

# Cornell FSRDC Newsletter

*A report of activities of the Cornell Federal Statistical Research Data Center*

## Eye on Researchers: Sanders



Seth Sanders joined the Cornell department of economics this fall. Though he is new to the Cornell RDC community, he is no stranger to RDC research, having worked on several strands of significant research at the Triangle RDC.

“Workplace Concentration of Immigrants,” published in *Demography* in 2014, uses the Longitudinal Employer-Household Dynamics data linked to 2000 Decennial Census long form. He and his coauthors find that immigrants have more immigrant co-workers than non-immigrants. Since factors like industry, language, and residential segregation explain only some of the tendency to work with others from the same country, they argue that social networks or other unmeasured country-specific factors play a role.

More recently, “The Methuselah Effect: The Pernicious Impact of Unreported Deaths on Old-Age Mortality Estimates” demonstrates that imperfect matching between death records and survey data can have large impacts on mortality measurement. Published in *Demography* in 2017, this work utilizes the National Longitudinal Survey of Older Men linked to death records in the U.S. Vital Statistics and the Social Security Death Index.

## Data Update: Bureau of Economic Analysis

Researchers who apply to use BEA data under their special sworn researcher program can now access the data through the FSRDC system instead of traveling to BEA headquarters. BEA data covers foreign direct investment, multinational enterprises, and international trade in services. To access these data, submit a proposal through BEA, and let Nichole Szembrot ([nichole.e.szembrot@census.gov](mailto:nichole.e.szembrot@census.gov)) and Warren Brown ([wab4@cornell.edu](mailto:wab4@cornell.edu)) know that you would like to use the Cornell RDC. For more information about the available data and proposal process, visit <https://www.bea.gov/research/special-sworn-researcher-program>.



---

## New Active Projects

### **“Research and Development Laboratories in the Production Process”**

Kristy Buzard (Syracuse), Gerald Carlino (Philadelphia Fed), Robert Hunt (Philadelphia Fed), Tony Smith (University of Pennsylvania), and Anna Tranfaglia (Philadelphia Fed)

This research will benefit the U.S. Census Bureau by expanding the understanding of how research and development (R&D) investments contribute to the productivity of firms--both the firms that undertake the investments directly as well as neighboring firms that benefit from spillover effects. Specifically, combining a unique data set of geocoded R&D labs with firm-level data from Compustat and restricted-use Census data at the firm- and establishment-level, this project will first describe how firms locate their R&D labs relative to their own production and administrative establishments as well as the labs of other firms. It will then estimate the impact of R&D labs, the locations and characteristics of those labs and the clusters in which the labs are located, and the position of labs within the production structure of the firm on total firm productivity. The researchers will test whether these estimates vary by industry, firm size, or presence of multiple labs within the firm. In addition, the researchers will use their proprietary data on R&D clusters and the strength of knowledge spillovers within them to produce population estimates of the impact of being located within R&D clusters or near the labs of other firms in one's own industry.

### **“Economic Impact of Science and Engineering Workers” (Boston)**

Holden Diethorn (SUNY Albany) et al

The project will investigate the use of big data machine learning methods to improve Title 13, Chapter 5 data programs, and to generate estimates of the population of Research and Development (R&D) performing firms and establishments, and of Science, Technology, Engineering and Math (STEM) workers. The project will including analyses to improve R&D survey data, identify postdoc and graduate student survey respondents, to estimate the returns to education for STEM PhDs and postdocs, knowledge-flows from universities to industry, the determinants of STEM labor demand, R&D-productivity, STEM mobility, wages, and spillovers, regional economic impacts of federally funded science and the impact of technology on older workers. The project will use newly available big data machine learning techniques to explore methods to classify R&D performers in the Longitudinal Business Database (LBD) using Business Research and Development and Innovation Survey (BRDIS), Survey of Industrial Research and Development (SIRD), the Standard Statistical Establishment List (SSEL) Company Organization Survey (COS) and Compustat data as a training dataset. Using machine learning methods will be used to classify LBD firms based on common attributes available in universe files such as the LBD, Standard Statistical Establishment List (SSEL), Longitudinal Employer-Household Dynamics



(LEHD), Economic Census and USPTO patent data. The outcome will be a predicted probability of R&D classification. The project will also predict postdoc and graduate student status using data from the Census Innovation Measurement Initiative (IMI). The IMI UMETRICS data will be used as a training dataset to employ random forests and boosted trees machine learning routines to classify ACS and decennial respondents based on common attributes. The relationship of the training data outcome to linked ACS, decennial, LEHD EHF, ICF and numident predictive characteristics such as age, gender, country of origin, PhD education, industry and quarterly earnings is then projected onto the larger ACS-decennial-LEHD sample to form a predicted probability of being a postdoc and a graduate student that can be used to classify respondents when the true status is unknown.

**“The Link between Education Quality and Access and later life health and Mortality” (Internal)**  
Seth Sanders (Cornell) et al

The project will exploit local school quality measures and changes in State school policies from the 1920s and 1930s. We will link the 1940 Decennial Census to the NLMS so that we can estimate the effects of school quality on mortality and health outcomes at older ages. The 1940 data allows us to understand what the local school environment was for children 0-18 in 1940 and also allows us to observe the education and occupation of their parents as these children will overwhelmingly be living at home. This allows us to estimate the effects of school quality and school policies on late life health outcomes and to evaluate whether the impacts differ by parental levels of education and socioeconomic status.

## **Continuing Active Projects**

### **Cornell University:**

“Does Access to Health Care Affect Teen Birth Rates and School Dropout Rates? Evidence from School-based Health Centers”

Michael Lovenheim, Randall Reback (Barnard), and Leigh Wedenoja (Rockefeller Institute of Government; former Ph.D. student)

“Government Welfare and Private Provision of Public Goods: Evidence from SNAP”  
Nicolas Bottan

“Experimental Estimates of the Long-Run Impacts of Welfare Reform” (Internal)  
Amanda Eng, Pauline Leung, Zhuan Pei, Katherine Wen, et al

“Multigenerational Exposure to Neighborhood Disadvantage and its Effects on Well Being”  
Steven Alvarado and Alexandra Cooperstock

“The Heterogeneous Effects of Education on Health and Productivity” (Internal)  
Seth Sanders et al



- 
- “The Impact of Childhood Nutrition Assistance on Child Health and Well-Being: Lessons from WIC and the School Breakfast and National School Lunch Programs” (Berkeley)  
Barton Willage (Louisiana State; former Ph.D. student), Marianne Bitler (UC-Davis), Hilary Hoynes (UC-Berkeley), Krista Ruffini (UC-Berkeley), and Lisa Schulkind (UNC-Charlotte)
- “Race in Rural America: Differentials in Teenage Motherhood and High School Completion” (Triangle)  
Seth Sanders and Laurel Wheeler (Duke)
- “Developing and Validating New Statistics on Intergenerational Inequality, Poverty, Mobility, and Opportunity” (Internal)  
Laura Tach et al
- “Human Capital and Corporate Financing, Restructuring, and Governance”  
Hyunseob Kim, Warren Bailey, Joao Vitor Costa, Antonio Falato (Fed), Edith Liu (Fed), Song Ma (Yale), David Matsa (Northwestern), Brian Melzer (Dartmouth), and Till von Wachter (UCLA)
- “The Role of Educators on Reporting Child Maltreatment”  
Maria Fitzpatrick, Cassandra Benson (U.S. Air Force Academy; former Ph.D. student), Samuel Bondurant (Census), and Christopher Wildeman
- “Immigration Impacts, Immigrant Well-Being, and Geographic Context using the 1996-2008 Survey of Income and Program Participation”  
Matthew Hall and Emily Greenman (Census)
- “How Housing and Labor Market Conditions Influence the Progression of Romantic Relationships”  
Laura Tach, Sharon Sassler, Mariana Amorim (Washington State; former Ph.D. student), and Emily Parker
- “The Political Economy of Medicaid: Ideology, Eligibility, and the Consequences of Cost-saving Measures”  
Jee-Hun Choi and Claire S.H. Lim
- “The LBD Initiative - Collaboration with RDC Researchers” (Internal)  
Lars Vilhuber, Chih-Chung Lin, et al
- “Census Longitudinal Infrastructure Project - Linkage Development” (Internal)  
Seth Sanders et al
- “The Incidence of a Local Labor Demand Shock with One-Sided Migration: American Indian-Owned Casino Gaming” (Triangle)  
Seth Sanders, Martha Stinson (Census), and Laurel Wheeler (Duke)
- “Effects of Education on Health and Other Outcomes Using Variation in Social Security Benefits for Children”  
Barton Willage (Louisiana State University; former Ph.D. student)

“Estimating the Determinants of Private School Enrollment: The Role of School Attendance Zone Racial and Ethnic Composition”

Arnab Basu, Chia-Hua Lin, Salvatore Saporito (William & Mary), and Anna Choi (Pepperdine; former Ph.D. student)

“Effects of DI on Health and Well-being”

Sarah Prenovitz (Abt Associates; former Ph.D. student)

“SIPP 2014 Internal Research and Collaboration Project” (Internal)

Lars Vilhuber, Bill Block, Warren Brown et al

“Vietnam Veterans and Obesity”

Daniel Grossman (West Virginia University; former Ph.D. student) and John Cawley (adviser)

“LEHD Data Infrastructure Project” (Internal)

Lars Vilhuber, Seth Sanders, and Sylvérie Herbert et al

“Research Program to Determine the Possibilities from Linking ACS & LEHD Data” (Internal)

Lars Vilhuber et al

“Demographic Survey Improvement Project” (Internal)

Lars Vilhuber et al

### **Syracuse University:**

“SNAP and the Human Capital Investments of Young Adults”

Sarah Hamersma, Warren Brown (Cornell), Rhea Acuna, and Elmer Li (Cornell)

“The New York State Homestead Tax Option: Tax Incidence and Equity”

David Schwegman and John Yinger (adviser)

“The Relationship between SNAP and Mortality” (Kentucky)

Colleen Heflin, Samuel Ingram (Kentucky), Chaegyung Jun (Missouri), and James Ziliak (Kentucky)

“A Binational Perspective on Selective Migration among Korean Immigrants” (Missouri)

Colleen Heflin, Claire Altman (Missouri), and Chaegyung Jun (Missouri)

“Disparities in Healthcare Access and Utilization among Children with and without Special Healthcare Needs, and their Caregivers” (Missouri)

Colleen Heflin, Nancy Cheak-Zamora (Missouri), and Chinedum Ojinnaka (Arizona State)

“Knowledge of Health Status and the Timing of Retirement and Disability Claims”

Perry Singleton

“Education, Geography, and U.S. Adult Mortality Risk”

Jennifer Montez, Blakelee Kemp, and Julene Cooney

“The Effects of Pregnancy-Related Medicaid Expansions on Maternal, Infant, and Child Health”

Sarah Hamersma and Melanie Guldi (University of Central Florida)



---

“Occupational Variation in Health Care Coverage” (Wisconsin)

Colleen Heflin, Joan Hermsen (Missouri), and Leslie Hodges (Wisconsin)

“A Second Look at the Health Effects of Military Service Using the Vietnam-Era Draft Lottery as a Potentially Invalid Instrumental Variable”

Alfonso Flores-Lagunes and Xintong Wang (Slippery Rock University of Pennsylvania)

### **University at Albany:**

“Productivity and Wage Dispersion in the Great Recession and Beyond” (Boston)

Holden Diethorn, Erling Barth (Institute for Social Research), Wang Jin (MIT), Sari Kerr (Wellesley), Kristina Steffenson McElheran (MIT), Namrata Narain (Harvard), Andrew Wang (Harvard), Xiupeng Wang (MIT)

### **Contacts:**

Nichole Szembrot, Cornell RDC  
Administrator  
[nichole.e.szembrot@census.gov](mailto:nichole.e.szembrot@census.gov)  
607-255-8603

William Block, Executive Director  
[block@cornell.edu](mailto:block@cornell.edu)  
Warren Brown, Research Director  
[wab4@cornell.edu](mailto:wab4@cornell.edu)

