

Linking City, County, State, and Federal datasets

Robert M. Goerge, Ph.D.

**WORKSHOP ON STATE AND LOCAL GOVERNMENTS USE OF  
ALTERNATIVE AND MULTIPLE DATA SOURCES**

June 1, 2016

The National Academy of Sciences, Medicine and Engineering  
Washington, DC

**ChapinHall** at the University of Chicago  
Policy research that benefits children, families, and their communities

# Approach

---

- Began with a database development effort within Illinois where the primary study population is Chicago Public School students
- Expanding it to use Federal data of out necessity
- The future is expanding to new study population—other places with k-12 data
- **A successful linkage project is 99% about accessing the “right” data**

# Collaboration and Funding

---

- Derek Neal, Department of Economics
- UChicago Consortium for Chicago School Research - Elaine Allensworth
- Chicago Public Schools
- Milgrom Fund
- Spencer Foundation

# Initial research questions

---

What are the empirical relationships between the characteristics of families, neighborhoods, and schools that Chicago children experience and the subsequent outcomes they experience as adults?

- Given the lack of significant improvement in high school graduation rates over much of the last three decades and the fact that unemployment, welfare program use, and imprisonment are heavily concentrated among those who do not finish high school, we will devote special attention to predictors of **high-school completion**.
- Given the dramatic growth in U.S. Prison populations during the last 35 years, we are particularly interested in understanding what measures of family environment, neighborhoods, and school experience predict **arrests and subsequent incarceration in young adulthood**.
- Given the sharp declines in male employment rates among less-educated men in recent decades, we seek to understand both the predictors of adolescent and adult employment and also how involvement in the criminal justice systems affects **transitions into stable employment**.

# Goals

---

- Make data available to other researchers (and ourselves) in a format that is more conducive to social science research so that more research and information can be produced to build actionable knowledge
- To improve the K-12 and adult outcomes of children in Chicago.

# Priorities

---

- Transparency
  - Document the data quality (linkage)
  - Document the decisions made with raw data
- Make data available to other researchers in ways that are conducive to their research
- Develop a sustainable model
- Extend the data
- Build on existing public sector relationships by ensuring that they benefit from any research

# Events

---

- high school graduation
- Post-secondary attendance
- Post-secondary persistence (attendance after the first year of enrollment)
- Post-secondary graduation
- transition to stable employment
- arrest
- juvenile and adult incarceration

# Dataset to be linked at individual-level

---

- ***Chicago Public Schools Master Database:*** The database tracks all children enrolled in CPS. Chapin Hall's data holdings currently cover the **1991-2015 school years**. The records contain information on **residential census tract, school assignment and mobility, days absent from school, disciplinary events in school, grades, test scores, grade progression, and graduation outcomes** for all CPS students. We are able to track students over time and create a complete history of their schooling experiences. We also have data on where students go when they leave CPS prior to graduation. This data accounts for over **1.2 millions students**.

# Data (cont'd)

---

- ***Illinois Longitudinal Public Assistance Research Database (ILPARD)***: ILPARD is contains of program eligibility and participation of **over 13 million individuals** in the SNAP (Food Stamps), TANF and Medicaid programs in Illinois from 1989 until the present.
- ***Unemployment Insurance Wage Data***: Chapin Hall obtains this data from extracts of the Illinois Department of Employment Security's (IDES) **using the SSN as the link field**. The data contain quarterly wage records that firms report to IDES.

# Data (cont'd)

---

- ***National Student Clearinghouse (NSC)***: This database provides enrollment history and degree attainment data for current and former students from more than 3,400 colleges and universities that enroll over 96 percent of all students in public and private U.S. institutions.
- ***Cook County Court Clerk System Delinquency File***: The Cook County Juvenile Court, Juvenile Justice Division data contain information on juvenile delinquency cases from the 1981 through 2014.
- ***Chicago Police Department (CPD)***: CPD data provide records of all arrest events in Chicago over the years 1999 to 2015.
- ***Illinois Departments of Corrections (IDOC)***: All incarcerated individuals from 1990-present.

# Roadblock

---

- Roadblock is a temporary installation for managing traffic
- CPS was unable to secure NSC and UI wage data for us
  - Their focus was on graduates
  - They lacked the technical capacity to work with NSC and Illinois Department of Economic Security to obtain the data that was required
  - They simply lack the staff time as they are currently in a crisis management mode
- All other datasets are in hand

# 2005 U.S. Census Bureau data project

---

- In 2005, TANF, Food Stamps, and child care subsidy administrative data was transferred to Census from Illinois, Maryland, Minnesota and Texas
- Linked to Census 2001 Supplementary Survey (SS01)—the first ACS
- Linked to UI wage data in possession of Census Bureau through LEHD project
- Used to understand the impact of child care subsidy on maternal employment and comparison between self-report and administrative report of Food Stamps

# Data available now (or possibly\*) available at Census (CARRA)

---

- Illinois UI wage data
- Illinois SNAP data
- Illinois prison data
- National Student Clearinghouse \*
- Census survey data (ACS)
- Federal IRS data
- NUMIDENT file
- Other UI wage data that could be used for research
- National Mortality data \*

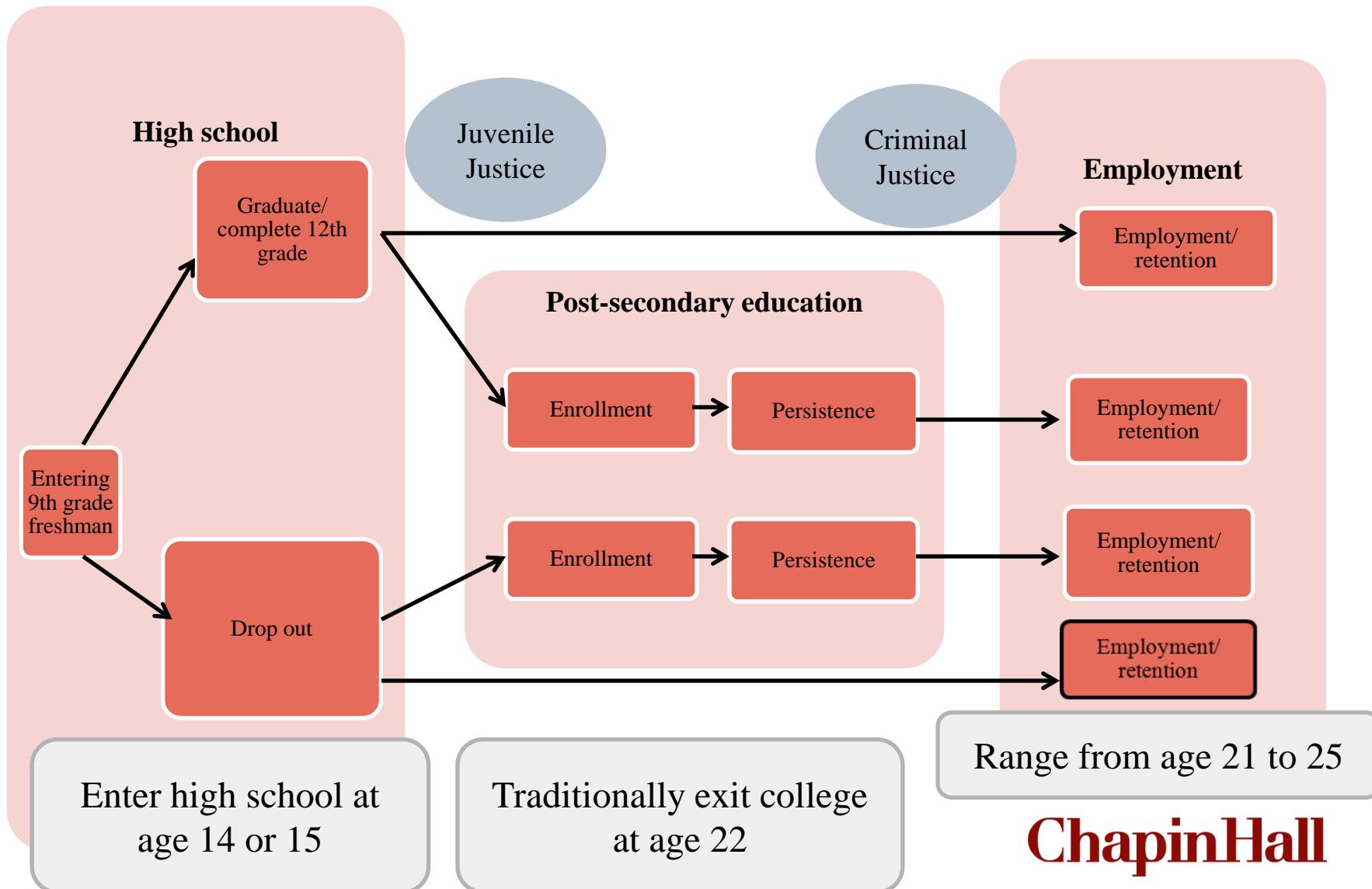
# Benefits of linking to federal data

---

- Outcomes go beyond state borders
- Ability of Census Bureau to attach SSN to data without SSNs and conduct better record-linkage
- Secure storage and access
- Sustainability

# Education to Employment Pathway

Fall 00/01 ----- Spring 04/05----- 2009/2010



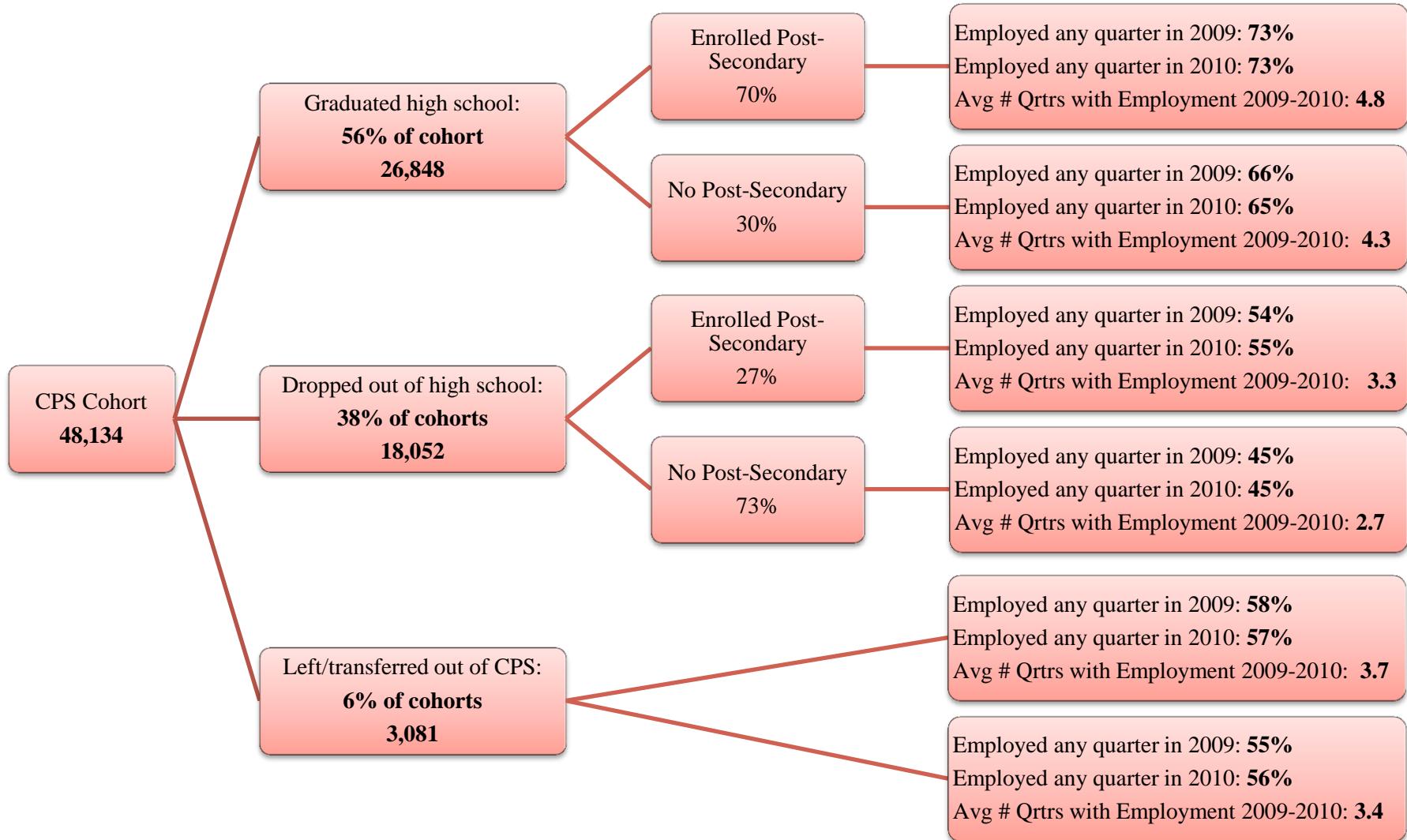
# Sample

---

- Initial sample
  - First-time 9<sup>th</sup> grade students in Chicago Public Schools in 2000-01 and 2001-02 cohorts: *58,810*
- Given data limitations, sample was limited to:
  - Students with valid SSNs (for data linking): *53,899*
  - Students confirmed to be living in Illinois during 2009-10: *48,134*
- Overview of study sample: *48,134*
  - Students who graduated from high school: *26,848 (55.8%)\**
  - Students who ever enrolled in post-secondary education: *23,807 (49.5%)*
  - Students who had wages any time in 2009 or 2010: *33,483 (69.6%)*

# Employment in 2009-2010 was Highest for CPS Graduates Who Enrolled in College

---



# Summing up

---

- Getting the right data to the right place to do the analysis you need to do
- Understanding the limitations
- Technical aspects are not that complicated, but
  - Still need eyes on the data to make sure you’re linked it correctly (even machine learning requires human supervision)
- There is still no divine right to data, but the hope is that the Commission will address this.