The Delphi Group at Carnegie Mellon University has for years developed tools to forecast epidemics like influenza. In early 2020, it faced a new challenge: How can you forecast a pandemic with only patchy disease surveillance data? Through a unique partnership with Facebook and the University of Maryland, it set out to collect its own.

Delphi's COVID-19 Trends and Impact Survey (CTIS) has since received nearly 25 million responses in the United States – or 40,000 per day for over a year -- combined with over 50 million responses received by the University of Maryland's international effort. This is the largest research survey ever conducted outside of a national census. Aggregate data is publicly available, updated daily, and has already been used to develop forecasts and guide national responses.

In this talk, I'll introduce the survey and describe its expansion from tracking COVID symptoms to collecting information on testing, mental health, social distancing, mask use, and most recently, attitudes to COVID vaccination. I'll talk about the statistical challenges in running such a large survey, open statistical problems in analyzing its results, and what the data engineering process has taught us about the role of statistics in emergency response.