

# The Rise of Administrative Data: Paths to Privacy and Public Good

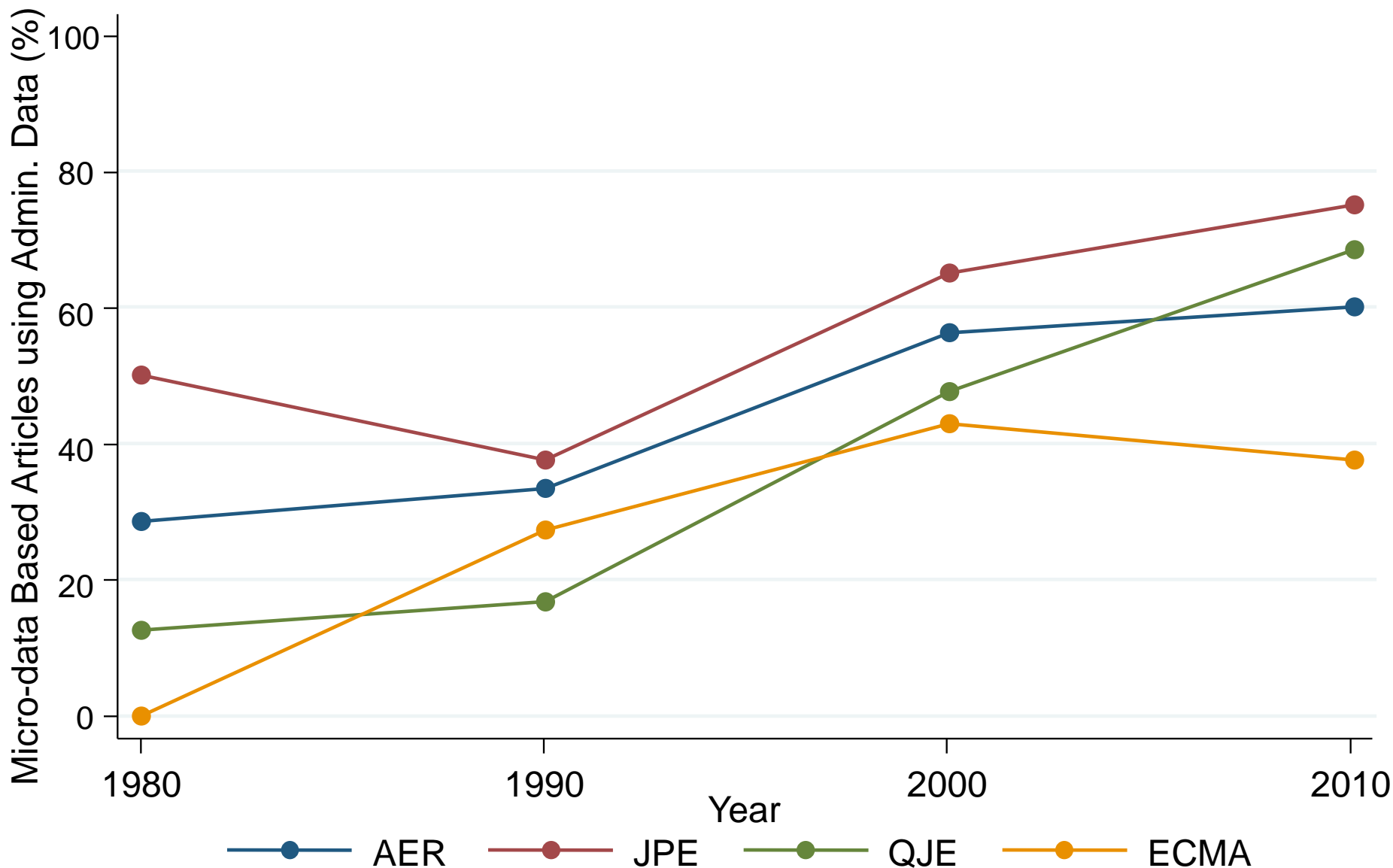
William F. Tate  
Washington University in St. Louis  
The Graduate School  
Council of Graduate Schools Annual Meeting  
Washington, DC

December 9, 2016

# Acknowledgements

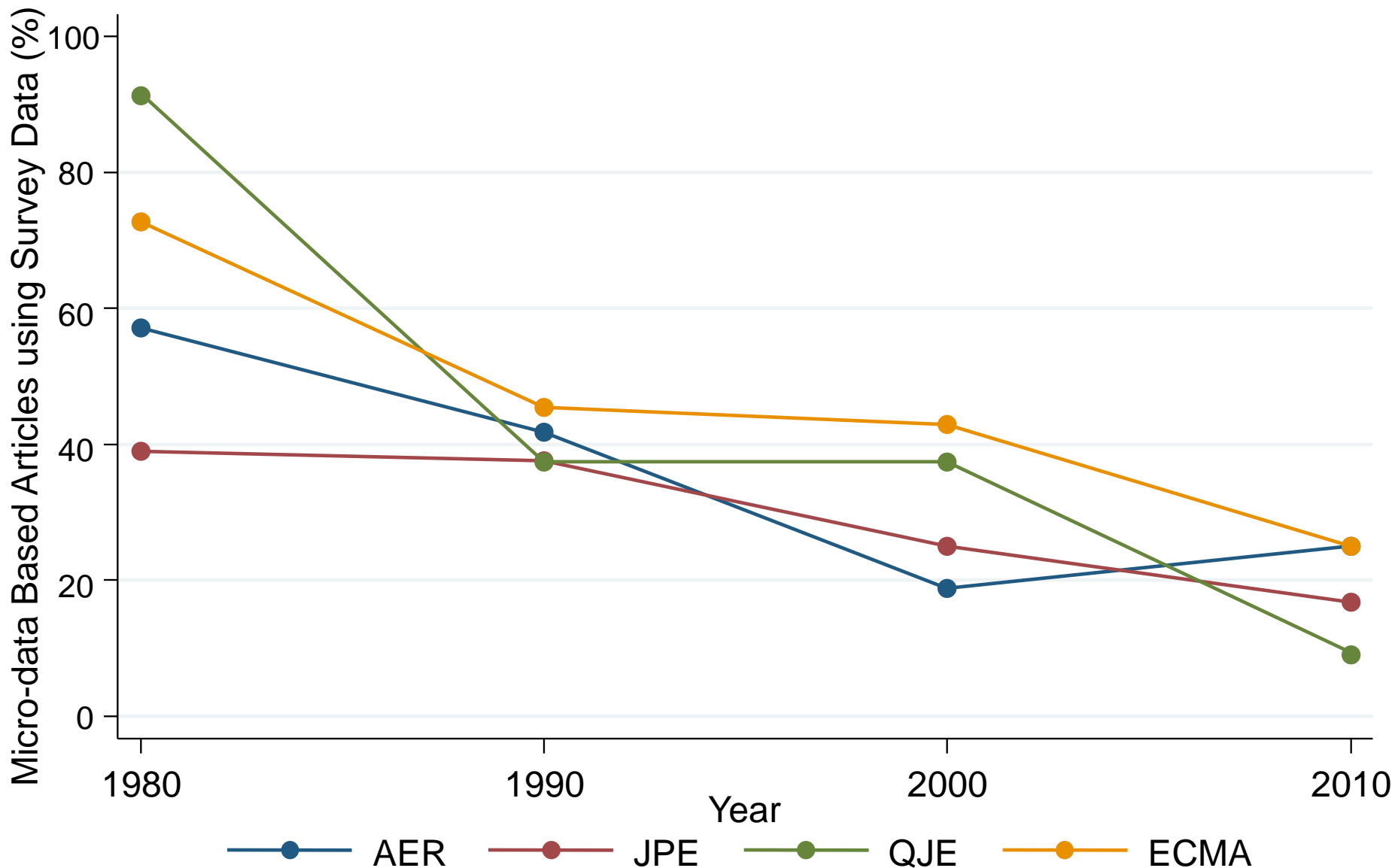
- Amy Berman, National Academy of Education
- John Friedman, Brown University, Department of Economics
- Andrew Ho, Harvard Graduate School of Education

# Use of Administrative Data in Publications in Leading Journals, 1980-2010



Note: "Administrative" datasets refer to any dataset that was collected without directly surveying individuals (e.g., scanner data, stock prices, school district records, social security records). Sample excludes studies whose primary data source is from developing countries.

# Use of Pre-Existing Survey Data in Publications in Leading Journals, 1980-2010



Note: "Pre-existing survey" datasets refer to micro surveys such as the CPS or SIPP and do not include surveys designed by researchers for their study. Sample excludes studies whose primary data source is from developing countries.

# Administrative Data in the United States

- Papers using administrative data increasing
- As demand for administrative data increases, European countries are experiencing growth as sites of investigations
  - ↑ Population registers in Sweden, Austria, Germany, Norway, and Denmark
  - U.S. starting to build this infrastructure, but remains fragmented
- Administrative data can help advance research in many fields and disciplines.
  - Education
  - Economics and Business
  - Social Work
  - Public Health

# Data Use Models: Education Data

- Data Use Model involves establishing MOU between a learning organization (school district, university, etc.) and individual researcher or institution, and a data transfer.
- Model: Scholar maintains (potentially identifiable) data on individual systems.
- Pros: Researcher can work on personal machines, highly flexible and efficient
- Cons: Costly to replicate secure environments for data

# Data-Use Models: Federal Statistical Research Data Centers

- Partnership between federal statistical agencies and research institutions. Secure facilities providing authorized access to restrict-use microdata for statistical purposes.
- Model: (1) Dedicated closed system, (2) statistics may only be removed after human review, and (3) data is deidentified with exception of data mergers with outside data
- Pros: Access is feasible with 24 centers around the country. Standardize process for access
- Cons: Statistical review is time consuming.

# Data-Use Model: Statistics Denmark

- Unified national statistical agency (think IRS, Census, HHS, NCES, BLS combined)
- Pros (advantages): Ability to work with aggregated data outside the closed system and it allows for use with a personal computer
- Model: Researchers VPN into Statistics Denmark Systems
  - Statistics are portable after computer review, with occasional human audits.
  - Data is deidentified, Statistics Denmark will merge outside data.
- Cons (disadvantages): Loss of U.S. centric analyses and related concerns about the translation of findings to domestic context.



# Going Forward

- Public Education: Explaining the benefits of research that uses administrative data, while also putting forth models that address privacy issues.
- Legislative Action: Murray, Ryan introduce Bill to expand data use in evaluating federal programs, tax expenditures