Assessing incomplete sampling of disease transmission networks

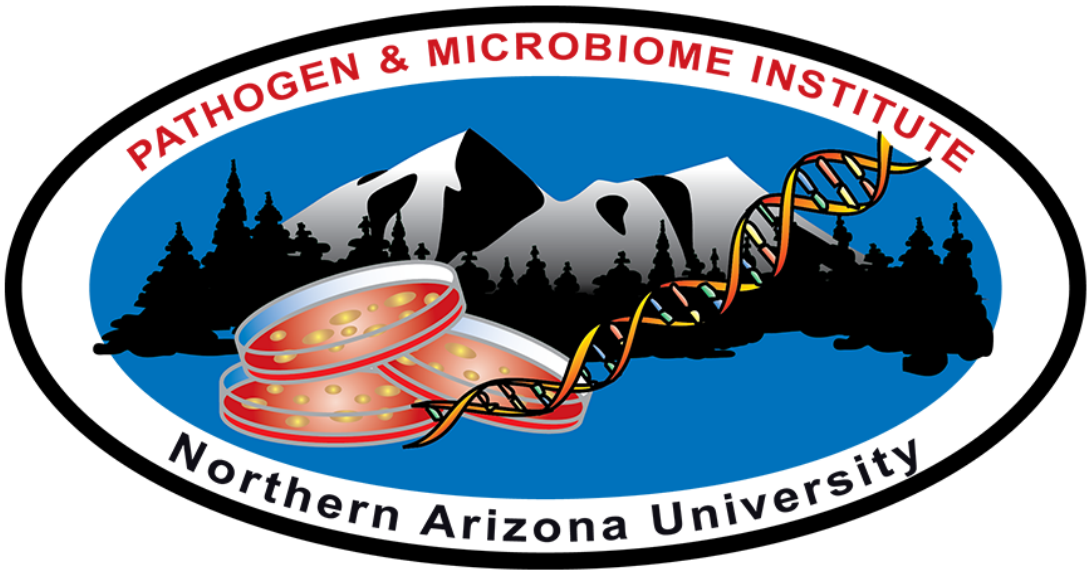
Meeting with Paul

Derek Sonderegger, PhD - Northern Arizona University

April 9, 2019

Colaborators

- Work that I have done with the Pathogen and Microbiome Institute at NAU and we are just a couple months into the project.



- Dr Paul Keim
- Dr Jason Sahl

Background Information

Two worrisome Healthcare Aquired Infections (HAIs)

- MRSA
  - Methicillin-resistant Staphylococcus aureus
  - Resistant to many common antibiotics
- *C. Diff*
  - *Clostridioides difficile*
  - Our disease of interest

Clostridioides difficile

- A spore-forming bacteria
  - Spores can survive for months in the environment
  - Bacteria die when exposed to oxygen.
  - Very difficult to work with in the lab.
- *C. diff* is widely distributed
  - Spores are widely found in the environment
  - People and animals can be asymptomatic carriers
- Resistant to many commonly used antibiotics

Human Infection

- Causes diarrhea, fever, nausea, and abdominal pain
- Spread through fecal contamination
- Additional $4.8 billion each year in health care costs
  - 290,000 Americans sickened by the bacteria in a hospital or other health care facility each year.
  - 27,000 people in the U.S. die while infected with *C. diff* annually.

Common infection cycle

- In a healthy gut biome, *C. diff* can't strongly establish due to bacterial competition.
- In patients under a common antibiotic treatment, *C. diff* can flourish.
- Prescribed antibiotics for some other reason (e.g. pneumonia)
  - *C. diff* might already be present in the patient.
  - Come into contact with *C. diff* via live bacteria or spores from another patient.

Medicare Implications

- Won't reimburse costs for treating infections acquired at a healthcare facility
- If the rate of Healthcare Acquired Infections (HAIs) is too high, Medicare will deduct one percent from their OVERALL reimbursements to the facility.
- Medicare defines any diagnosis of *C. diff* that occurs 3 days after admission as "healthcare acquired".

Goal: Estimate HAI rate

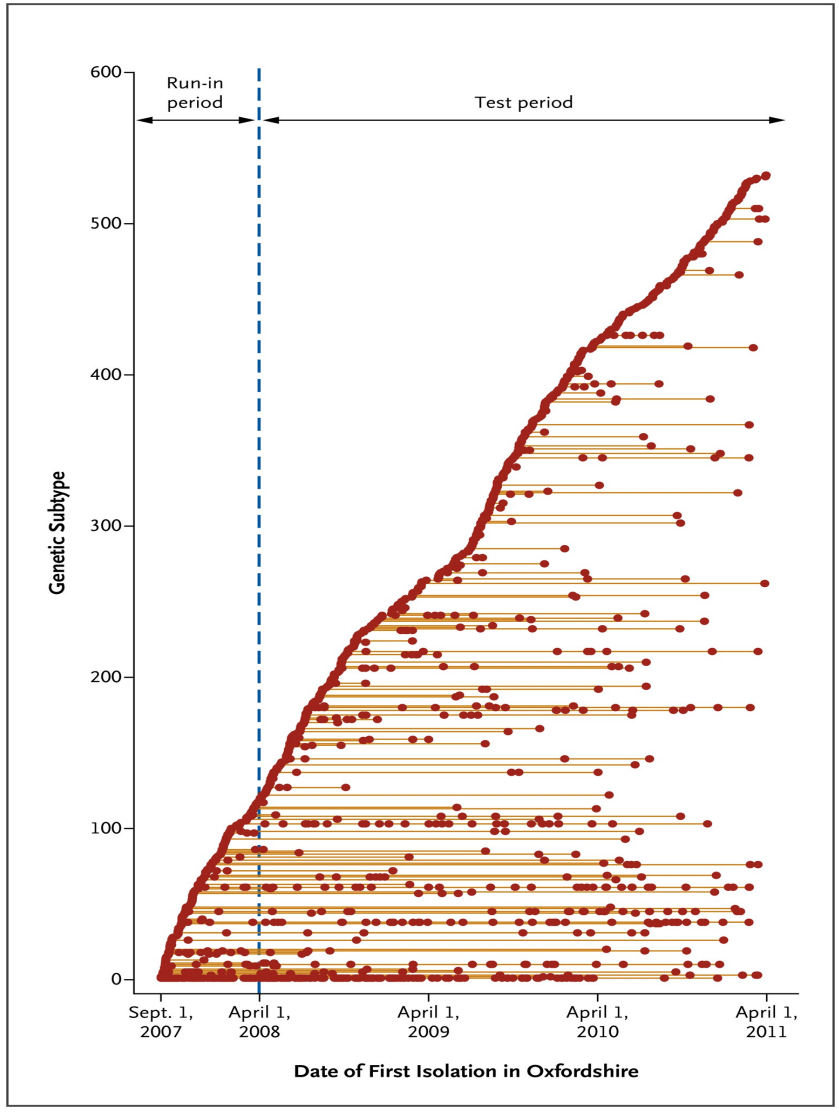
- Individual patients have the genome of their strain of *C. diff* sequenced.
- Group strains into clusters if they differ by at most 2 SNPs.
  - Use Single-Linkage clustering method: represents evolution along a chain of infections
  - Within patient variability suggests that maybe this needs to be evaluated.

Data!

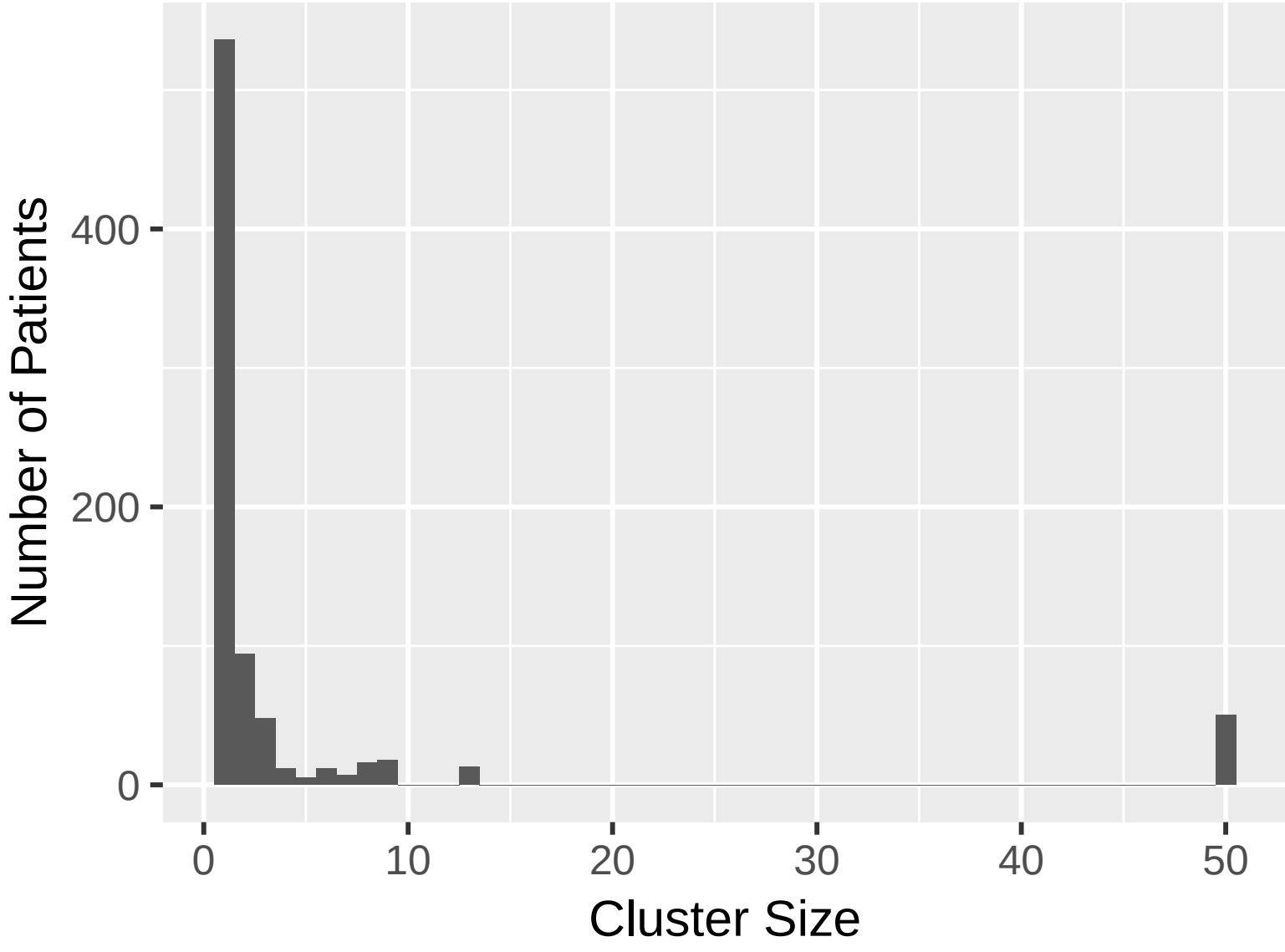
Oxfordshire Data

- Eyre *et al* 2013 describes a study which genotyped nearly all cases of *C. diff* in over three years in Oxfordshire, UK.
- Of the 1250 cases that were evaluated, $N = 1223$ were successfully genotyped.

Oxfordshire Time/Clusters



Oxfordshire Cluster Size Distribution



Defining HAI rate from full data

- For each cluster, the first time a strain is observed it is considered environmentally acquired.
- The second (or third, or fourth, ..) time a strain is observed, it is healthcare acquired.

$$HAI = \frac{N - ||\mathcal{I}||}{N} = 1 - \frac{||\mathcal{I}||}{N}$$

$$N = \text{ Number of Patients}$$

$$\mathcal{I} = \text{ Set of strain identifiers}$$

$$||\mathcal{I}|| = \text{ Actual Number of Clusters/Strains}$$

- Knowing $||\mathcal{I}||$ is the key to calculating HAI rate!

Observed Number of Clusters/Strains under Simple Random Sampling

$$\widehat{HAI}_{naive} = 1 - \frac{||I||}{n}$$

$$n = \text{ sample size}$$

$$I = \text{ Set of observed strains}$$

$$||I|| = \text{ Observed Number of Clusters/Strains}$$