Opportunities for criminal justice research using restricted data

Ling Wu, Ph.D.
Research scientist
Criminal justice data and spatial analyses
Texas Federal Statistical Research Data Center (TXRDC)
Texas A&M University, College Station
3/25/2022
Content

• Federal Statistics Research Data Center (RDC)
• Criminal Justice Administrative Records System (CJARS)
• National Crime Victimization Survey (NCVS)
• Other restricted data for criminal justice research
• Summary
Federal Statistics Research Data Center (RDC)

- RDCs are research facilities and network based on collaboration among Census Bureau, other federal statistical agencies, and universities.
- 32 RDC locations sited at leading research institutions.
- 70% R1 universities host an RDC or have a formal affiliation with an RDC location.
- Texas Federal Statistical Research Data Center (TXRDC)
Advantages of TXRDC to researchers

➢ Access to restricted use data
  • Access to restricted microdata in many disciplines (*Business, Economics, Management, Health, Demography, Geography, Sociology, Urban Planning, Criminal Justice*)
  • Key variables not in public versions of datasets, or detailed versions of variables
  • Larger data sets (full population counts and/or larger samples)
  • Less restrictive top-coding
  • Ability to link data

➢ Secure computing lab

➢ Proposal development and administrative assistance

➢ Beneficial to long-term research plans

➢ Research and facility network

➢ Recent developments: More data available in RDCs and shrinking data in the public domain
Criminal Justice Administrative Records System (CJARS)

• Nationally integrated repository of data following individuals through the criminal justice system (2016-now)
• CJARS collects various administrative data held by a wide range of federal, state, and local CJ agencies.
• Data holdings: 36 million unique individuals’ 175 million CJ events (4 million+ events per year)
• 1st research-ready data released in 2020Q4 vintage
• CJARS is solely distributed to researchers on approved projects through RDC network.
CJARS data holdings

- Procedural coverage (arrest, court, probation, prison, parole)
- Temporal coverage
- Geographical coverage
CJARS coverage (Texas)

- CJARS collects state-level data from Texas Department of Criminal Justice (TDCJ), Texas Department of Public Safety (TxDPS), and iDocket.

- CJARS collects local-level data from County Clerk’s Office and County District Clerk’s Office in 69 counties (27% of all counties in Texas), 6 county sheriff’s offices, and one municipal police department.
**CJARS data linkage**

**Data linkage within CJ system**
- roster table at individual-level
- five relational tables at event-level

**cjars_id**
- person-level linkage across agencies within a jurisdiction
- person-level linkage across jurisdictions

<table>
<thead>
<tr>
<th>CJ Event Variable</th>
<th>Scale</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Offense type</td>
<td>person; property; drug; DUI; public order; other</td>
<td>arrest table</td>
</tr>
<tr>
<td>Statutory severity</td>
<td>felony-level; misdemeanor-level charge</td>
<td>adjudication table</td>
</tr>
<tr>
<td>Charging characteristics</td>
<td>the number of charges pressed</td>
<td>linking roster, arrest, and adjudication tables</td>
</tr>
<tr>
<td>Initial plea</td>
<td>plea of guilty; plea of not guilty</td>
<td>adjudication table</td>
</tr>
<tr>
<td>Charge alteration</td>
<td>charge increase; decrease; no change</td>
<td>linking CJARS arrest and adjudication tables</td>
</tr>
<tr>
<td>Case dismissal</td>
<td>dismissal; non-dismissal</td>
<td>adjudication table</td>
</tr>
<tr>
<td>Adjudication disposition</td>
<td>conviction; non-conviction</td>
<td>adjudication table</td>
</tr>
<tr>
<td>Incarceration sentence</td>
<td>custodial sentence; community sentence</td>
<td>adjudication table</td>
</tr>
<tr>
<td>Level of CJ contact</td>
<td>arrest-only; community sentence; incarceration</td>
<td>linking arrest and adjudication tables</td>
</tr>
<tr>
<td>Criminal history</td>
<td>prior arrest; prior incarceration</td>
<td>linking CJARS in different time periods</td>
</tr>
</tbody>
</table>
**CJARS data linkage**

- The anonymized CJARS roster contains both the `cjars_id` and the Protected Identification Keys (PIK) for each CJ-involved individual.

<table>
<thead>
<tr>
<th>PIK allows linkage to Census data</th>
<th>cjohn_id</th>
<th>cjohn_id allows linkage to CJARS tables</th>
</tr>
</thead>
<tbody>
<tr>
<td>555555555</td>
<td>000001</td>
<td></td>
</tr>
<tr>
<td>555555555</td>
<td>000001</td>
<td></td>
</tr>
<tr>
<td>777777777</td>
<td>000002</td>
<td></td>
</tr>
<tr>
<td>777777777</td>
<td>000002</td>
<td></td>
</tr>
</tbody>
</table>

- Link CJARS and **other social, economic, and demographic data** in RDCs at person-level through PIKs.

**Restricted Survey data:** Decennial Census (DC), American Community Survey (ACS), American Housing Survey (AHS), Rental Housing Finance Survey (RHFS), Current Population Survey (CPS), National Crime Victimization Survey (NCVS), National Longitudinal Mortality Survey (NLMS), National Longitudinal Survey (NLS), Survey of Income and Program Participation (SIPP)

**Restricted Administrative data:** Longitudinal Employer-Household Dynamics (LEHD), public program data (Medicaid and Medicare enrollment, Social Security programs, HUD assistance, SNAP, TANF, WIC)
Example 1: Neighborhood environment and criminal justice involvement

- Neighborhood correlates of criminal justice involvement
- Trajectories of neighborhood attainment of CJ population
**CJARS data linkage**

**Example 2: Labor market outcomes after a criminal justice intervention**

- Measure the employment and earnings of justice-involved individuals

CJARS data harmonization

- CJARS variable harmonization
- All *dates* in 5 relational tables are standardized

<table>
<thead>
<tr>
<th>Table</th>
<th>Standardized variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arrest &amp; booking</td>
<td>arrest offense classification</td>
</tr>
<tr>
<td>Adjudication</td>
<td>offense grade; legal code; charge offense classification; disposition description; disposition offense classification; sentence details</td>
</tr>
<tr>
<td>Probation</td>
<td>conditions; end status</td>
</tr>
<tr>
<td>Incarceration</td>
<td>facility type, entry status, exit status</td>
</tr>
<tr>
<td>Parole</td>
<td>end status</td>
</tr>
</tbody>
</table>

### Offense classification scheme

<table>
<thead>
<tr>
<th>Offense type</th>
<th>Offense code</th>
<th>Offense code description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Violent</td>
<td>1010</td>
<td>Murder</td>
</tr>
<tr>
<td>Violent</td>
<td>1011</td>
<td>Attempted murder</td>
</tr>
<tr>
<td>Violent</td>
<td>1012</td>
<td>Conspiracy to commit murder</td>
</tr>
<tr>
<td>Violent</td>
<td>1020</td>
<td>Unspecified homicide</td>
</tr>
<tr>
<td>Violent</td>
<td>1021</td>
<td>Unspecified homicide, attempted</td>
</tr>
<tr>
<td>Violent</td>
<td>1022</td>
<td>Unspecified homicide, conspiracy</td>
</tr>
<tr>
<td>Violent</td>
<td>1030</td>
<td>Voluntary manslaughter</td>
</tr>
<tr>
<td>Violent</td>
<td>1031</td>
<td>Voluntary manslaughter, attempted</td>
</tr>
<tr>
<td>Violent</td>
<td>1032</td>
<td>Voluntary manslaughter, conspiracy</td>
</tr>
<tr>
<td>Violent</td>
<td>1040</td>
<td>Vehicular manslaughter</td>
</tr>
<tr>
<td>Violent</td>
<td>1041</td>
<td>Vehicular manslaughter, attempted</td>
</tr>
<tr>
<td>Violent</td>
<td>1042</td>
<td>Vehicular manslaughter, conspiracy</td>
</tr>
<tr>
<td>Violent</td>
<td>1050</td>
<td>Involuntary manslaughter</td>
</tr>
<tr>
<td>Violent</td>
<td>1051</td>
<td>Involuntary manslaughter, attempted</td>
</tr>
<tr>
<td>Violent</td>
<td>1052</td>
<td>Involuntary manslaughter, conspiracy</td>
</tr>
<tr>
<td>Violent</td>
<td>1060</td>
<td>Kidnapping</td>
</tr>
<tr>
<td>Violent</td>
<td>1061</td>
<td>Kidnapping, attempted</td>
</tr>
<tr>
<td>Violent</td>
<td>1062</td>
<td>Kidnapping, conspiracy</td>
</tr>
<tr>
<td>Violent</td>
<td>1070</td>
<td>Rape</td>
</tr>
<tr>
<td>Violent</td>
<td>1071</td>
<td>Rape, attempted</td>
</tr>
<tr>
<td>Violent</td>
<td>1072</td>
<td>Rape, conspiracy</td>
</tr>
</tbody>
</table>
National Crime Victimization Survey (NCVS PUF/RUF)

➢ National-level estimates of personal and household criminal victimization
➢ 240,000 interviews involving 160,000 unique persons in 95,000 households each year.
➢ NCVS collects information on characteristics of the offender and crime, and victim reporting and experiences with the CJ system.

➢ NCVS Supplements (PUF/RUF):
   • Identity Theft Supplement (ITS)
   • Police Public Contact Survey (PPCS)
   • Supplemental Victimization Survey (SVS)
   • School Crime Supplement (SCS)
   • Supplemental Fraud Survey (SFS)
### National Crime Victimization Survey (RUF vs. PUF)

<table>
<thead>
<tr>
<th>NCVS (RUF) metadata: Scope and coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Collection time period</strong></td>
</tr>
<tr>
<td><strong>Spatial coverage</strong></td>
</tr>
<tr>
<td><strong>Universe</strong></td>
</tr>
<tr>
<td><strong>Unit of observation</strong></td>
</tr>
<tr>
<td><strong>Access modality</strong></td>
</tr>
</tbody>
</table>

- **NCVS RUF geographic variable:**
  - Smaller geographic identifiers (census tract, block group)
National Crime Victimization Survey (RUF)

- **Data linkage:**
  - Link NCVS and other data at finer geographic levels through geocode
  - Link NCVS and other data at person-level through PIK
**National Crime Victimization Survey (RUF)**


<table>
<thead>
<tr>
<th>Category</th>
<th>Variable</th>
<th>Data source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent variable</td>
<td>Violence reporting</td>
<td>NCVS</td>
</tr>
<tr>
<td>Independent variable</td>
<td>Neighborhood immigrant concentration</td>
<td>Decennial Censuses; ACS (linked by census tract identifiers in NCVS RUF)</td>
</tr>
<tr>
<td>Control variable</td>
<td>Victim characteristics (age, sex, marital status, education, income, homeownership)</td>
<td>NCVS</td>
</tr>
<tr>
<td></td>
<td>Crime characteristics (type, status, seriousness, victim–offender relationship)</td>
<td>NCVS</td>
</tr>
<tr>
<td></td>
<td>Neighborhood concentrated disadvantage (SES, racial composition, residential instability)</td>
<td>Decennial Censuses; ACS (linked by census tract identifiers in NCVS RUF)</td>
</tr>
<tr>
<td></td>
<td>County police force size</td>
<td>Law Enforcement Agency Roster (linked by county identifiers in NCVS RUF)</td>
</tr>
</tbody>
</table>
# NCVS vs. CJARS

<table>
<thead>
<tr>
<th></th>
<th>NCVS</th>
<th>CJARS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collection method</td>
<td>survey</td>
<td>administrative</td>
</tr>
<tr>
<td>Information</td>
<td>personal and household victimization</td>
<td>offenders' involvement in cj process</td>
</tr>
<tr>
<td>Crime type</td>
<td>7</td>
<td>all</td>
</tr>
<tr>
<td>Temporal coverage</td>
<td>1973-now</td>
<td>1975-now</td>
</tr>
<tr>
<td>Geographical setting</td>
<td>national</td>
<td>national, state, county</td>
</tr>
<tr>
<td>Smallest geographic unit</td>
<td>block group</td>
<td>residential address</td>
</tr>
<tr>
<td>PUF available</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Data linkage</td>
<td>link to other data at neighborhood-level</td>
<td>link to other data at person-level</td>
</tr>
<tr>
<td>Linkage key</td>
<td>geo-identifier; PIK (in certain year)</td>
<td>cjars_id; PIK</td>
</tr>
</tbody>
</table>
Other opportunities for CJ research using restricted data

➢ National Longitudinal Surveys of Youth (NLSY79, NLSY97)
  • Measure: crime, incarceration, & substance use
  • Examine smoking behaviors among younger women

➢ American Housing Survey (AHS)
  • Measure: perception on neighborhood crime
  • Examine relationship between residential instability and perceived and actual crime rates in neighborhoods

➢ Federal Justice Statistic Program (FJSP) PUF/RUF
  • Federal criminal case processing in arrest, prosecution, pretrial release, adjudication, sentencing, appeals, and corrections
Other opportunities for CJ research using restricted data

RDC project example: Criminal Offenders in the Labor Market (2019- )
• Impacts of CJ involvement (incarceration, arrest, criminal charging) on labor market outcomes (employment, earnings, firm characteristics).

• Ten restricted datasets utilized:
  American Community Survey (ACS)
  Decennial Census (CEN)
  Current Population Survey (CPS)
  Current Population Survey (March/ASEC)
  LEHD Employer Characteristics File (LEHD-ECF)
  LEHD Employment History File (LEHD-EHF)
  LEHD Individual Characteristics File (LEHD-ICF)
  LEHD Quarterly Workforce Indicators (QWI)
  LEHD Unit-to-Worker (LEHD-U2W)
  State–Education (STATE_EDU)
Summary: Popularity of crime and justice research

ICPSR’s 50 Most Popular Search Terms

1. Survey Data
2. Covid
3. Education
4. Race
5. Mental Health
6. **Crime**
7. MIDUS
8. Depression
9. Health
10. Income
11. India Human Development
12. Gender
13. Age
14. Add Health
15. China
16. Social Media
17. General Social Survey
18. Police
19. Sex
20. Immigration
21. Suicide
22. Alcohol
23. Monitoring the Future
24. Religion
25. **Domestic Violence**
26. Stress
27. CPES
28. Anxiety
29. Advanced Cognitive Training
30. Diabetes
31. **Recidivism**
32. Census
33. Happiness
34. American Community Survey
35. **Prison**
36. State
37. Trauma
38. **Violence**
39. Cancer
40. **Criminal Justice**
41. Outlook on Life
42. Discrimination
43. Racism
44. **LEMAS**
45. Sleep
46. Transgender
47. **NIBRS**
48. Marijuana
49. Poverty
50. Parent
Summary: Where to start?

• Metadata for active and completed projects

All metadata for active and completed projects as of February 17, 2022.

- Project title, years, RDC, status, PI, abstract, datasets

RDC new projects

- Annual projects by year
- Cumulative projects (N=758)
Summary: Where to start?

Find and apply to use restricted data from the Federal Statistical System

- 226 datasets can be accessed through RDCs
Summary: Where to start?

➢ Relevant documentation and literature:


• Publications utilizing restricted NCVS

➢ Funding opportunities:

• Illegal immigration, immigration enforcement policies, and American citizens’ victimization risk. NIJ (2020-2022, $673,685)

• New directions in research on immigration and crime. NIJ (2019-2021, $713,114)


• Ethnicity and English-language proficiency and experiences with crime and police: A multi-level analysis of restricted data from NCVS. Russell Sage Foundation (2016, $73,874)

• CJARS Fellowship Program (active, $35,000)
Summary: We are here to help you!

TXRDC Administrator – Karin Johnson (karin.johnson@census.gov)
- Direct liaison to Census Bureau
- Assists with proposal development & project planning
- Oversees disclosure review
- Manages the secure computing lab

Director – Mary Campbell (m-campbell@tamu.edu)
- Coordinates with the RDC Admin to oversee the facility
- Oversees contracts with U.S. Census Bureau
- Promotes awareness and shares information
- Nurtures/grows the TXRDC consortium
- Approves lab access for projects

Research scientist – Ling Wu (ling.wu@tamu.edu)
- Spatial analysis for all projects
- Criminal justice research
Opportunities for criminal justice research using restricted data

Q&A

Thanks!