



WELCOME  
TO THE  
NDACAN  
SUMMER  
TRAINING  
SERIES!

- The session will begin at 12pm EST.
- This session is being recorded.
- See ZOOM Help Center for connection issues:  
<https://support.zoom.us/hc/en-us>
- If issues persist and solutions cannot be found through Zoom, contact Andres Arroyo at [aa17@cornell.edu](mailto:aa17@cornell.edu).

# NDACAN SUMMER TRAINING SERIES

National Data Archive on Child Abuse and Neglect  
Bronfenbrenner Center for Translational Research  
Cornell University



# Children's Bureau

An Office of the Administration for Children & Families



# NDACAN SUMMER TRAINING SERIES SCHEDULE

- July 1, 2020 - Introduction to NDACAN
- July 8, 2020 - Historical Data
- July 15, 2020 - Research Example using Historical Data
- July 22, 2020 - Administrative Data (NCANDS, AFCARS, NYTD)
- August 5, 2020 - Research Example using Linked Administrative Data
- **August 12, 2020 - Linking Administrative Data in SPSS**

# SESSION AGENDA

- Barriers to Linking
- Process of Linking
- SPSS Linking Walkthrough

# DEFINITIONS

- Administrative data – data collected by government agencies or large organizations, usually created just for the purpose of record keeping (rather than for statistical purposes)
- Data linking – combining two datasets that share at least one common variable/entity
- By “tables”, we mean a format where data are arranged in rows and columns, rows are cases/records
- A “key” or “linking variable(s)” may be defined distinctly by one or more columns (e.g. Child ID or Child ID x State x Year)
- Columns are variables (fields). We’ll use “variables” in this presentation

## ADMINISTRATIVE DATASETS

- NCANDS: National Child Abuse and Neglect Data System
- AFCARS: Adoption and Foster Care Analysis Reporting System
- NYTD: National Youth in Transition Database

# BARRIERS TO LINKING

# BARRIERS TO LINKING

- The data structure is different between datasets.
  - The AFCARS Foster Care File has one row per child.
  - The NCANDS Child File has one row per report/child.
  - The NYTD Outcomes file has one row per child per wave
- There are entry/recording errors in administrative data.
  - While you should be able to link based on the unique identifier, it is important to add checks to make sure you're linking the right child.
  - For example: you can have a more conservative match if the birth date and sex must be the same as well.

# BARRIERS TO LINKING

- Missing data in key fields
  - If data is missing in the variables that are being used to match then you can't match those rows.
- Changes in record keeping
  - Similar to data structure, data recording may change drastically over the years (e.g. more variables recorded, change in variable codes)
- You have to be familiar with the datasets.
  - The user guides and code books can be of great help in this context! Both are available on our website.

# PROCESS OF LINKING

## LINKING TABLES

- Data tables can only be linked if they share a variable of the same entity
- An “entity” is the object that the variables contain information about
- Linking variables aren’t necessarily named the same
- In the AFCARS set of data files (Foster Care, Adoption, NYTD, and Child File), the common entity is a *child*
- The AFCARS linking variable found in each of the data files is *StFCID* : a unique child identifier comprised of the state abbreviation + the AFCARS foster care ID (e.g. record number)

## STEPS IN LINKING

- Clarify your hypothesis: What variables do you need to do your analysis?
- Specify the datasets you'll need
- Remove all un-needed variables from each dataset
- Subset/filter based on scope of interest (e.g. range of years, states, other characteristics)
- Resolve tables to a single row per unique identifier, here it will be one row per child
- Save results as a new table
- Link the restructured tables

## RESOLVING THE AFCARS FOSTER CARE FILE

- Each year of the Foster Care File has one row per child
- Filtering Rows: You may want just kids who entered, or are in FC at the end of the year, or who aged out, etc
- If using multiple years, duplicate IDs (*StFCID*) will be present – once for each year the child is in foster care.
- Most commonly, the most recent case is kept (use *FY*). It preserves information about the first entry, reason for entry, etc.

## RESOLVING THE NYTD OUTCOMES FILE

- A full Outcomes set (FY2011 or FY2014 Cohort) has three waves.
- The dataset has one row per *StFCID/Wave*. This format is called “long”/stacked.
  - For each child there will be as many rows as waves in which they responded, e.g. up to 3 rows per child
- Data includes children who did not respond to the survey at all – may want to filter these out.
- Resolving to one row per child: create variables as needed for each wave, (e.g. homeless.w1, homeless.w2, etc.)

## RESOLVING THE NYTD SERVICES FILE

- The Services File has one or two records per child per year
- There's a record for each 6-month period the child received services
- The 2018 Services File has data for FY2011 through FY2018
- Unwanted years should be dropped
- To get one record per child, the two six-month records should be consolidated
- Consider if you want only kids who received particular services
- **Not all children from Outcomes file can be found in services file**

## RESOLVING THE NCANDS CHILD FILE

- **The Child File has one row per Report-Child**
  - A report can have multiple children
  - A child can be on multiple reports
- **Each fiscal year is a separate data table and should be (carefully) stacked depending on years of interest**
- **To get one row per child, it may be necessary to aggregate information to populate variables**
  - For example, the number of Reports a child has appeared on

## STEPS IN LINKING

- Clarify your hypothesis: What variables do you need to do your analysis?
- Specify the datasets you'll need
- Remove any unnecessary variables from each dataset
- Resolve tables to one row per child
- Save results as a new table
- Link the restructured tables

# LINKING NYTD 2014 COHORT DATA WITH AFCARS FOSTER CARE

## WHAT DATASETS/VARIABLES DO YOU NEED?

- Want to look at relationship of homelessness of the 2014 NYTD Cohort and foster care, while controlling for sex and race.
- From NYTD Outcomes File:
  - Wave, StFCID, Sex, Homeless, Race, RepYr (created from original variable RemIDt)
- From Foster Care File:
  - StFCID, FY, Sex, Race, CurPISet, Entered, Exited, TotalRem, PHYABUSE, SEXABUSE, NEGLECT.

## PREPARE THE OUTCOMES FILE FOR LINKING

\* Prepare the NYTD file.

```
GET FILE = '/Users/Sarah/Downloads/NDACAN summer series/nytd_summerseries2020.sav'.  
DATASET NAME nytd2014cohortOutcomes WINDOW = FRONT.
```

\* Create a new file with only the relevant variables:.

```
SAVE OUTFILE = '/Users/Sarah/Downloads/NDACAN summer series/nytd14subset.sav'  
  /KEEP Wave StFCID Sex Race Homeless RepYr SubAbuse Incarc.  
DATASET CLOSE nytd2014cohortOutcomes.
```

\* Reload file just saved.

```
GET FILE= '/Users/Sarah/Downloads/NDACAN summer series/nytd14subset.sav'.  
DATASET NAME Outcomes WINDOW=FRONT.  
DATASET CLOSE Outcomes.  
SORT CASES BY StFCID Wave.
```

\* Convert the file from “Long” to “Wide”:

```
CASESTOVARS  
  /ID=StFCID  
  /INDEX=Wave.  
SORT CASES BY StFCID.
```

```
DELETE VARIABLES Incarc.1 Incarc.2 Incarc.3 SubAbuse.1 SubAbuse.2 SubAbuse.3 .
```

```
SAVE OUTFILE = '/Users/Sarah/Downloads/NDACAN summer series/Cohort14Link.sav'.  
DATASET CLOSE ALL.
```

# PREPARE THE FOSTER CARE FILE FOR LINKING

\* Prepare the Foster Care File:.

```
GET FILE = '/Users/Sarah/Downloads/NDACAN summer  
series/fc_summerseries2020.sav'.
```

```
DATASET NAME fosterCare WINDOW=FRONT.
```

```
SAVE OUTFILE = '/Users/Sarah/Downloads/NDACAN summer  
series/FC14Link.sav'
```

```
    /KEEP StFCID FY Sex Race CurPlSet Entered Exited TotalRem  
PHYABUSE SEXABUSE NEGLECT.
```

```
DATASET CLOSE ALL.
```

# LINK THE TABLES

\* Open the files that you prepared for linking:.

```
GET FILE = '/Users/Sarah/Downloads/NDACAN summer series/Cohort14Link.sav'.
```

```
DATASET NAME Cohort14 WINDOW=FRONT.
```

```
GET FILE= '/Users/Sarah/Downloads/NDACAN summer series/FC14Link.sav'.
```

```
DATASET NAME FC14 WINDOW=FRONT.
```

```
DATASET ACTIVATE FC14.
```

\* Join the tables:.

```
STAR JOIN
```

```
  /SELECT t0.FY, t0.Entered, t0.Exited, t0.TotalRem,t0.CurPlSet,t0.PHYABUSE,  
          t0.SEXABUSE,t0.NEGLECT, t1.sex, t1.race,  
          t1.Homeless.1, t1.repYr.1 ,  
          t1.Homeless.2, t1.repYr.2 ,  
          t1.Homeless.3, t1.repYr.3 ,
```

```
  /FROM FC14 AS t0
```

```
  /JOIN Cohort14 AS t1
```

```
    ON t0.StFCID=t1.StFCID
```

```
    IN=t1
```

```
  /OUTFILE FILE= '/Users/Sarah/Downloads/NDACAN summer series/Joined.sav'.
```

```
GET FILE = '/Users/Sarah/Downloads/NDACAN summer series/Joined.sav'.
```

```
DATASET NAME Joined Window = FRONT.
```

\* Check to see if the join was successful:.

```
SELECT IF (t1=1).
```

```
FREQUENCIES VARIABLES=t1.
```

# QUESTIONS?

SARAH SERNAKER:  
STATISTICIAN,  
SS3376@CORNELL.EDU