DARTS - Detection of Anomalies in Real Time Surveillance Ongoing research on anomaly detection in public health

- Sarah Holte, PhD. Mathematician/Biostatistician. Fred Hutch Cancer Center and U of Washington.
- Roxanne Kerani, PhD. Epidemiologist and Public Health Scientist. U of Washington
- Hao Yan, PhD. Industrial Engineering Statistics, School of Computing, Informatics, and Decision Systems Engineering, Arizona State University.
- Yajun Mei, PhD. Industrial and Systems Engineering, Georgia Institute of Technology
- Joshua Herbeck, PhD. **Phylogenetics** Bill and Malinda Gates Foundation.
- Steven Erly, PhD. Epidemiology and Surveillance, WA DOH State.
- Roberto Rafftery. Informatics and Software Development Senior at U of Washington.

Scientific Research Motivation

- Rapidly detect unexpected groups/anomalies/clusters of individuals with a specific disease or other public health risk in time and location
- Existing cluster detection methods for HIV (and previously for COVID) are based on **molecular phylogenetics** and contact tracing.
- For HIV, **molecular data is missing** in about half the population. Often no molecular data from marginalized populations such as those living homeless or without health insurance.
- According to WA DOH, the CDC's (required) *HIV-TRACE-phylogenetic* cluster approach is **lacking and difficult to implement**.

Scientific Research Focus (details)

- In 2017, the WA DOH missed a new HIV cluster in North Seattle among injection drug users using HIV-TRACE.
- In Washington state, for HIV, nearly half (~47%) of the incident population is missing sequence data.
- Sequencing analyses and cluster ID takes weeks or months. For HIV, sequence results take, on average, 6 months to complete. Clusters could have dissipated, grown, shrunk, or merged.

Our current goal

Identify HIV clusters using real time, streaming case-report data from the WA DOH.

- Rapidly identify spatial/temporal clusters from individuals who seek health care.
- Note: The transmission chain is lost.
- Therefore, we aim to combine our approach with transmission/phylogeny whenever available.
- Rigorous assessment and comparison is major part of the ongoing research.

Our approach adapts to *many* different settings and diseases

- Multi-drug resistant gonorrhea
- HIV
- Monkey pox
- Covid
- Syphilis in women
- Major crime occurrences
- Tainted drug overdose, Fentanyl overdoes deaths is a major public health problem in Seattle. Currently contacting Seattle/King County Public Health to obtain data.

Multi-disciplinary team with 5+ years collaboration.

Sarah Holte, PhD. Deterministic mathematical models, nonlinear dynamical systems and associated statistical methods. Fred Hutch Cancer Center and U of Washington.

Roxanne Kerani, PhD. Epidemiology, Phylogenetics and Public Health Science. U of Washington

Hao Yan, PhD. Stochastic Process Development, Computational challenges, Tensor decomposition of high dimensional data. Arizona State University.

Yajun Mei, PhD. Engineering Statistics and Biostatistics, Statistical Change Point Methods Georgia Institute of Technology

Josh Herbeck, PhD. Phylogenetics Bill and Malinda Gates Foundation.

Steven Erly, PhD Epidemiology and Surveillance. WA DOH State.

Roberto Rafftery Informatics and Software Development Senior at U of Washington.

Dr Holte's contributions

Expertise in nonlinear deterministic dynamical systems

Complicated dynamics, description of the underlying deterministic transmission process.

Formal statistical assessment and inference

Research on model validation.

Used to develop the population trend without anomalies or *deterministic changes that result in anomalies.*

Coordinate the activities. Ι χαν σπεακ α λιττλε βιτ οφ αλλ τηεσε σχιεντιφιχ λανγυαγεσ.

Dr Holte's contributions

Expertise in nonlinear deterministic dynamical systems

Complicated dynamics, description of the underlying deterministic transmission process.

Formal statistical assessment and inference

Research on model validation.

Used to develop the population trend without anomalies or *deterministic changes that result in anomalies.*

Coordinate the activities. I $\chi \alpha \nu \sigma \pi \epsilon \alpha \kappa \alpha \lambda_1 \tau \tau \lambda \epsilon \beta_1 \tau \circ \phi \alpha \lambda \lambda \tau \eta \epsilon \sigma \epsilon \sigma \chi_1 \epsilon \nu \tau_1 \phi_1 \chi$ $\lambda \alpha \nu \gamma \nu \alpha \gamma \epsilon \sigma$. (I can speak a little bit of all these scientific languages.)

Dr Yan's contributions

Expertise in Stochastic Process Development including appropriate statistical stochastic process (e.g Normal, Poisson).

Expertise in computation and tensor decompositions, a data reduction and feature extraction tool. High dimensionality of spatial/temporal data.

Dr Mei's contributions

Collaborate with Dr Yan on determining the appropriate statistical approach for the stochastic processes used.

Expertise on *changes in statistical distribution that result in anomalies* with the shortest detection delay while maintaining a certain false alarm rate requirement.

Dr Kerani's contributions

Expertise in Epidemiology of HIV and sequence analysis.

Facilitate interaction with Washington Department of Health

Dr Erly's contributions

Provide real-time data on HIV case report from Washington State Expertise on currently used surveillance methods for comparison.

Will be the end-user of the approach(es) we develop.

Mr Rafftery's contributions

Develop R-shiny and web page interface for the end users and demonstrations

Tensor Spatial-temporal Data Decomposition Method

Goal: Sparse clustered anomaly detection under smooth spatio-temporal background



 $\operatorname{argmin}_{\theta,\theta_a} \quad \lambda \theta' R \theta + \gamma |\theta_a|_1 + ||e||^2, s.t. \ y = (B_s \otimes B_t) \theta + (B_{as} \otimes B_{at}) \theta_a + e$





Spatial/Temporal mean: Holte (deterministic), Yan and Mei (stochastic)

Anomalies: Holte (Deterministic in underlying model dynamics), change in statistical distribution (Mei), development of the anomaly process (Yan and Mei) computation using tensor methods for dimension reduction (Yan).

Noise: Holte, Yan and Mei

Results: Tensor Decomposition for Anomaly Detection

- Gonorrhea: The second most commonly reported notifiable disease in USA
- Goal: Anomaly detection to better allocate resources.



Additional Results and Future Plans

Published COVID in Washington State

Major Crimes (Murder, Rape, Aggravated Assault) NYC. Publicly available data. FBI.

In Preparation HIV in Washington State

Future Plans

Overdoses and Deaths due to Fentanyl or other drug contamination.

Syphilis in Women.

Thank you for your time and attention and to the Symposium organizers and participants. Contact: Sarah Holte, <u>sholte@fredhutch.org</u>, sholte@uw.edu



Supported by:

- UW/FH Center for AIDS Research (CFAR) P30AI027757-34 and -36
- Novel threat detection methodology to detect HIV outbreaks in Washington
- R21R21 AI157618-01A1