in treatment	Factor	
polymer_dose	Dose of polymer used in treatment	Numeric	mg/L
pre_oxidant_type	Type of pre oxidant used in treatment	Factor	
pre_oxidant_location	Location pre oxidant is added in treatment	String	
pre_oxidant_dose	Dose of pre oxidant used in treatment	Numeric	mg/L
gac_used	Whether Granular Activated Carbon is used in treatment	Binary	{Y, N}
gac_ebct	Empty Bed Contact Time for GAC treatment	Numeric	hours
uv_used	Whether Ultraviolet disinfection is used in treatment	Binary	{Y, N}
uv_dose	Dose of Ultraviolet disinfection used in treatment	Numeric	mJ/cm^2

Sampling results

Variable	Description	Data Type	Units	Source
sample_date	Date sample was collected	Date	YYYY-MM-DD	COC
plant_id	Alpha code denoting plant sample was collected from	String (single letter)	NA	COC
disinfectant_type	Type of residual disinfectant used	Factor	{Cl2; NH2Cl}	Utility Survey
sample_location_number	Number denoting sample location - specific to each plant	Integer	NA	COC
sample_location_type	Type of sample location	Factor		COC
sample_site	Brief sample location description	Factor		COC
sample_id	Unique sample identification ID	String	NA	COC
hrt	Hydraulic retention time	Numeric	days	Analytical Results
disinfectant_residual	Concentration of residual disinfectant in sample	Numeric	mg/L	Analytical Results
ph	Measured pH of sample	Numeric	NA	Analytical Results
doc	Concentration of dissolved organic carbon found in sample	Numeric	mg/L	Analytical Results

Planned Analyses: Sampling Data

- **Spearman correlation analysis – WQ variables**
  - Assess co-occurrence between measured variables
- **Principal Component Analysis (PCA) – WQ variables**
  - Identifies components which explain the most variance in data
- **Generalized linear, LASSO and RIDGE regression – WQ variables and epidemiological data**
  - Establish predictive relationship between all WQ variables and Legionellosis case data
  - LASSO allows for L1 regularization (reducing variance) to prevent overfitting
  - RIDGE allows for L2 regularization (addressing multicollinearity) to prevent overfitting

Variable	pH	Temp	DO	Total.Cl	Free.Cl	TOC	DOC	Alka	TTHM	TCC	HPC
Temp	0.09	-									
DO	-0.30	-0.40	-								
Total.Cl	-0.20	-0.51	0.27	-							
Free.Cl	-0.19	-0.41	0.22	0.79	-						
TOC	0.19	0.49	-0.42	-0.35	-0.25	-					
DOC	0.17	0.53	-0.45	-0.42	-0.27	0.97	-				
Alka	0.07	0.31	-0.14	-0.21	-0.12	0.36	0.36	-			
TTHM	0.24	0.20	-0.34	-0.29	-0.33	0.65	0.62	0.32	-		
TCC	-0.06	0.48	-0.23	-0.28	-0.14	0.53	0.56	0.56	0.26	-	
HPC	0.19	0.46	-0.35	-0.30	-0.16	0.61	0.60	0.50	0.37	0.70	-
Leg.sp	0.15	0.35	0.08	-0.48	-0.32	0.53	0.52	0.54	0.33	0.54	0.62

Spearman correlation analysis – example data from previous project

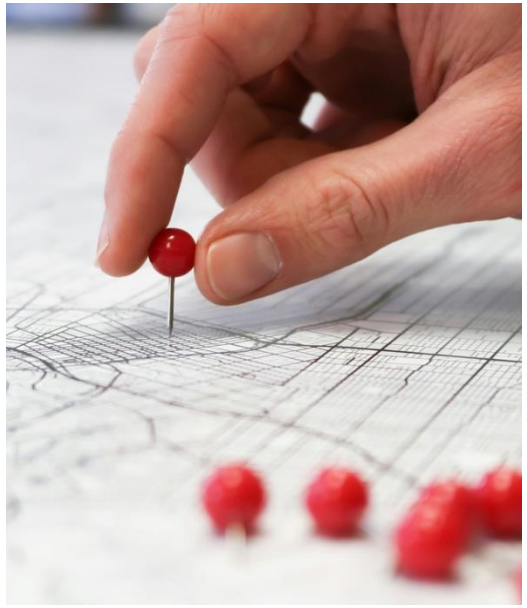
Current Projects

Goal: Determine statistical link between water quality parameters and Legionellosis cases

1. Sampling WQ data and CDC case counts

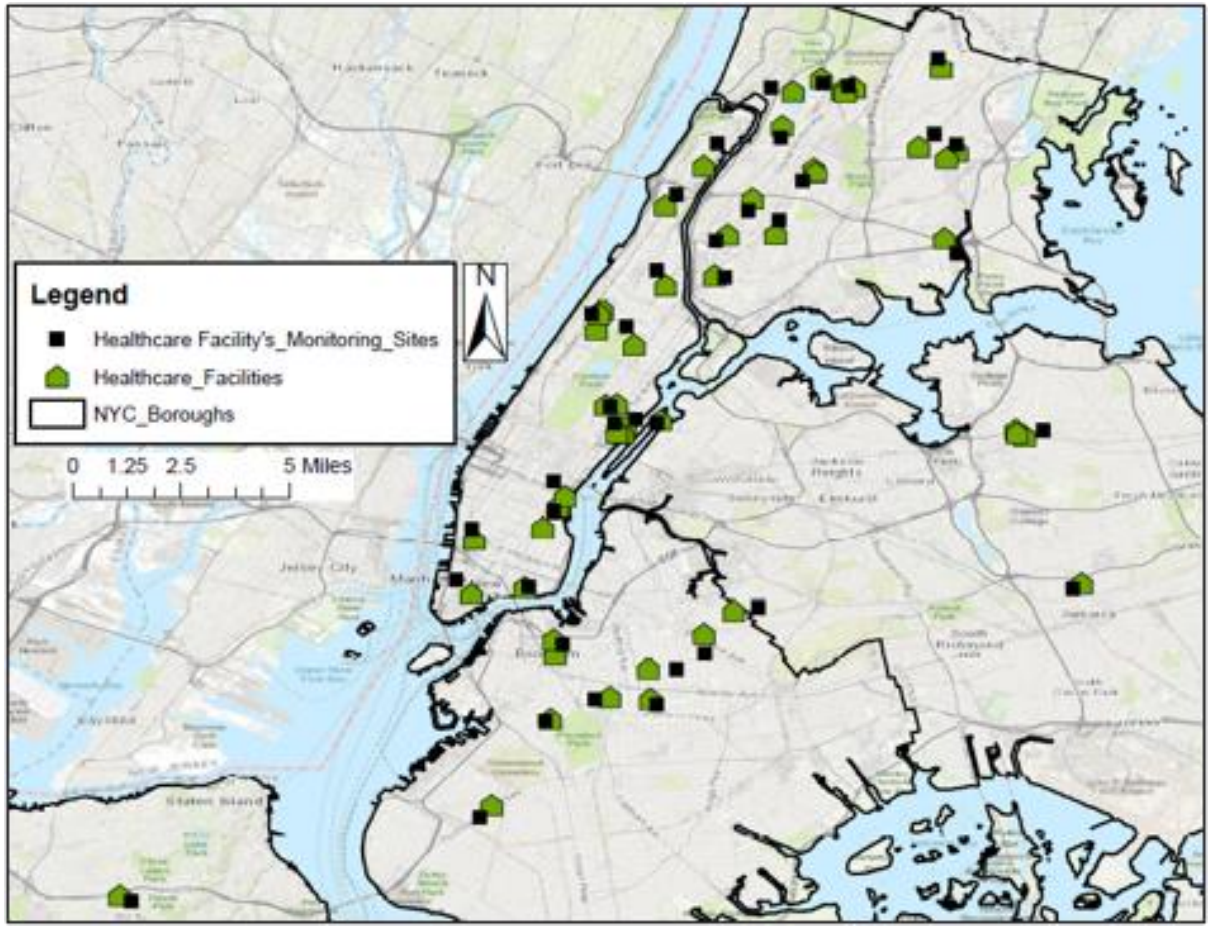
2. NYC sidewalk WQ samples and outbreaks in healthcare facilities

3. County-wide utility WQ data with county-wide case counts



Progress: NYC Project

- Currently no strong relationships between PWS WQ parameters and HCF *Legionella* percent positivity
- *Legionella* occurrence varies among facilities with different in-house water treatment types
  - Type of in-house treatments (chlorine dioxide, chlorination, monochloramine, CSI, none) might have significant influence

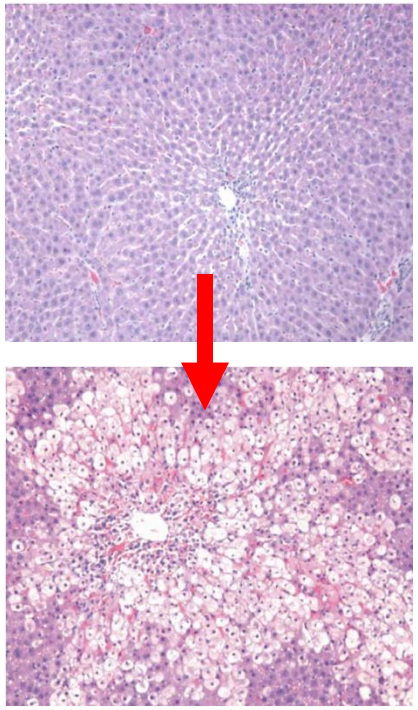


NEW YORK STATE OF OPPORTUNITY. Department of Health

Risk Assessment – Step One

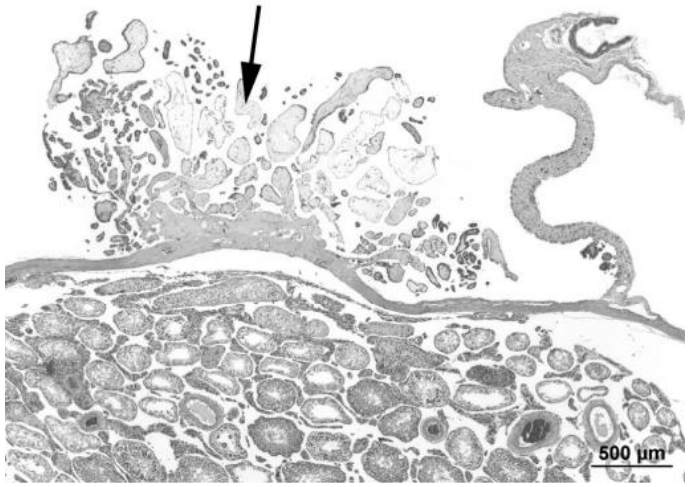
Identify possible human health hazards associated with each DBP and OP

Chloroform
Liver necrosis



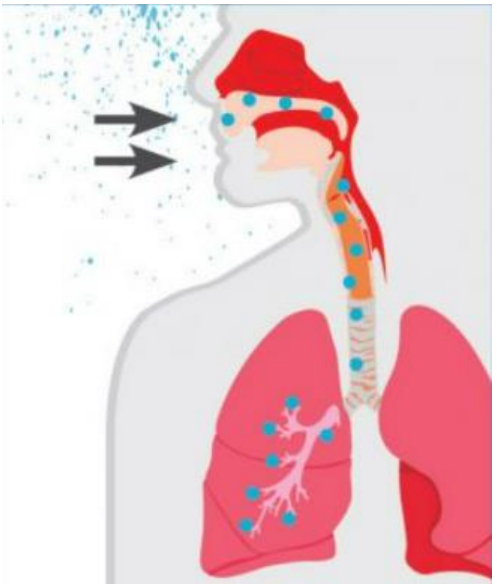
Late administration of COX-2 inhibitors minimize hepatic necrosis in chloroform induced liver injury. 2003. Begay & Gandolfi.

Bromochloroacetic acid
Malignant mesothelioma



NTP technical report on the toxicology and carcinogenesis studies of bromochloroacetic acid (CAS no. 5589-96-8) in F344/N rats and B6C3F1 mice (drinking water studies). 2009. NTP

Legionella pneumophila
Legionnaire's Disease



State of Hawaii Department of Public Health. https://health.hawaii.gov/docd/disease_listing/legionellosis-legionnaires-disease/

Risk Assessment – Step Two

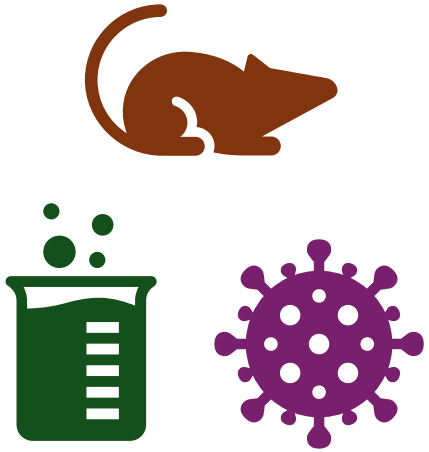
Predict probability of response (illness, death, etc.), given a known dose

Published reference doses (DBPs)

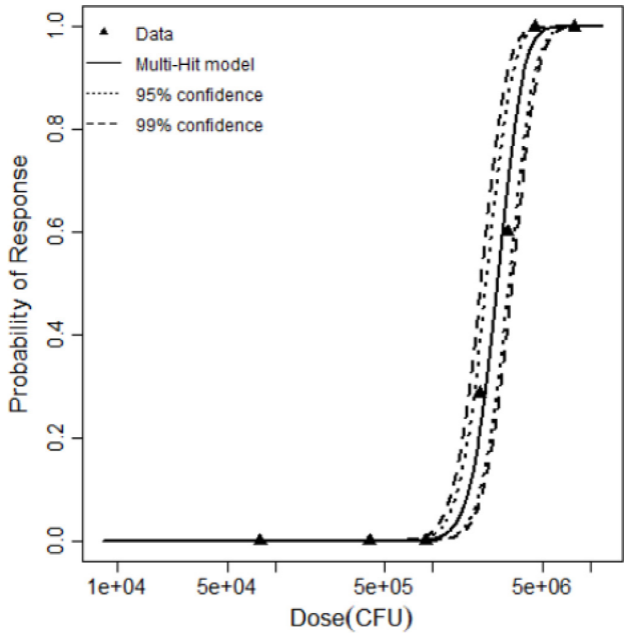
Critical Effect	Experimental Doses*
Moderate/marked fatty cyst formation in the liver and elevated SGPT	NOAEL: none LOAEL: 15 mg/kg/day (converted to 12.9 mg/kg/day)
Dog, chronic oral bioassay	
Heywood et al., 1979	

Chloroform; CASRN 67-66-3. 2001. IRIS, US EPA.

Animal study data (DBPs & OPs)



Existing models (OPs)



A dose response model for the inhalation route of exposure to *P. aeruginosa*. 2020. Dean & Mitchell.

Risk Assessment – Step Three

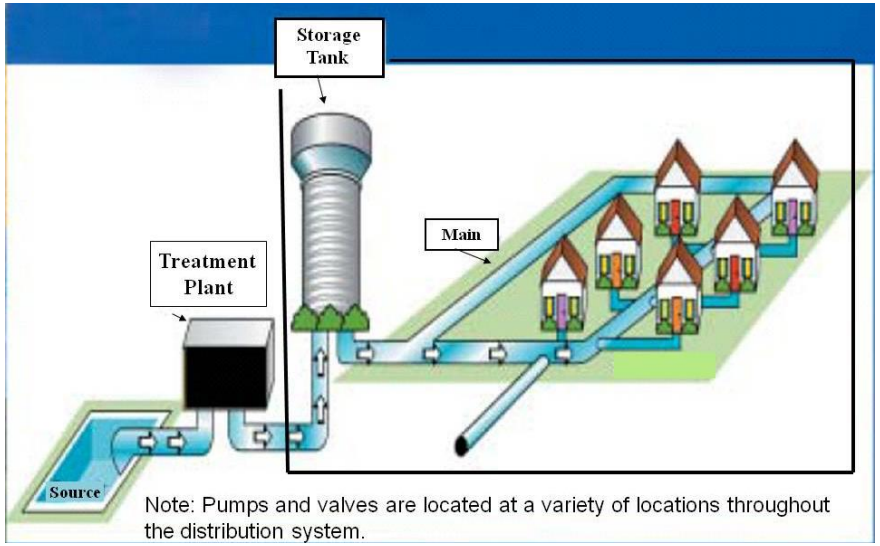
Determine amount (dose) of DBP/OPs communities are exposed to via drinking water

Occurrence

Sampling data

Transport

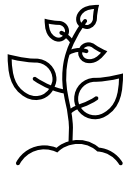
Consider distribution system



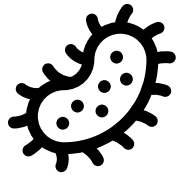
Drinking Water Distribution Systems. 2023. EPA

Removal/persistence

Consider factors affecting dose



Adsorption



Native community

Use and Exposure Factors

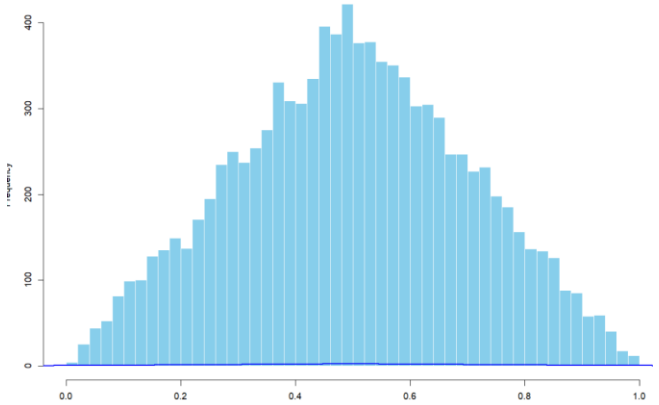
Consider factors affecting exposure

Risk Assessment – Step Four

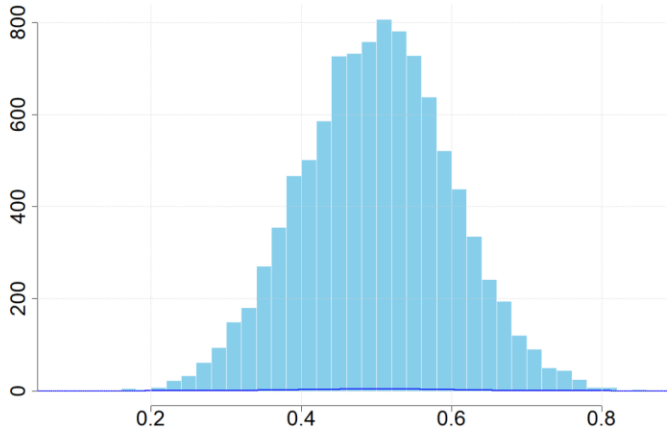
Estimate risk based on exposure dose

Consider…

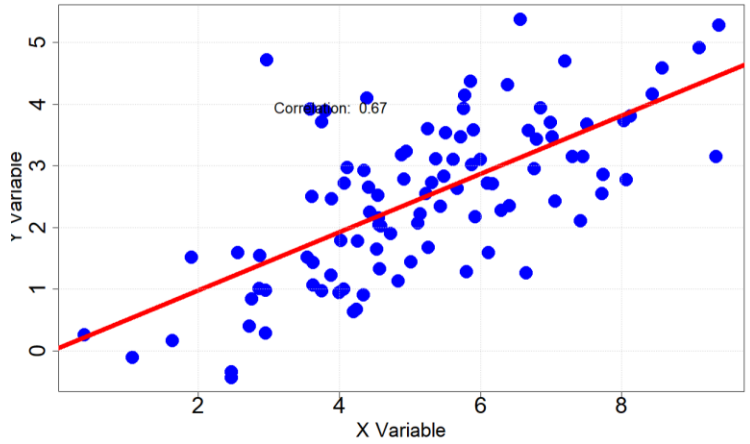
Ranges for possible inputs
Ex: Values for TOC



Risk distribution



Correlations between variables
Ex: WQ and risk



Risk Assessment: Ongoing

DBP dose-response models

- Literature review for 39 DBPs
- Whole-animal data to create models for 13 DBPs
  - Multiple endpoints (carcinomas, adenomas, necrosis, etc.)
- Beginning modeling process

TABLE 10
Incidences of Malignant Mesothelioma in Male Rats in the 2-Year Drinking Water Study of Dibromoacetic Acid

	0 mg/L	50 mg/L	500 mg/L	1,000 mg/L
Malignant Mesothelioma[a]				
Overall rate[b]	3/50 (6%)	1/50 (2%)	0/50 (0%)	10/50 (20%)
Adjusted rate[c]	6.9%	2.4%	0.0%	22.6%
Terminal rate[d]	2/34 (6%)	1/24 (4%)	0/30 (0%)	2/28 (7%)
First incidence (days)	591	729 (T)	—[f]	512
Poly-3 test[e]	P<0.001	P=0.325N	P=0.137N	P=0.035

Example data included in analysis

NTP. 2007. Dibromoacetic acid (Cas no. 631-64-1) in F344/N rats and B6c3f1 mice (Drinking water studies)

Ongoing Work: Collaboration and Engagement

Collaboration and Engagement

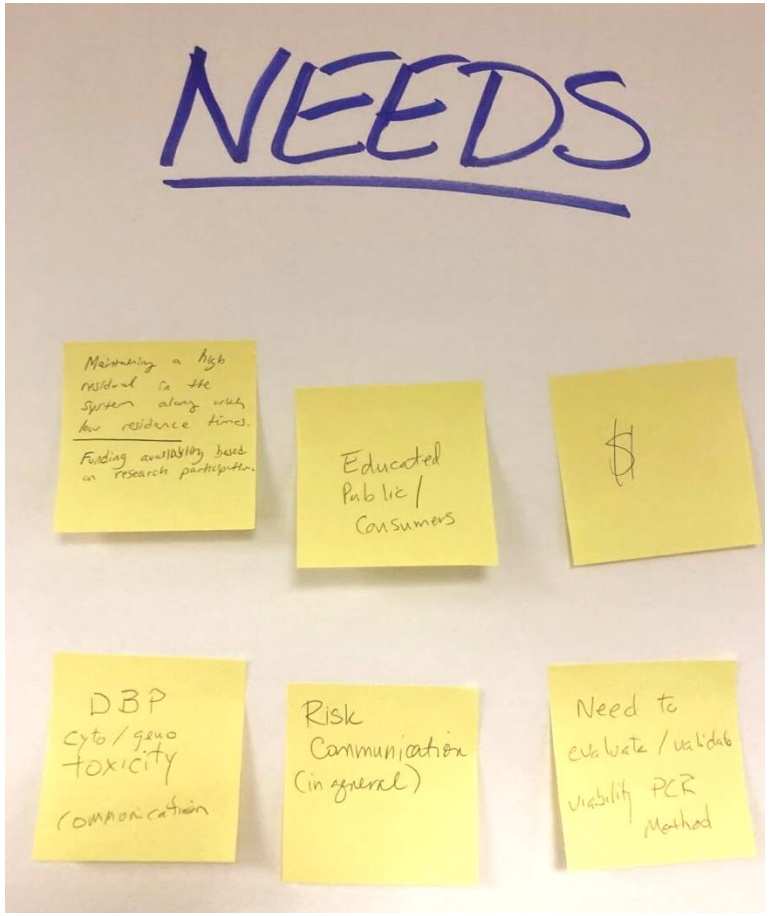
Key takeaways from June workshop with 14/25 utility partners

Challenges

- New regulations
- Poorly understood testing requirements
- Changing system characteristics
- Sustainable staffing
- Lack of resources
- Communication

Needs

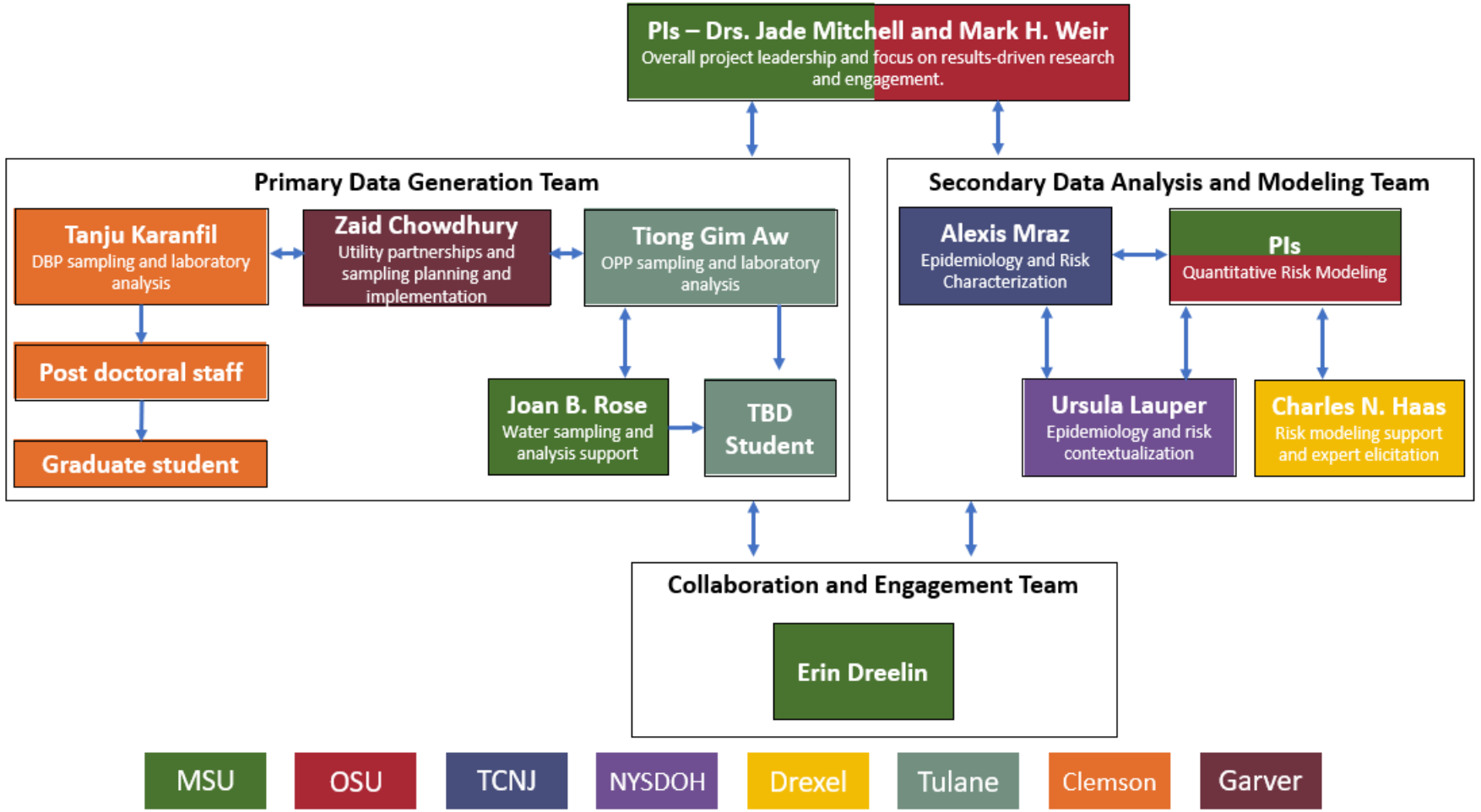
- Better understanding of relationships between DBPs, OPs, & other contaminants
- Co-treatment options
- Technical assistance
- Workforce development
- Risk-based communication strategies



Future Work

- Complete full-scale summer/fall sampling and data analysis
  - Conduct further sampling in utilities with positive DBP/OP levels (Year 2-3)
- Begin DBP risk analysis and DBP-OP risk trade-off assessment
- Develop risk communication documents with utility partners
  - Put occurrence values in context of risk
- Webinars
- Consider centralized database and supporting unified analysis of project results

Thank You!



Contact us: bae.drinkingwater@msu.edu

Regressions Explained

- All regressions are a means of describing data mathematically to make accurate predictions in the real work
- This inherently causes bias and variance, both diverge from the real values.
- If there are more than one variables then an effect called multicollinearity can cause poor estimates.
- LASSO is a regression method that reduces variance on the estimates.
- RIDGE is a regression method that reduces the impact of multicollinearity on the estimates
- Both of these are done on the raw data with no need for pre-processing, standardization, normalization or other data manipulation methods.