Welcome to the NDACAN Summer Training Series!

- The session will begin at 12PM Eastern Time
- Please turn video off and mute the line.
- 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.

NDACAN Summer Training series

National Data Archive on Child Abuse and Neglect

Bronfenbrenner Center for Translational Research

Cornell University

New Horizons for Child Welfare Data

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)
July 29, 2020 - Linking Administrative Data in SPSS
August 5, 2020 - Research Example using Linked Administrative Data

Session Agenda

- Why use historical data?
- Organizing NDACAN data for historical analysis
- Illustration: long-term trends in children in substitute care

WHY USE HISTORICAL DATA?

What is historical data?

- Definition 1: Data that describes past attributes or events
  - E.g. National Incidence Study of Child Abuse and Neglect, 2006
- Definition 2: Data that describes attributes or events over historical time
  - E.g. Adoption and Foster Care Analysis and Reporting System, 1995-2018

Why should I use historical data?

- Documenting trends in historical time
  - E.g. substantiations of child neglect per 1,000 children in the U.S., 1990–2018
- Measuring stasis or change in relationships
  - E.g. association between child poverty rates and child neglect rates in U.S., 1990–2018
- Identifying causal relationships
  - E.g. the effect of welfare generosity on child neglect rates in the U.S., 1990–2018
- Predicting future outcomes
  - E.g. substantiations of child neglect by state, U.S., 2020–2025

What NDACAN Data are good for historical research?

- Existing administrative data
  - Adoption and Foster Care Analysis and Reporting System (AFCARS), 1995–2018
  - National Child Abuse and Neglect Data System (NCANDS), 1990–2018
- New administrative data
  - Voluntary Cooperative Information System (VCIS), 1982–1995
  - Children's Bureau Statistical Series (CBSS), 1962–1975

New historical Administrative data

- VCIS
  - Annual state-level data on children entering, in, and exiting substitute care
  - Cross tabulated by race/ethnicity, age, sex, living arrangement, etc.
- CBSS
  - Annual state-level data on children in substitute care
- Limitations
  - Reliability
  - Missing data

ORGANIZING NDACAN DATA FOR HISTORICAL ANALYSIS

NDACAN data structures

- Data already aggregated by state-by-year
  - NCANDS agency file
  - Voluntary Cooperative Information System (VCIS)
  - Children's Bureau Statistical Series (CBSS)
- Individual-level data that can be aggregated
  - AFCARS foster care file: state, county, and year IDs
  - AFCARS adoption file: state and year IDs
  - NCANDS child file: state, county, and year IDs

AGGREGATING individual-level data: STATA

- Count of all children:
  - `collapse (count) id, by(state year)`
- Count of children by race:
  - `collapse (count) id, by(state year race)`

AGGREGATING individual-level data: R (tidyverse)

- Count of all children:
  - `df %>% group_by(state, year) %>% summarize(n = n())`
- Count of all children by race:
  - `df %>% group_by(state, year, race) %>% summarize(n = n())`

AGGREGATING individual-level data: SPSS

- Count of all children:
  - `AGGREGATE`
    - `/OUTFILE = 'data.sav'`
    - `/BREAK = state year`
    - `/n = NU(id)`
- Count of all children by race:
  - `AGGREGATE`
    - `/OUTFILE = 'data.sav'`
    - `/BREAK = state year race`
    - `/n = NU(id)`

Constructing useful measures

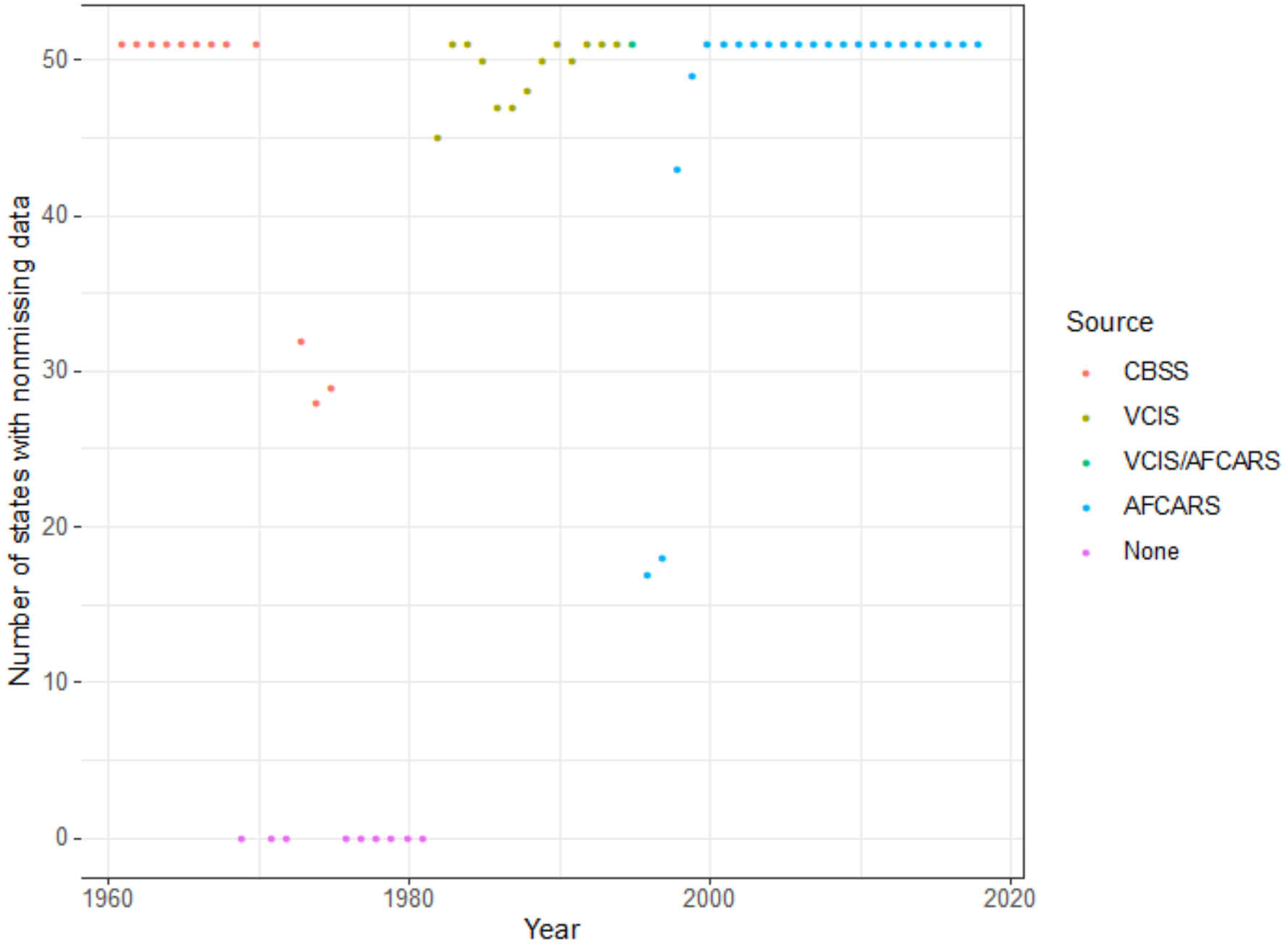
- Instead of counts, we often want rates or proportions
- Sometimes NDACAN data themselves provide useful denominators
- Other times, other data is needed:
  - SEER: U.S. Census Bureau intercensal estimates of population by county, age, and race/ethnicity
  - NHGIS: Decennial Census (DC) and American Community Survey (ACS) data at various levels of geographic aggregation
  - IPUMS-USA: DC and ACS microdata
  - IPUMS-CPS: Current Population Survey microdata

Cautionary notes

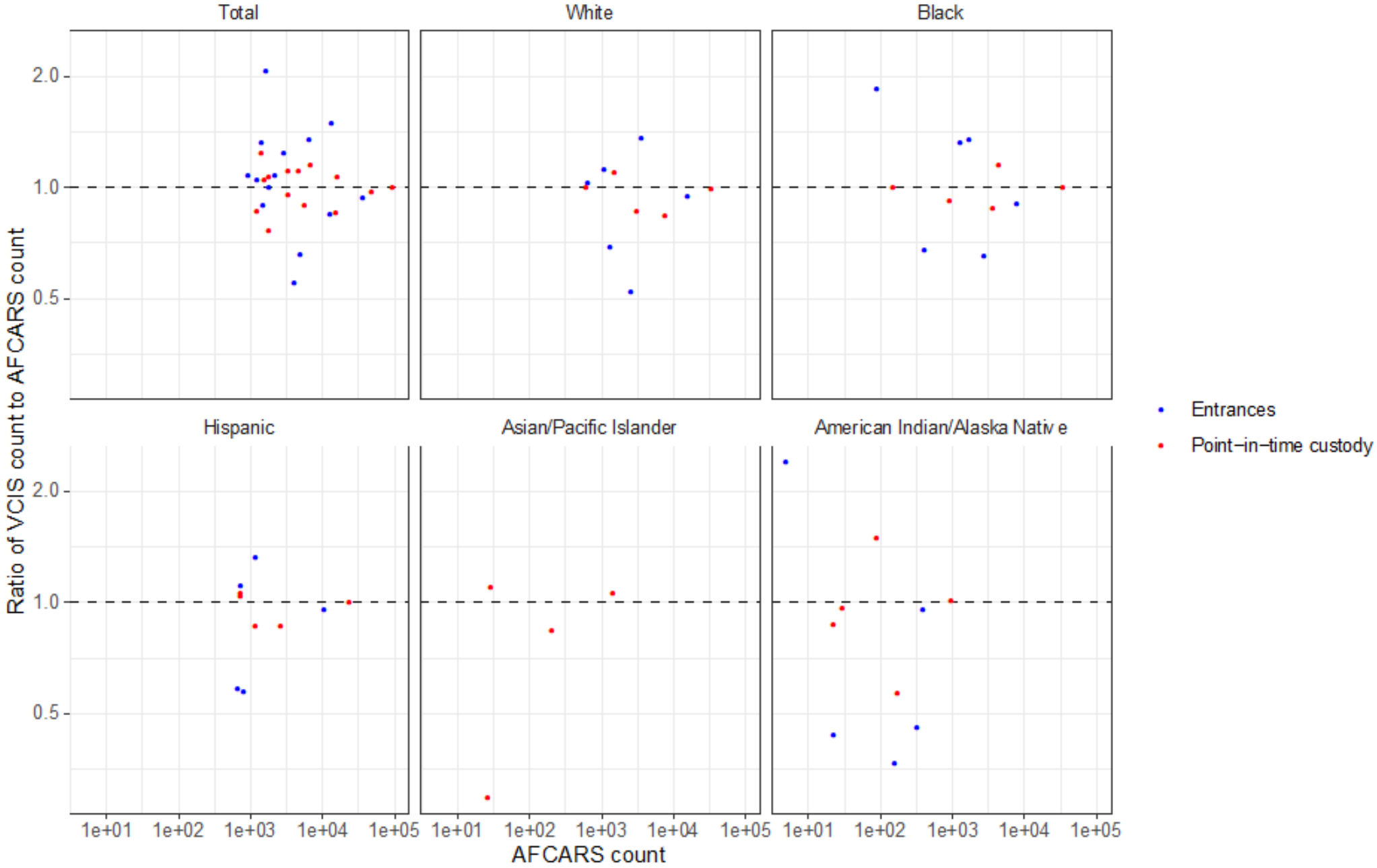
- Are data consistent across time?
  - E.g. AFCARS measures race/ethnicity differently before/after 2000
- Are data consistent across sources?
  - E.g. AFCARS includes runaways, VCIS and CBSS do not
- Does data missingness, suppression, or sampling lead to measurement error?
  - E.g. state–years missing irregularly from all administrative data sources
  - E.g. AFCARS/NCANDS suppress county ID if county contains fewer than 1k cases or if case involves child death
  - E.g. estimates based on ACS, CPS should account for error in variables (EIV)

ILLUSTRATION: LONG-TERM TRENDS IN CHILDREN IN SUBSTITUTE CARE

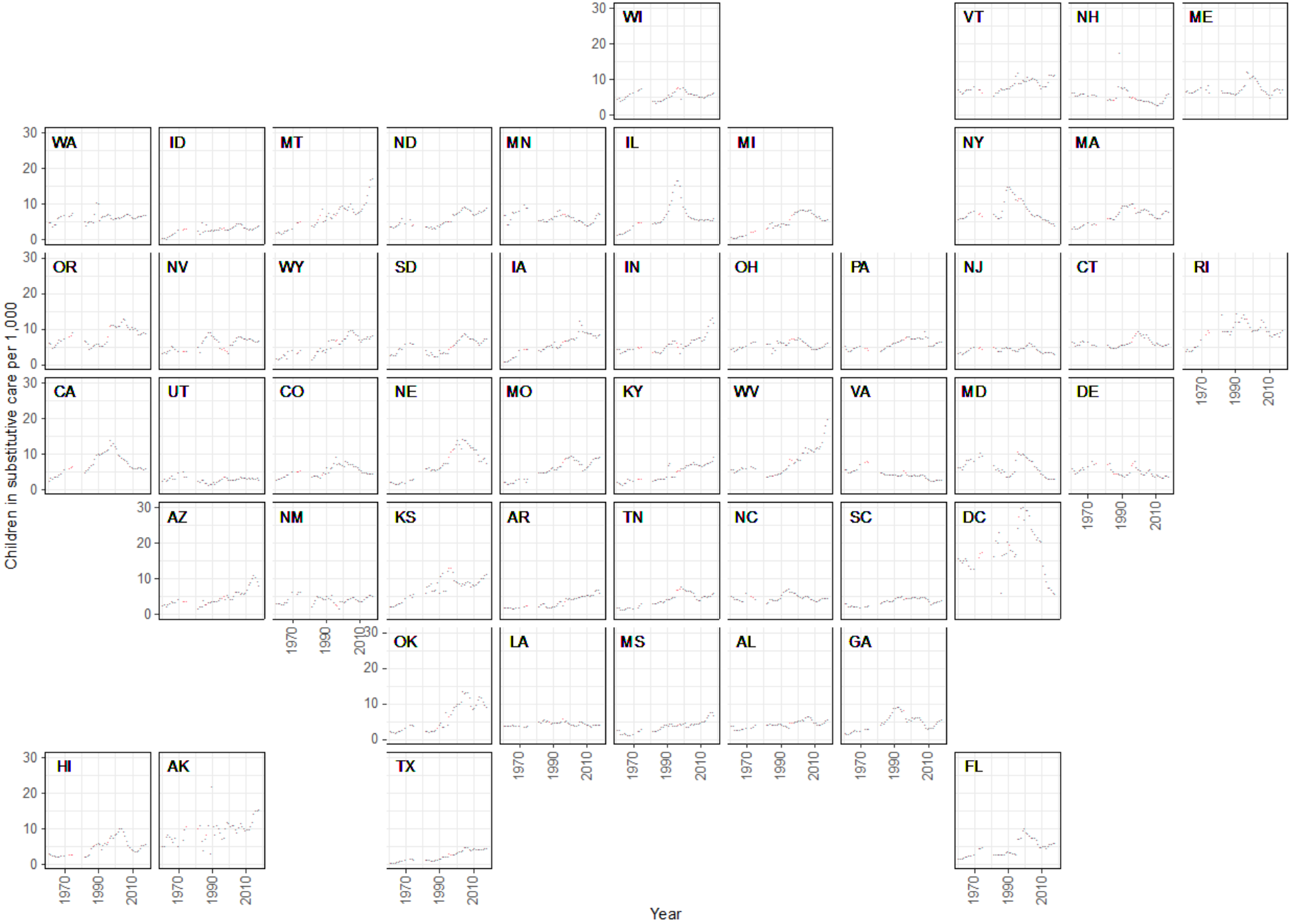
Availability of data



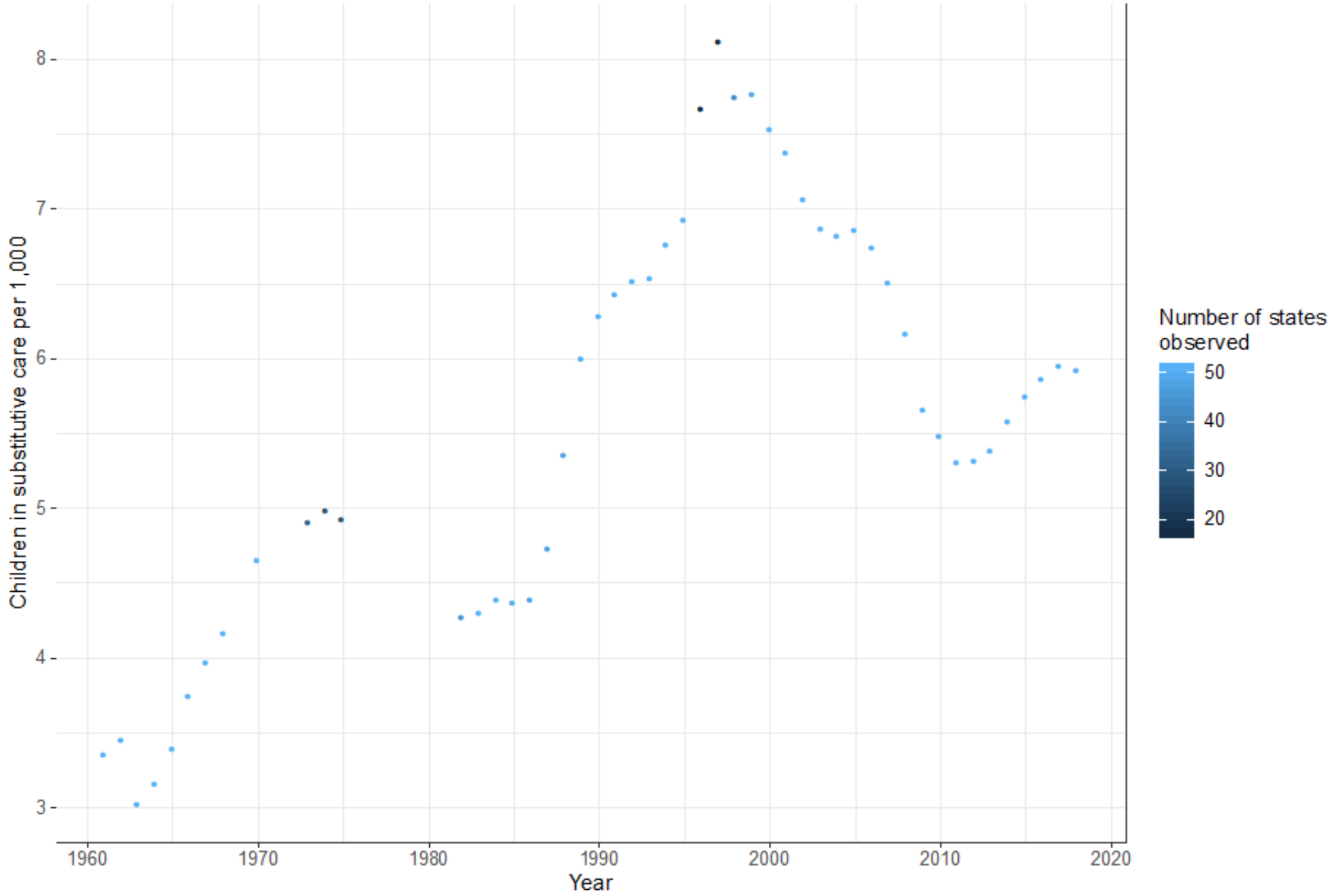
Reliability of measures



State trends in substitute care



National trend in substitute care



QUESTIONS?

ALEX ROEHRKASSE
POSTDOCTORAL ASSOCIATE

AFR66@CORNELL.EDU

Next week…

July 15, 2020
Research Example Using Historical Data Historical Data

Presenter(s):
Alex Fort Roehrkasse, Ph.D.