Disability Benefit Receipt by State and County, 1970-2017: Description and Dataset

Timothy Moore (Purdue University and NBER)

Program and Focal Areas: The program areas are *Social Security Disability Insurance (DI)*, *Supplemental Security Income (SSI)* and *Disability Support (DI and SSI)*. The focal area directly addressed is *Disability Incidence in the Population*, while the new dataset could be used to examine other focal areas (e.g., *Program Interactions with the Affordable Care Act* and *State and Local Government Pensions*).

Keywords: DI; SSI; disability incidence; geographic differences in disability receipt; state; county.

Abstract: In this project, a dataset will be created with the rate of SSDI beneficiaries and SSI recipients in the working-age population for each county and year from 1970 to 2017. DI/SSI payment information will also be included. An article will be written to showcase the dataset and discuss key insights and puzzles that arise from looking at disability incidence across counties over a long period of time. The dataset and a codebook will be made freely available (e.g., via the SSA and/or NBER websites). The long-term payoff will be the creation of a detailed, publicly available data asset that can be used to examine a wide variety of disability policy questions.

Research using geographic variation in the receipt of federal disability benefits has been hampered by the lack of readily available longitudinal data. SSA publishes *OASDI Beneficiaries by State and County* and *SSI Recipients by State and County* each year, but these are underutilized because they are in paper form until 1998 and subsequently as online state-specific PDF/Excel files. I have collected and digitized these publications back to 1970, and will use the output to create a dataset with ~150,000 county-year observations spanning 48 years. Disability policy researchers will be able to use it for standalone analysis, merge it with other socioeconomic data, or use it to explore the potential of a topic before seeking more detailed administrative data.

The project will deliver a dataset, a detailed codebook and an overview article. To do this:

- Data will be checked for accuracy, and variables and county codes made consistent over time;

- Detailed population data compiled by the Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute will be merged in and disability benefit rates calculated;

- A codebook will be written explaining the structure of the dataset and describing the variables;

- Data will be analyzed to examine characteristics such as: the counties/regions with the highest/lowest DI, SSI, and disability benefit rates in different periods (e.g., 1970s, 1990s, 2010s); the persistence of disability benefit rates within counties; the correlation between DI and SSI incidence rates within counties; and the regions and characteristics of counties with high/low disability benefit rates;

- An article will be written describing the most important and interesting facts that come out of the analysis. It will also describe the dataset. I publish it as a NBER working paper and in a widely accessible policy journal.

Significance: There is substantial geographic variation in disability benefit receipt. For example, between 1993 and 2009, the fraction on DI in the state with the highest rate (Mississippi) was roughly three times larger than the state with the lowest rate (Utah) (Coe et al., 2011). The variation is even greater at the county level, with rural counties in the South and Appalachia region having DI and SSI rates in 1990 that were several times higher than the national average (McCoy, Davis and Hudson, 1994). These are among a small number of papers that examine geographic variation in disability benefit receipt, mostly at the state level (e.g., Strand, 2002; McVicar, 2006).

It is important to examine this variation in more detail. States have been the most commonly used geographical unit. Yet there is enormous variation within states that is missed in such analysis; for example, Virginia has counties with some of the highest as well as the lowest rates of DI receipt in the United States. McCoy, Davis and Hudson (1994) examine county-level differences, but focus only on 1990. Easily accessible county-level data over a long period of time will allow more research to understand geographic variation.

Such data also allow the empirical investigation of disability policy issues with state or county variation over time. There are some examples of how county-level data has allowed researchers to do this. For example, Black, Daniel and Sanders (2002) examine how economic conditions related to the coal boom and bust affected disability receipt; Autor, Dorn and Hanson (2013) examine how Chinese import competition affects local labor markets and DI participation; and Foote, Grosz and Stevens (2018) examine how mass layoffs affect DI applications and allowances. A broader, more accessible data series will stimulate more research of this type.

Methods: Since the 1950s, extracts of the Master Beneficiary Record have been used to produce a snapshot of DI beneficiaries in each state and county in December of each year. This was called *Social Security Beneficiaries by State and County* until 1985, and has since been called *OASDI Beneficiaries by State and County*. A similar publication, *SSI Recipients by State and County*, has been created for SSI recipients using extracts from the Supplemental Security Record since SSI started in 1974. Online versions are available from 1999 for the OASDI publication and 1998 for the SSI publication.

I started locating copies of these publications about ten years ago. With the help of SSA library and publications staff, I was able to find all publications from 1970 except for *Social Security Beneficiaries by State and County, 1981*. I copied them and have used research funds to contract with Digital Divide Data (DDD) to key the data into Excel files. (DDD has been used by other NBER researchers, and guarantees 99.95% accuracy by having two separate operators key the data while a third conducts quality control.) My original intention was to use these data for a range of DI/SSI research projects, but other research has taken priority. This project will enable broad access to the data, building on the foresight and efforts of SSA staff to produce consistent statistical publications over long periods of time.

In order to make the data accurate and easy to use, and to complete other aspects of the project, I will:

- Create consistent variables over time, including creating data flags that make it easy to understand the different confidentiality protection rules applied in each publication;

- Create counties with consistent geographic borders from 1970 to 2017 and attach FIPS codes that will allow linking to other data sources;

- Merge in annual intercensal population estimates from the Census Bureau (compiled by SEER);

- Calculate disability benefit receipt rates for DI, SSI and both together (adjusting for overlap);

- Develop a codebook describing the dataset. This will be circulated among potential users to ensure it is comprehensive and understandable, and refined in response to feedback. Assistance and feedback

will also be sought from NBER staff who curate and document data. The document will acknowledge the underlying data sources as being produced by SSA and SSA staff;

- Examine characteristics such as:

  - The counties with the highest and lowest disability benefit rates in different periods;

  - The persistence of disability benefit rates within counties;

  - The correlation between DI and SSI incidence rates;

  - The characteristics of counties with high/low disability benefit rates.

  In order to make the examination accessible, the main methods used in the article will be to calculate and report summary statistics and to assess the statistical significance of differences over time, across regions, etc.

- Write an article describing the most important and interesting facts that come out of the analysis. It will also describe the dataset.

- Make the dataset, codebook and article available to SSA, and work with SSA to make sure all are in formats that can be easily accessed and used by other researchers.

Statement of Complementarity or Overlap: There is overlap with "Understanding the Geographic Variation in Social Security Disability Insurance," a current DRC project being undertaken by Moore with Amanda Michaud and David Wiczer. In that project, we are using SSA administrative data from 1996 to describe the county-level geographic concentration of DI, how it has evolved over time, and its correlation with socioeconomic factors that changed over time. This project complements that project by producing a more accessible research article and providing data that others can use to examine this topic. These projects are part of a long-term agenda in understanding geographic variation in Social Security programs. In the future, I hope to produce similar datasets on seniors (Retirement/Survivors Insurance and Aged SSI) and children (Social Security dependents and SSI for Children).

References

Autor, David, David Dorn and Gordon Hanson. 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review*, 103(6): 2121–2168.

Black, Dan, Kermit Daniel and Seth Sanders. "The Impact of Economic Conditions on Participation in Disability Programs: Evidence from the Coal Boom and Bust." *American Economic Review*, 92(1): 27-50.

Coe, Norma, Kelly Haverstick, Alicia Munnell, and Anthony Webb. 2011. "What Explains State Variation in SSDI Application Rates?" *Center for Retirement Research at Boston College Technical Report 2011-23*.

Foote, Andrew, Michel Grosz and Ann Huff Stevens. 2018. "Locate Your Nearest Exit: Mass Layoffs and Local Labor Market Response." *ILR Review*, forthcoming.

McCoy, John L., Miles Davis, and Russell E. Hudson. 1994. "Geographic Patterns of Disability in the United States." *Social Security Bulletin*, 57(1): 25–36.

McVicar Duncan. 2006. "Why do disability benefit rolls vary between regions? A review of the evidence from the USA and the UK." *Regional Studies*, 40: 519–533.

Strand, Alexander. 2002. "Social Security Disability Programs: Assessing the Variation in Allowance Rates." *SSA Office of Research Evaluation and Statistics Working Paper No. 98*.

***Deliverables and Timeline*:**

Timeline	Deliverable
September 30, 2019	Electronic dataset of DI, SSI adult disability numbers and rates, 1970-2017
September 30, 2019	Codebook for dataset
September 30, 2019	One final paper
September 30, 2019	One non-technical summary report