Methodology

Dhaval Dave

Department of Economics

DDave@Bentley.edu

Data and Research Methods Matter

- **Dispassionate science**

- Data can quantitative ("quantities"; numbers)
  or qualitative (descriptive, observations, textual)
- Data can be primary or secondary
- Data can be observational or experimental
- Data can be time-series, cross-sectional
  or panel / longitudinal

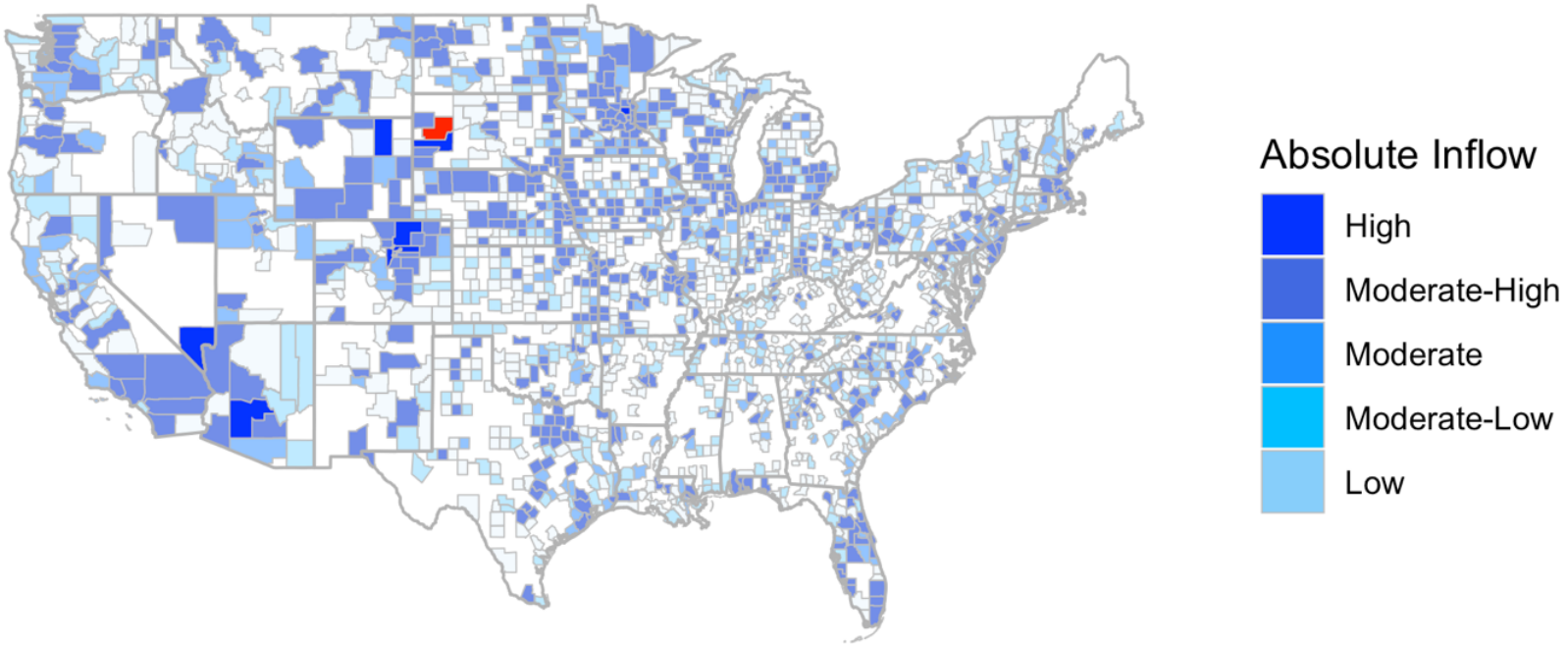
- Empirical methods can descriptive – answering questions of "What is?"
  or more advanced and causal – answering questions of "What if?"

Descriptive: Sturgis Motorcycle Rally

(Source: Dave et al. 2020, NBER WP 27813)

Figure 1. Distribution of Inflows into Sturgis Motorcycle Rally

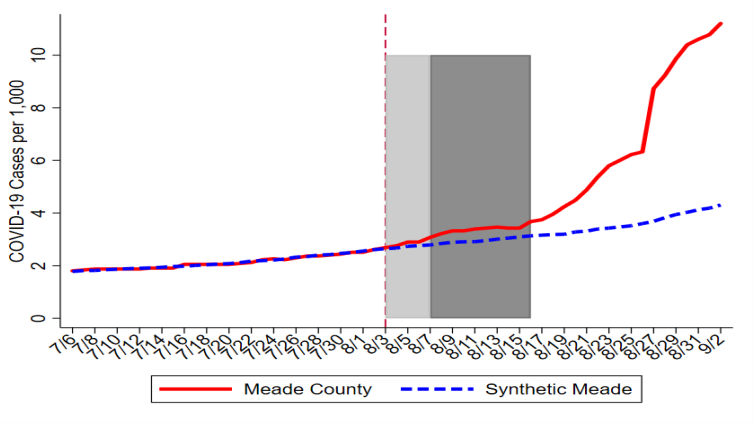
Panel (a): Absolute Intensity of Sturgis Attendee Inflow, by Resident County



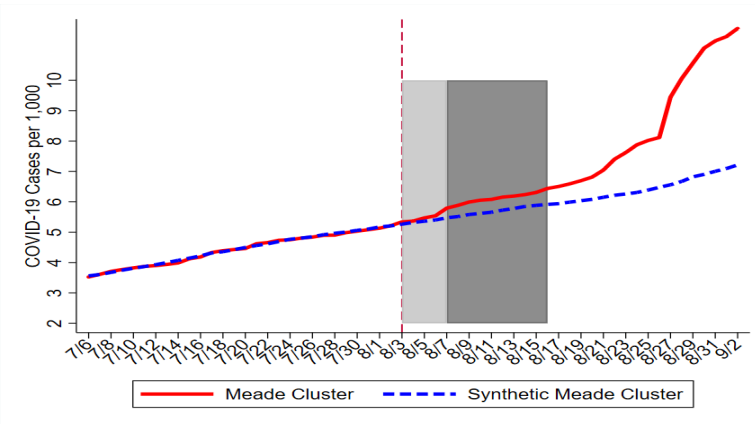
"Causal" Question: What is the impact of the Rally on Covid-19 Cases Internally & Externally?

(Source: Dave et al. 2020, NBER WP 27813)

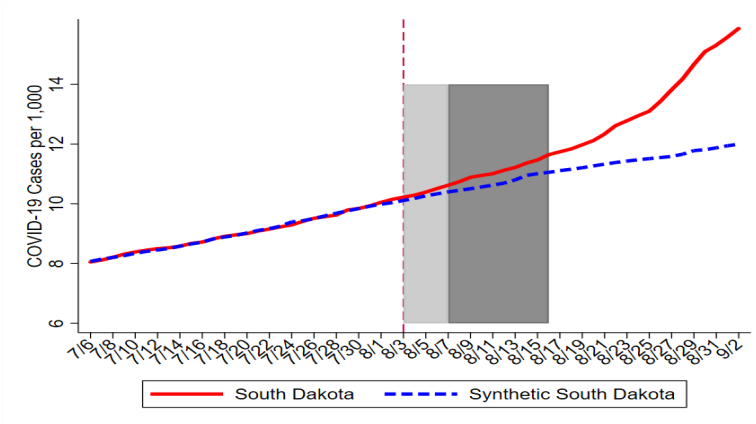
Panel (a): Meade County



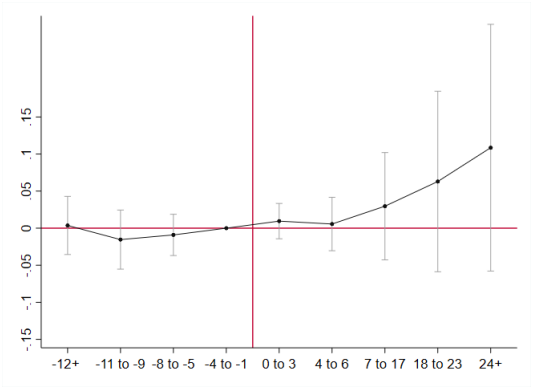
Panel (b): Meade County and Border Counties



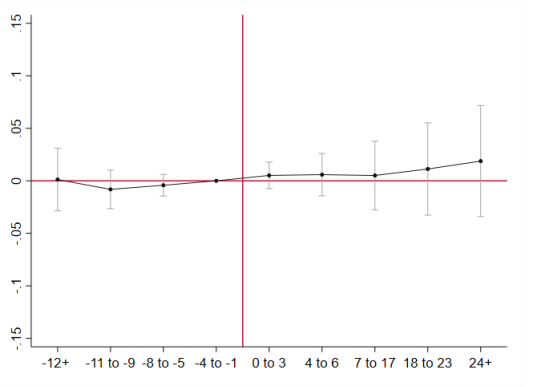
Panel (c): South Dakota



Panel (a): High Absolute Inflow



Panel (d): Moderate-Low Absolute Inflow



Methods: Synthetic Control Difference-in-Differences & Dose-Response Difference-in-Differences

Example: Cross-state Disparities in Health

Seth Wise

Data: Panel Data (1995-2014)

Methods: Panel data fixed effects methods

1) **Quantifying the factors contributing to state disparities in health outcomes?**

   Health insurance, spending, economic factors, lifestyle factors?

2) **Rank states with respect to "over-performers" and "under-performers"**

   Top: MN, CA, WI… **Bottom: AL, MS, WY…**

Example: How does a conditional cash transfer program affect children's well-being?

Alberto Usabiaga

Data: Longitudinal waves of household survey data (Mexico)

Methods: Triple Differenced (DDD) Model

1) **The program significantly improved health and schooling outcomes among children – particularly older children (17-21%)**

2) **Accounting for migration is important.**

   Not accounting for migration of households can understate the true effects of the program

Example: Does Energy Literacy translate into energy cost-savings?

Aaron Pinet

Data: Primary Data

Cross-sectional survey (25 questions) administered to National Grid customers (n=4306)

Methods: Multivariate regression models

1. **Higher energy literacy at the household level leads to energy cost-savings for that household.**
2. **Estimates can be used to impute to the national level. What if there was an investment in energy literacy programs to expand energy literacy among the population? How much would this reduced the demand for energy – and resulting cost-savings?**

Combining Secondary & Primary Data

- How did the EVALI outbreak in late 2019 affect consumer risk perceptions of e-cigarettes vs. combustible cigarettes – as the outbreak was unfolding?

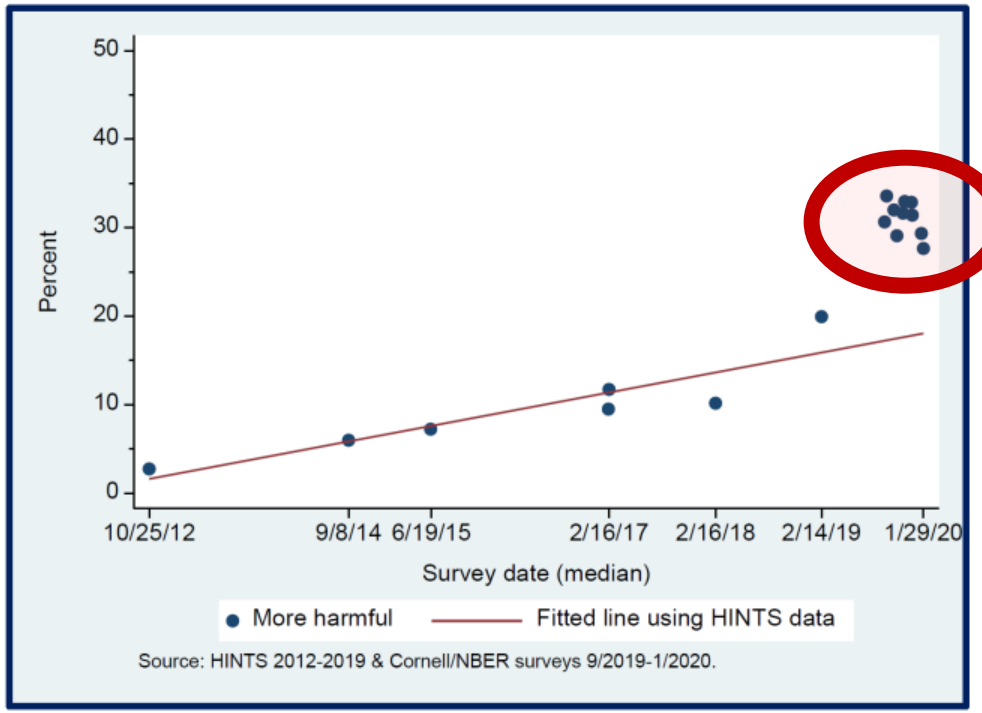
(Dave et al., Journal of Risk and Uncertainy 2020)

- Secondary Data: HINTS (collected by the FDA) in 2012, 2014, 2015, 2017, 2018, 2019

BUT these data only covered the situation BEFORE the outbreak.

- Primary Data: Fielded 10 bimonthly Google Surveys (September 2019 – January 2020)

Figure 5: Respondents who think e-cigarettes are more harmful compared to smoking cigarettes HINTS and GS data



Example: Economic Impact of Autonomous Vehicles

Cameron Wilson

- Combination of approaches and data ("Mixed methods")

- Reference to qualitative data (e.g. expert interviews)
- Projection of trends in transportation-related employment
- Time-series modeling

Takeaway

- Data and methods come in various shapes and sizes
- Not meant to be perfect
  - Every study has its strengths, weaknesses, and assumptions being made
  - Important to be transparent and recognize what the data/methods can and cannot answer
- Answers have a dose of uncertainty
  - Quantify that uncertainty and understand its sources.
- Data and methods matter!