WELCOME TO THE 2021 NDACAN SUMMER TRAINING SERIES!

- The session will begin at 12pm EST.

- Please submit questions to the Q&A box.

- This session is being recorded.

NDACAN SUMMER TRAINING SERIES

National Data Archive on Child Abuse and Neglect

Cornell University & Duke University

DATA STRATEGIES FOR THE STUDY OF CHILD WELFARE

NDACAN SUMMER TRAINING SERIES SCHEDULE

- July 7, 2021 - Introduction to NDACAN

- July 14, 2021 - Survey Based Data

- **July 21, 2021 - Administrative Data and Linking**

- July 28, 2021 - VCIS Data and Special Populations

- August 4, 2021 - Multilevel Modeling Workshop

- August 11, 2021 - Latent Class Analysis Workshop

INTRODUCTION
TAMMY WHITE, ACF

SESSION AGENDA

- Overview of Administrative Data (NCANDS and AFCARS)

- Steps in Linking

- Linking Walkthrough

OVERVIEW OF ADMINISTRATIVE DATA

WHAT IS ADMINISTRATIVE DATA?

- *Administrative data* – data collected by government agencies or large organizations, usually created for the purpose of record keeping (rather than for statistical analyses)

NDACAN ADMINISTRATIVE DATA

NCANDS
Child protective history

AFCARS
Foster care experience

NYTD
Transition out of care

Covers progression through the child welfare system

NCANDS

NCANDS - INTRODUCTION

- The National Child Abuse and Neglect Data System (NCANDS) was created as a *voluntary* system in response to a 1988 Child Abuse and Prevention and Treatment Act (CAPTA) amendment.

- Case-level data are collected for all children who received a response from a child protective services (CPS) agency in the form of an investigation or an alternative response.

NCANDS CHILD FILE

- A file containing:

  - child-specific *records* for each report of alleged child abuse and neglect that received a CPS response

  - complete *reports* that resulted in a disposition (or finding) during the reporting year.

- **Report-Child Pair = NCANDS Record**

  - Combines the Report ID and Child ID to uniquely identify a single record

CHILD FILE: REPORT AND RECORD CHARACTERISTICS

CHILD FILE REPORT

- Reports suspected child abuse

- May involve one or more children (records)

- May be substantiated, unsubstantiated, alternative response, or other dispositions

- May involve multiple perpetrators

CHILD FILE RECORD

- Data related to only one child in a given report

- Data representing a victim or nonvictim

- Data in all fields (for victim records)

- Data concerning perpetrators (up to three) for victim records

CHILD FILE: VARIABLES

REPORT VARIABLES

- Report Data (report IT, report date, disposition, investigation start date)

- Child Data (child ID, demographics, *note: child age recorded as "22" or "23" indicates sex trafficking maltreatment type*)

- Maltreatment Data (types such as sex trafficking and medical neglect, individual dispositions, fatalities)

- Child Risk Factors (sexual abuse, substance abuse, diagnosed disabilities)

CHILD VARIABLES

- Caregiver Risk Factors (substance abuse, financial problems)

- Services (foster care, adoption, counseling)

- Perpetrator Data (relationship to victim, demographics, sex trafficking)

- Additional Fields (AFCARS ID, date of death, plan of safe care, referral to CARA-related services)

NCANDS AGENCY FILE

- CAPTA required items

- Summary data

  - Prevention Services

  - Referrals and Reports

  - Additional information on Child Victims Reported in the Child File

  - Child Fatalities

  - Part C of Individuals with Disabilities Education Act (IDEA) reporting

AFCARS

AFCARS - INTRODUCTION

- Case-level information on children who are under the placement and care responsibility of of Title IV-E child welfare agencies

- Collection:

  - States document information in their electronic record system

  - States compile data and send to Children's Bureau

  - Children's Bureau works with states to correct errors

AFCARS VARIABLES

- AFCARS includes demographic, removal, placement, and other case related information. Examples include:

  - Date of birth of the child, caretakers of the child, and foster/adoptive parents.

  - Race information (self-identified) on the child and foster parents.

  - Date of first and recent removal, number of removals, and discharge date.

  - Date of placement, number of placements, and placement location.

  - Case plan goals, termination of parental rights dates, and sources of federal financial support.

WORKING WITH MULTIPLE YEARS OF AFCARS

- Years of the AFCARS files can be "stacked"

- When more than one year of the Foster Care File is used, there will be duplicated AFCARS IDs (StFCID)

- A child has a record for each year that they are in foster care

- If resolving data to one row per child, keep the most recent year

STEPS IN LINKING

DEFINITIONS

- *Administrative data* – data collected by government agencies or large organizations, usually created just for the purpose of record keeping (rather than for statistical analyses)

- *Data linking/joining* – combining two datasets that share at least one common variable

- By *tables*, we mean a format where data are arranged in rows and columns, where rows are *observations* or *cases/records* and columns are *variables/fields*.

    - Variable names will be referenced using `this font`

- A *case* or *key* is a variable or combination of variables that uniquely identifies a record/observation (e.g. Child ID or Child ID x State x Year)

LINKING DATA

- Data tables can only be linked if they share at least one variable of the same *entity*, i.e. the object that the variable(s) 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), commonly used 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

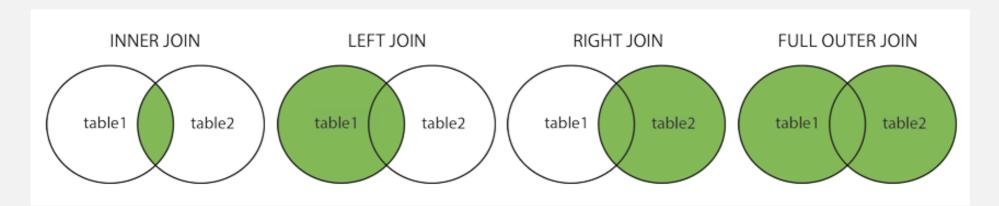1. Prepare each table separately before linking

   - What variables are needed/not needed in research? – remove unnecessary variables

   - Which datasets have variables of interest?

   - Take subset based on research question, ex. only children under 15, or only reports from Maryland

   - Append datasets of different years, ex. NCANDS child files are given as separate files for each year

STEPS IN LINKING

2. Data should be reduced to one row per *unique* identifier (may use more than one variable), e.g. one row per child per year if working with multiple years.

- Reduce/*resolve* data tables to one row that summarizes the child's full record within a year (e.g. going from *long* format to *wide* format)

- If data is not resolved to one row per unique identifier, duplicate rows may arise when linking

STEPS IN LINKING

3. Save cleaned data as new tables

4. Use functions within programming language of choice to perform linkage, specifying which variable(s) are the linking variables

- If linking more than 2 datasets: first link 2, save the linked set as a new table, run through checklist again before joining next data

JOINING THE DATA

- Most programing languages offer built-in functions to facilitate linkage; often referred to as *join/merge statements*

INNER JOIN	LEFT JOIN	RIGHT JOIN	FULL OUTER JOIN
table1 table2	table1 table2	table1 table2	table1 table2

Returns rows that have matching key variables in both tables.

Returns all rows from left table and only the records that match the key variables from the right table.

Returns all rows from the right table, and only the rows that match the key variables from the left table.

Returns all rows from both tables, matching key variables where possible and (usually) leaving missing values in anything that doesn't match

https://www.w3schools.com/sql/sql_join.asp

RESOLVING 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 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

RESOLVING AFCARS FOSTER CARE FILE

- Each year of the foster care file has one row per child

  - If using multiple years, a child may appear more than once, e.g. duplicate `StFCID`

- Take subset based on scope of research, e.g. children just entering FC in a year, children who left FC in a year

- Reduce each row to summarize all of the records of one child

- Most commonly, the most recent case is kept (use `FY`). It preserves information about the first entry, reason for entry, etc.

  - A child's record through FC is preserved from when they first enter FC, thus the most recent entry should contain information like services received, when they exited and/or re-entered

PROGRAMMING LINKAGE

- Here are examples of *functions* (written in `this font`) in commonly used languages that can be used to link data:
  - SAS
    - `PROC SQL`

  - SPSS:
    - `star join`
    - `match files`

LIMITATIONS

- In general, data privacy can limit linking

- Missing data in key fields

- Errors/inconsistencies in record keeping

- Changes in record keeping

SUMMARY

- Data linkage is possible in any two datasets that share at least one common, but unique, identifier

- The data to be linked should be separately cleaned and prepared before joining

- For child level analysis, most NDACAN administrative datasets can be linked using the variable `StFCID`

LINKING WALKTHROUGH

QUESTIONS?

CLAYTON COVINGTON

CLAYTON.COVINGTON@DUKE.EDU

SARAH SERNAKER
STATISTICIAN, DUKE
SARAH.SERNAKER@DUKE.EDU

NEXT WEEK…

Date: July 28, 2021

Topic: VCIS Data and Special Populations

Presenters: Alex Roehrkasse, Ph.D., and Frank Edwards, Ph.D.