

WELCOME
TO THE
NDACAN
SUMMER
TRAINING
SERIES!

Linking NCANDS, AFCARS, and NYTD Data

Presenter: Michael Dineen

August 14th, 2019

This webinar 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.

NDACAN SUMMER TRAINING SERIES

NDACAN SUMMER TRAINING SERIES SCHEDULE

July 17th, 2019 - Introduction to NDACAN

July 24th, 2019 - Overview of NCANDS Data

July 31st, 2019 - Overview of AFCARS and NYTD Data

August 7th, 2019 - Strategies for Managing Data

August 14th, 2019 - Linking NCANDS, AFCARS, and NYTD Data

August 21st, 2019 – Article Presentation: “Family surveillance:
Race, police and the reporting of child abuse and neglect.”
Frank Edwards, Rutgers University

SESSION INTRODUCTION

One unique quality of our data is the ability to link youth's child protective service (CPS) history and foster care experiences.

Furthermore, for those who are identified as likely to age out of foster care without permanent placement we can also link their experiences during the transition to independence and the services they receive in preparation to exit with their CPS and foster care experiences.

BENEFITS TO LINKING

Administrative data is a form of “big data”, which has revolutionized social science research. It has several advantages, and these are amplified when administrative data can be linked:

- Allows researchers to ask unique questions

- Adds details and context to individuals’ data in each system

- All-encompassing and specific population

BARRIERS TO LINKING

The data structure is different between datasets.

NYTD and AFCARS are organized by child.

NCANDS data are organized by report incident.

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.

You have to be familiar with the ends and outs of multiple datasets.

The user guides and code books can be of great help in this context! Both are available on our website.

LINKING ADMINISTRATIVE DATASETS

MICHAEL DINEEN, M.A.

ADMINISTRATIVE DATASETS

NCANDS: National Child Abuse and Neglect Data System

AFCARS: Adoption and Foster Care Analysis Reporting System

NYTD: National Youth in Transition Database

DEFINITIONS

By “tables”, we mean a format where data are arranged in rows and columns – a “flat file”, or a “dataset”.

Rows are cases (records)

A “case” may be defined distinctly by one or more columns

Columns are variables (fields). We’ll use “variables” in this presentation

LINKING TABLES

If the variables you need are in two different tables, you can link the tables

The variables you want must be about the same entity

An “entity” is the object that the variables contain information about

In the AFCARS datasets (Foster Care, Adoption, NYTD, and Child File), the common entity is a child.

Both tables must have the columns that define the entity – *StFCID* in our case

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

Resolve tables to one row per child (if needed)

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

Limiting 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*, so a youth may have one to three records. This format is called “long”.

To get one row per child, the table will have to be restructured to “wide”.

Limiting Rows: Youth who are not in the Cohort have no outcomes data. You may or may not want these.

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

RESOLVING THE NCANDS CHILD FILE

The Child File has one row per Report-Child

Each fiscal year is a different dataset. These can be combined.

A child can appear on more than one report

To get one row per child, it may be necessary to use summary functions to populate variables

For example, the number of Reports a child has appeared on, or the type of maltreatment.

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

Resolve tables to one row per child

Save results as a new table

Link the restructured tables

WHAT DATASETS/VARIABLES DO YOU NEED?

Hypothesis: Unstable experiences in foster care predict more negative outcomes

Dependent Variables, from the FY2014 NYTD Outcomes File: *Homeless, SubAbuse, Incarc, Children*

Independent Variables, from the FY2014 Foster Care File: *NumPlep, TotalRem, LifeLOS.*

Demographic variables will be pulled from the Foster Care File.

STATA WALK THROUGH

.do File will be posted on website with video of this presentation

STATA .DO FILE USED IN THE DEMO:

```
cd d:\temp
clear

***** Restructure the Outcomes Table -- Long to Wide *****
use cohort14w3

* Keep only the variables we're going to use
keep wave stfcid homeless subabuse incarc children fy14cohort elig19 elig21

* Convert long to wide
reshape wide homeless subabuse incarc children elig19 elig21, i(stfcid) j(wave)

* Now save this table to disk
save Cohort14_wide

***** Restructure the Foster Care Table *****
clear
use FC2014v7

* Keep only the relevant Foster Care Variables (include demographic vars)
keep fy stfcid state st fipcode sex raceethn race numlep totalrem lifelos

* Save the table
save FC14_vars

**** Link Files ****
use FC14_vars

merge 1:1 stfcid using Cohort14_wide, keep(match using)

save merged_FC_NYTD
```

QUESTIONS?

MICHAEL DINEEN:
STATISTICIAN,
MED39@CORNELL.EDU

NEXT WEEK...

August 21st, 2019 - Concluding Session