Large-scale Data Integration and Harmonization to Accurately Predict Sites Facing Future Health-based Drinking Water Crisis

Presenter - Jason Westra[1,2]

N. Tintle[1,2], A. Best[2], B. Krueger[2], K. Vande Griend[2], A. Slater[2]

1- Superior Statistical Research

2- Aquora Research and Consulting

Organizations

Superior Statistical Research (SSR)

Large Data applications

Statistical Methods Development

Aquora Research & Consulting

Mission:

- Equipping individuals, communities, and organizations to make data-driven decisions about their water

Setting

At least 45 million people per year in the U.S. are directly impacted by health-based drinking water problems.

Many of these water problems are the direct result of managerial negligence, inconsistent monitoring, and a lack of the ability to anticipate where problems may arise next.

Can we move past a reactionary approach and move to a pre-emptive approach.

Project Overview

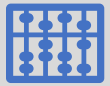
NIH via SBIR grant

Iowa Bioconnect grant

Aim 1. We will demonstrate that it is possible to predict community water systems with water quality-based health problems, utilizing existing, historical data on water quality and municipal water system performance

Aim 2. We will demonstrate how making these predictions can be leveraged to profitability.

Goals



Harmonize disparate data on water quality and water systems performance in Michigan and Iowa

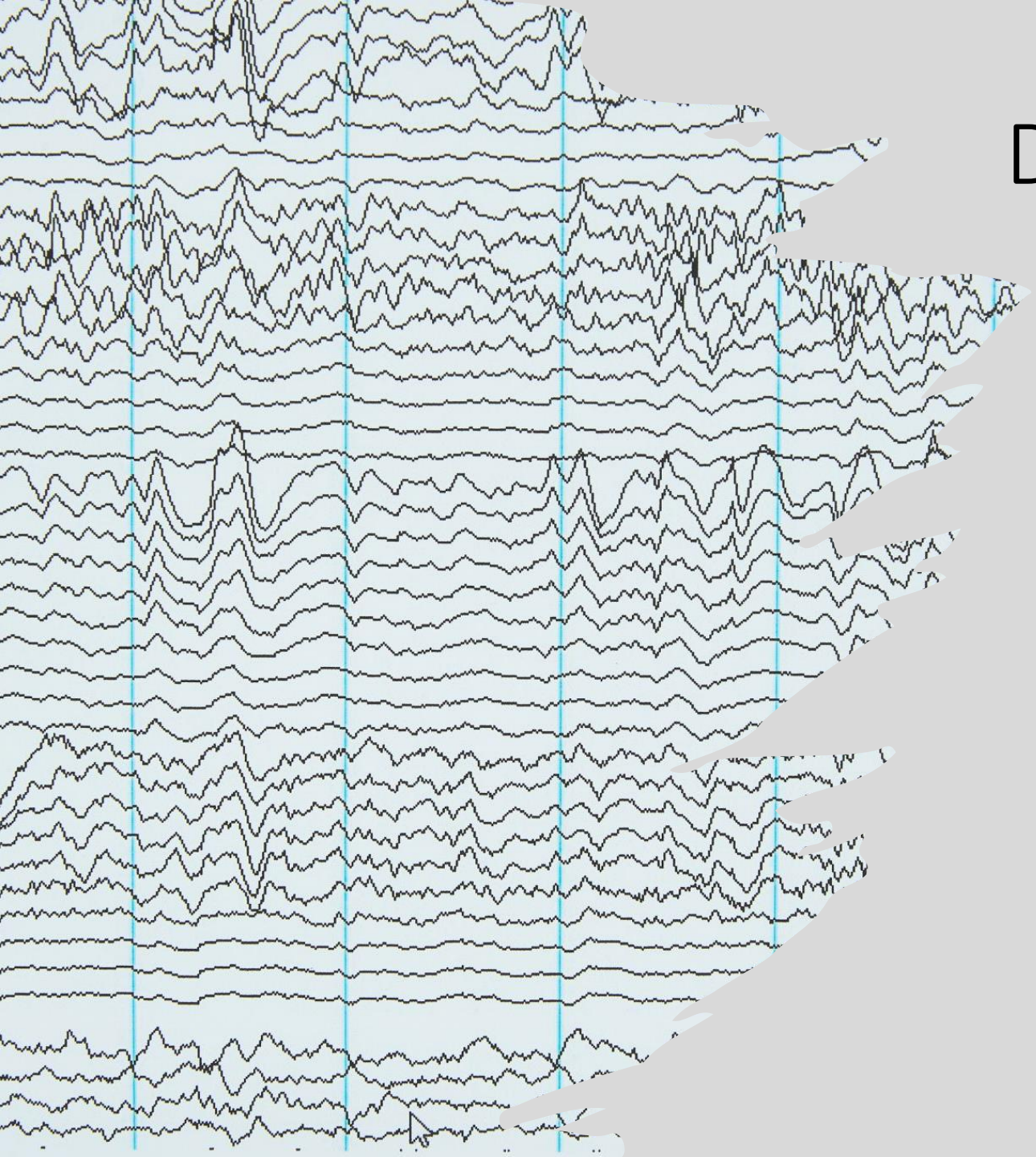
Use Machine Learning models to predict Health Violations

Included subthreshold water testing data to increase predictability of models

Validate models AUC>80% Ranking of Flint MI

Identify 10 municipalities at highest risk for future water quality health-based violations

Conduct systematic water sampling in the 10 communities



Datasets

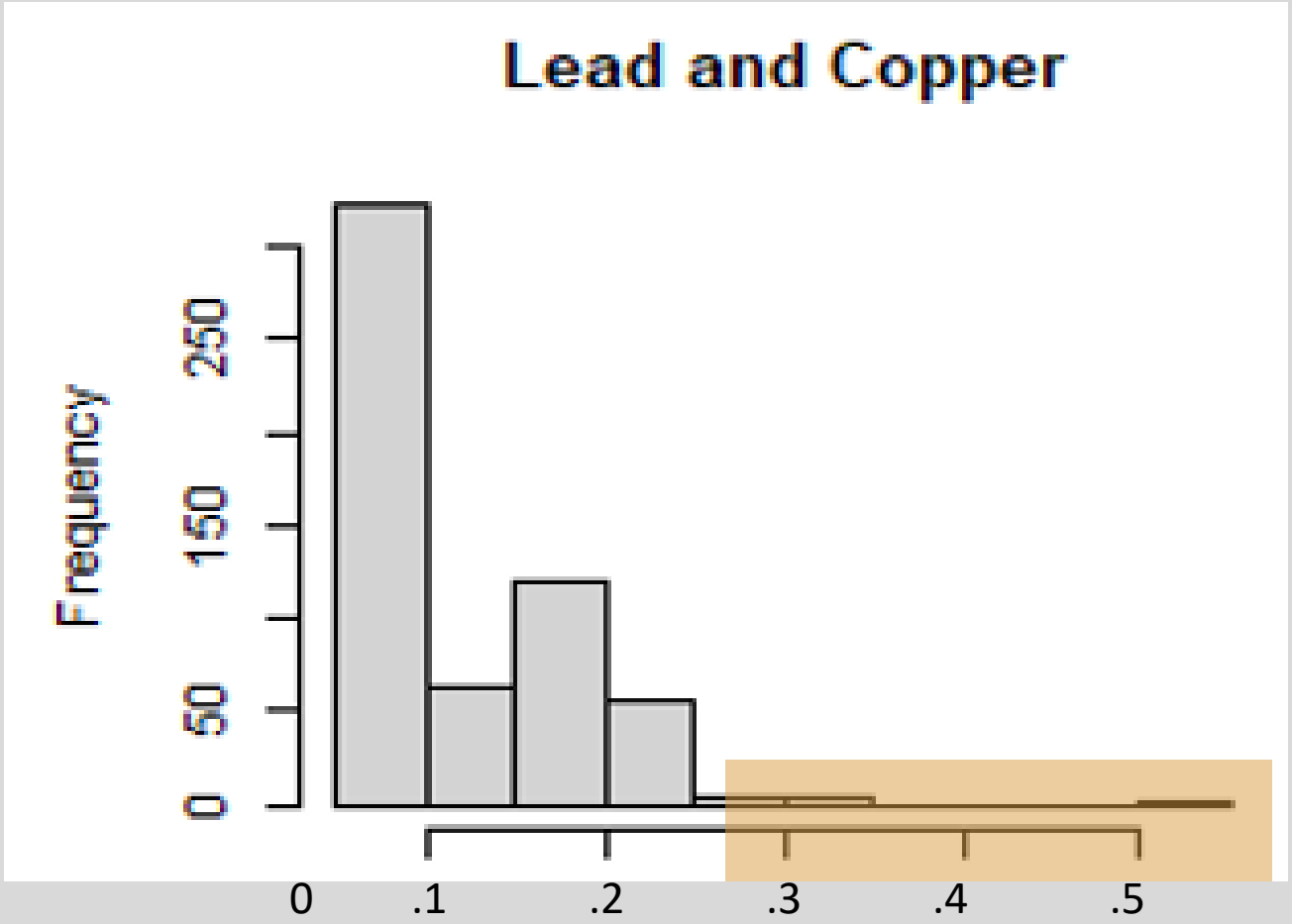
- Datasets
  - Safe Drinking Water Information System (SDWIS)
    - 80997 Violations in Michigan from 1979 – 2020
    - 26270 Violations in Iowa from 1975 - 2020
  - Publicly available contaminant measurements
    - 51858 data points from Iowa

Models

- Using machine learning we developed prediction models on which communities will have future health violations
- Developed models for different contaminants
- AUC for models ranged from 0.758 to 0.976.
  - Only one model below the goal of 0.800

Model predictions

- Predict probabilities of health-based violations
- Able to rank predictions
- Look for outliers

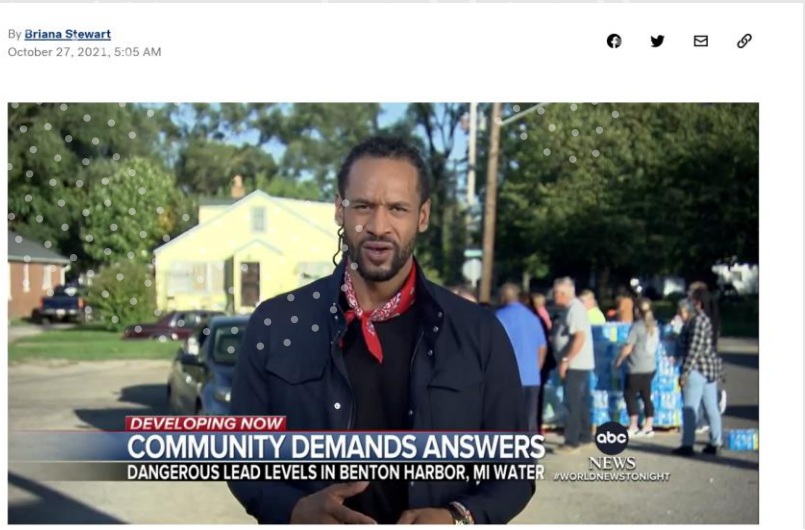


Model Validation

- Selected 10 communities with elevated risk
  - Mixture of urban and rural locations
  - Mixture of water profiles
  - 5 Michigan communities
  - 5 Iowa communities
- Selected 10 communities with low risk
  - Matched community profiles to elevated risk

Anecdotal Model Validation

- Models for Michigan were made 3/12/2021
- Largest outlier location for multiple contaminants was Benton Harbor
- Selected 4 other communities
- Prepared to send sampling teams out in fall of 2021
- Team arrived to test and survey 10/16/2021



abc NEWS VIDEO LIVE SHOWS ELECTIONS

Benton Harbor's water crisis highlights failing infrastructure's impact on the poor

Residents have been forced to use bottled water due to lead contamination.

By Briana Stewart
October 27, 2021, 5:05 AM

City of Benton Harbor Water Distribution

It is SMCAA's belief that clean water is vital to the health of communities and that everyone has the right to safe, clean water.

Since September 30, 2021, SMCAA has been offering bottled water to the residents of the City of Benton Harbor following reports of elevated levels of lead in the city-supplied water. The State of Michigan is providing free bottled water as City residents are being encouraged to use bottled water for cooking, drinking, brushing teeth, rinsing foods, and mixing powdered infant formula.

Helpful Links

- Michigan.gov's Mi Lead Safe Homepage
- Apply for Home Lead Services from MDHHS
- Benton Harbor Drinking Water Response

Water Tests and Surveys

Tested water in up to 10 homes and 5 business in each community

Sent water samples to an independent water testing facility

Targeted tests toward predicted contaminants

Administered water perception surveys

Trust in water system

Willingness to spend for filtration

Willingness to spend for testing

Water Testing Results

Michigan

- High Risk Communities
  - 3 of 5 communities had samples above the MCL in targeted contaminants
- Low Risk Communities
  - 0 of 5 communities had samples above the MCL in targeted contaminants

Iowa

- High Risk Communities
  - 4 of 5 communities had samples above the MCL in targeted contaminants
- Low Risk Communities
  - 0 of 5 communities had samples above the MCL in targeted contaminants

Survey Perceptions (N= 110)

Water satisfaction	Very dissatisfied 34.5%	Dissatisfied 16.4%	Satisfied 21.8%	Very Satisfied 27.3%	
Trust in water system	Don't 19.1%	Not Very 16.4%	Trust 34.5%	Much Trust 30%	
Likely to spend on testing	No money 25%	Nothing 35%	>$200 at least once 3%	Up to $200 at least once 13%	$25+/ month 25%
Likely to spend on filter	No money 19%	Nothing 28%	>$200 at least once 9%	Up to $200 at least once 10%	$25+/ month 34%

Future Opportunities

Expand predictions to more states

Expand predictions to other contaminates

Integrate more data sources into the models

Other strategic opportunities

Future Partnerships

- Align with organizations to help fix the issues
- Align with communities that need assistance and connect them to resources
- Educate communities about solutions that can provide relief now before issues arise, and before infrastructure change.

Acknowledgements

- NIH grant #1R43ES033134-01 (Tintle PI)
- Iowa Bioconnect