Infectious diseases have impacted civilisations throughout history, shaping interactions between individuals, and influencing the growth of populations. Despite advances in the control of disease, often stemming from the development and deployment of vaccines, they still pose a substantial threat to human life. To minimize disease burden it is essential to gain a greater understanding of the underlying dynamics of disease within populations.

Here I consider how the classification of infectious disease data provides both opportunities and uncertainty in our quest to characterize and control infectious disease outbreaks.

The COVID-19 pandemic threw the world into a state of uncertainty, shutting down university campuses, towns, and cities across the United States and wider world. Using disease exposure and behavioral data collected at the Pennsylvania State University, I demonstrate how the definition of exposure groups facilitated an understanding of the infection transmission dynamics. I implement clustering models to identify groups of individuals by risk behavior, and use mathematical models of disease transmission to show the potential impacts and limits of interventions aimed to reduce transmission in this population.

Exposure and risk classes are only one phenomena that can be defined from continuous data: infection status and the accumulation of cases in a population are commonly stratified. I show that the process of discretizing individuals into a binary positive or negative infection status has compounding effects on our detection and response to outbreaks. I develop simulation models of co-circulating diseases to illustrate the opportunities that exist for the use of imperfect diagnostic tests. In doing so, I map out the broader limits of detection given constraints imposed on disease surveillance systems, such as testing rates, and the relative magnitude of the target disease against the backdrop of the co-circulating pathogen.