

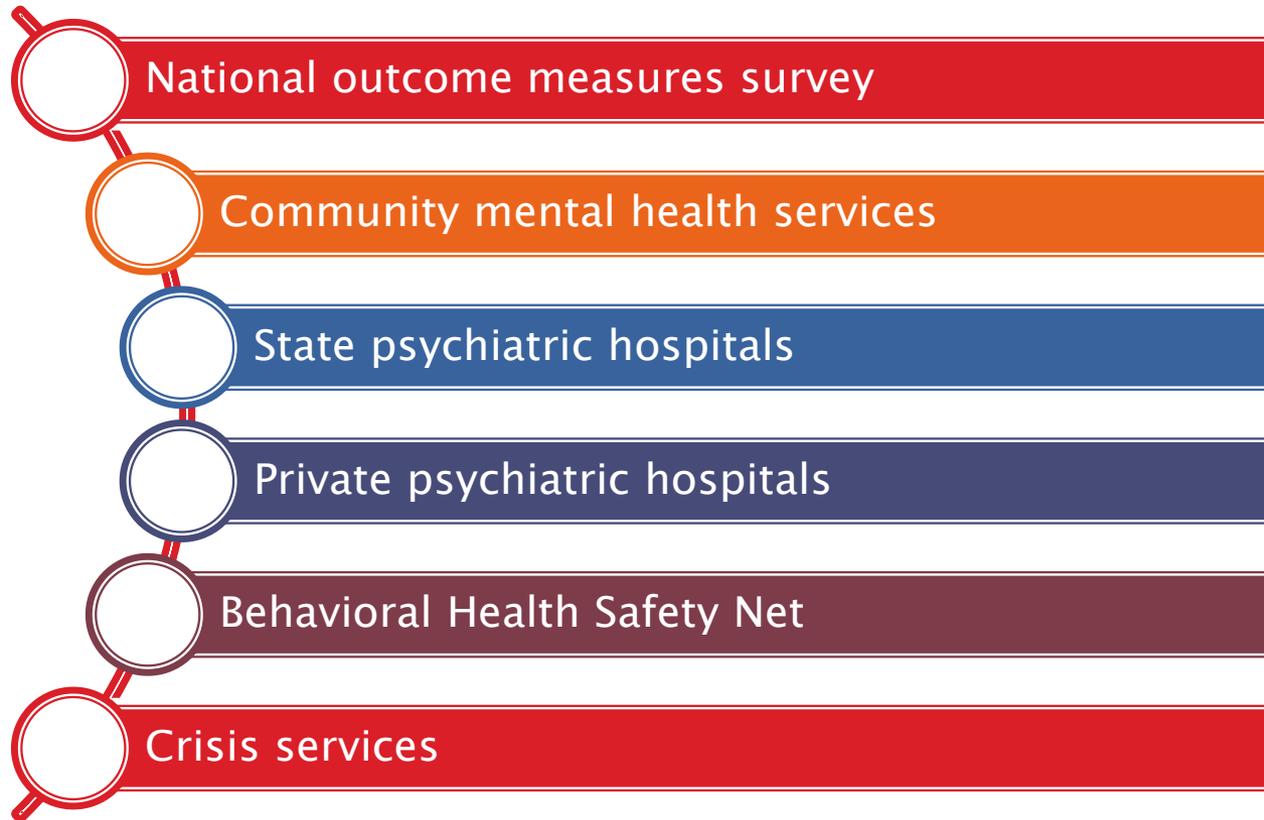


Creating unduplicated client counts

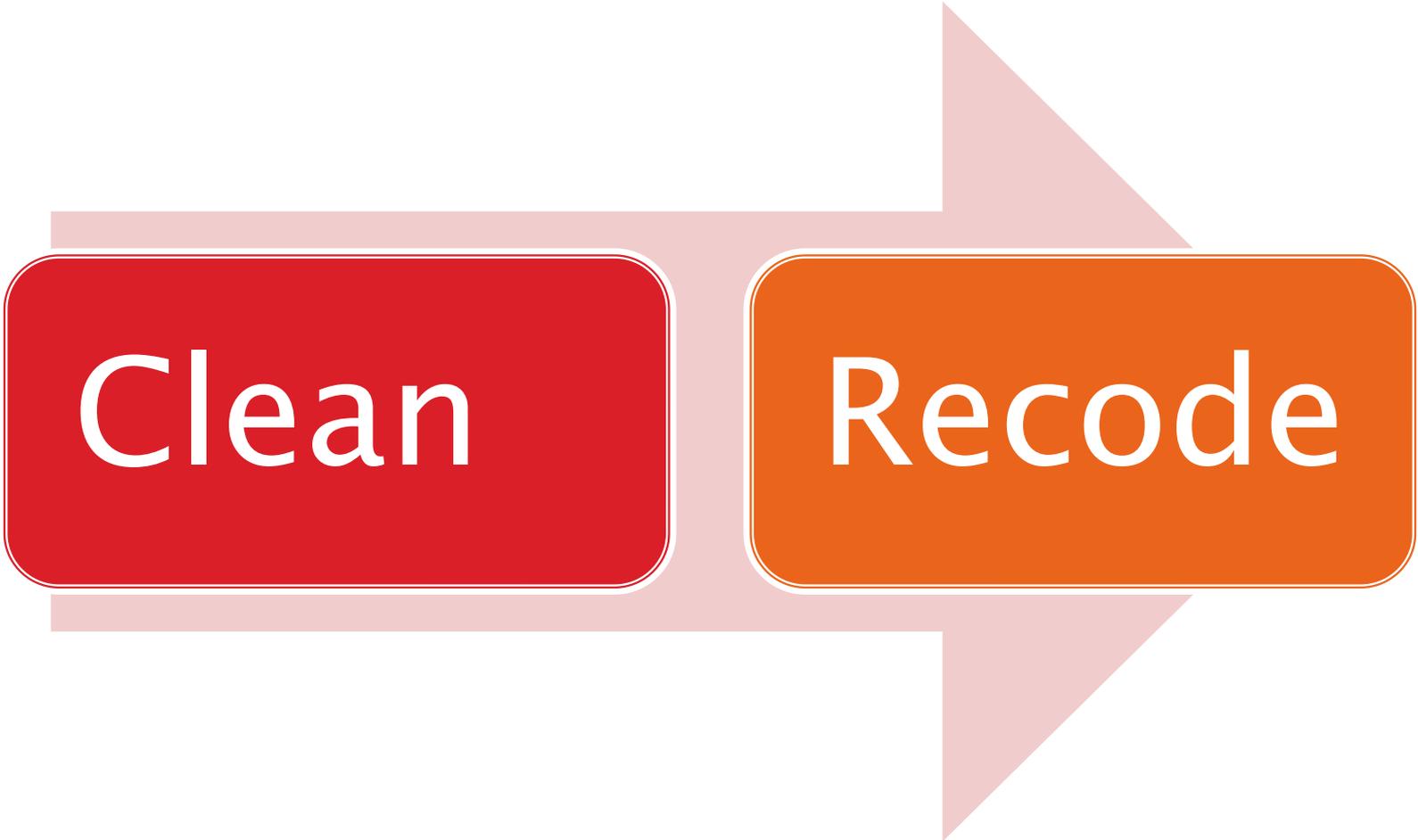
Prepared by
Lymari Benítez, Ph.D., Statistical Research Specialist
Karen Edwards, Ph.D., Director, Office of Research

April 13, 2015

Tennessee mental health data sources



Take each data set and...



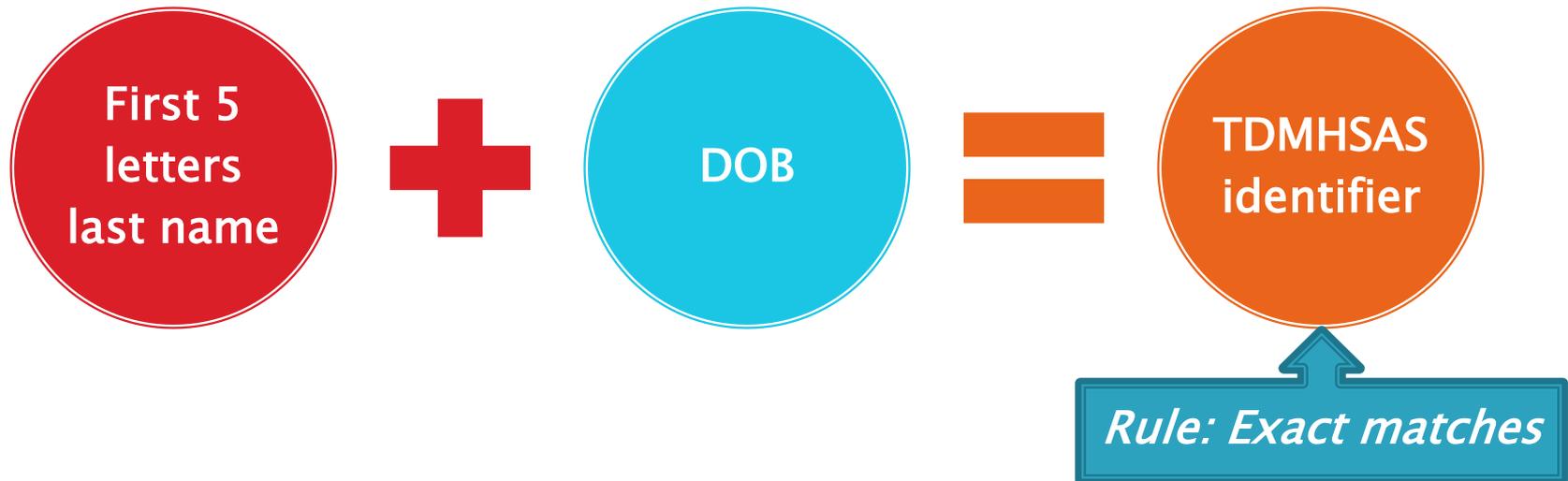
Clean

Recode

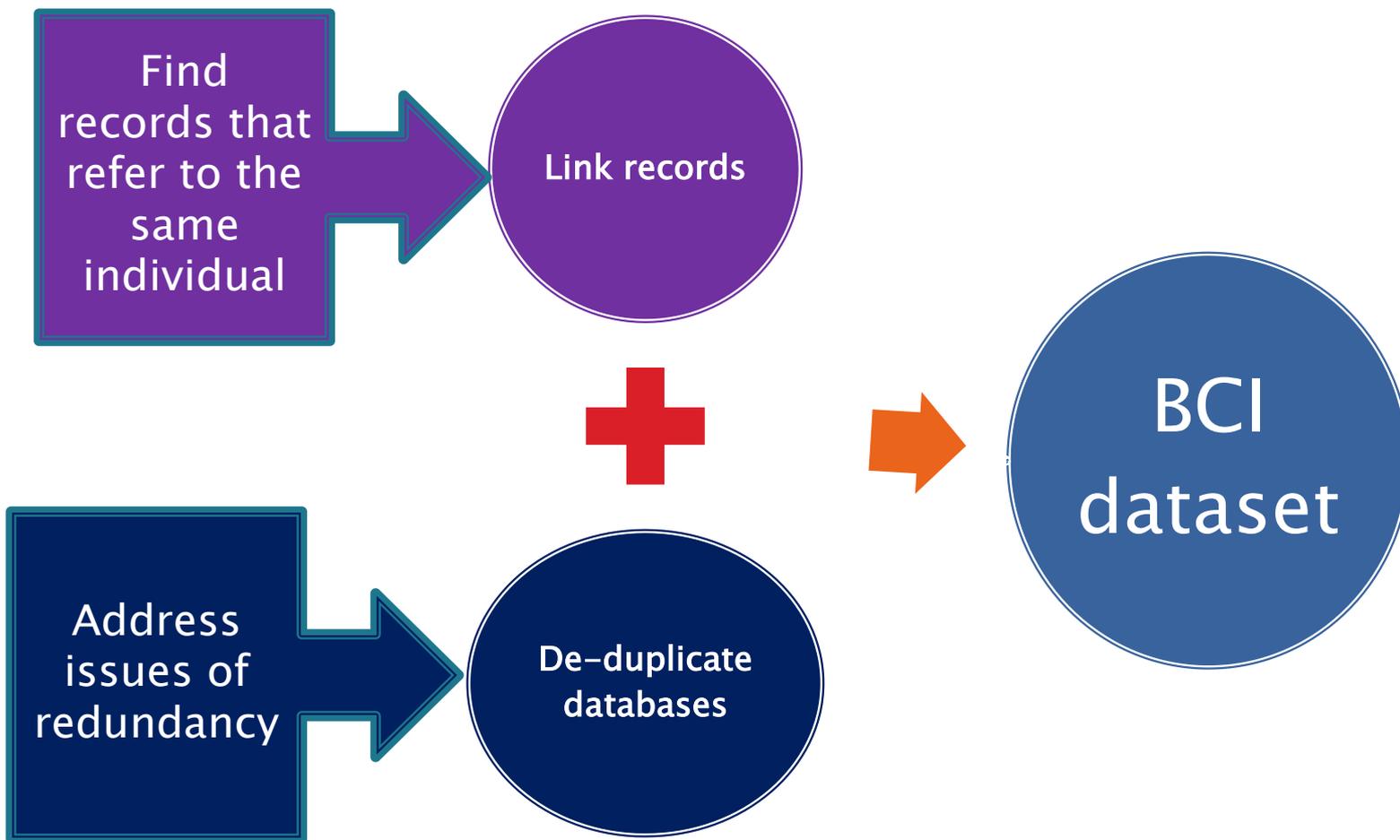
Assign an identifier



- ▶ TDMHSAS identifier
 - based on last name and date of birth (DOB)

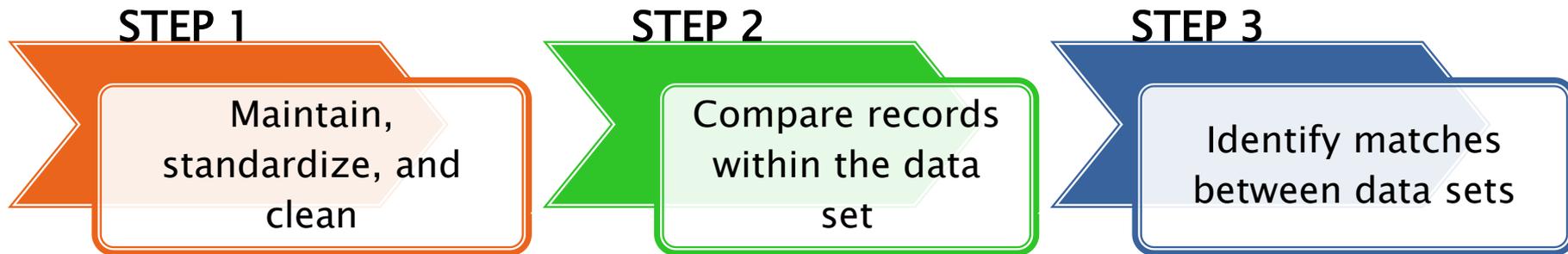


Link and de-duplicate records to create Basic Client Information Dataset (BCI)



TDMHSAS Process

Identify, validate and reconcile cases with the goal to identify the 'best match' within and between set(s).



Step 1: Maintain, standardize and clean

Create an identifier using SPSS syntax

- ▶ Separate first name from last name
- ▶ Eliminate blank spaces
- ▶ Get rid of periods, commas, apostrophes and hyphens in names
- ▶ Eliminate anything included in parenthesis
- ▶ Participants without last name are removed from reporting
- ▶ Identify if dummy variables were used
- ▶ Participants with invalid DOBs are removed
- ▶ Participants without DOBs are removed



Last name



Date of birth

Step 2: Compare records within the data set

Link records within a data set using available variables

- ▶ Use each dataset's unique identifier
- ▶ Define matching cases by last name, DOB, and gender using SPSS 'identify duplicate cases'
- ▶ Use SPSS syntax to create TDMHSAS identifier

Identify duplication, redundancy, inconsistency

▶ Identify duplication or redundancy

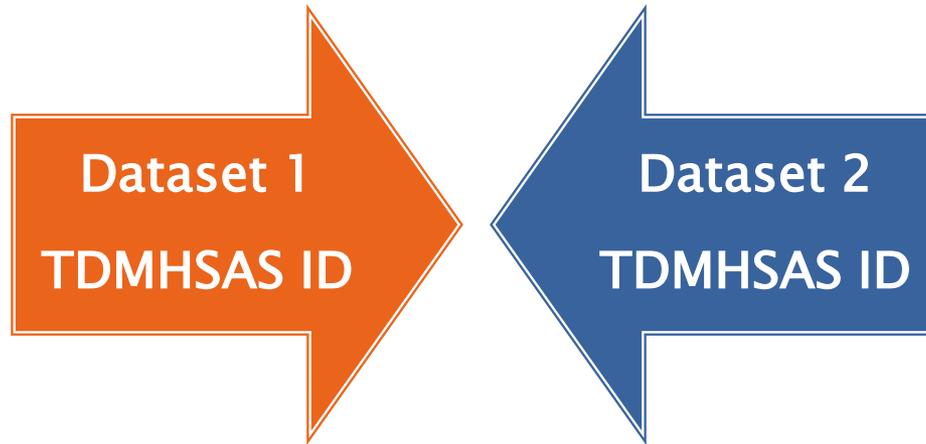
- Records with the same service dates and service
 - *Rule: most complete record is chosen*

▶ Identify inconsistency

- Records with same identifier but inconsistent demographic information.
 - *Rule: most recent record is chosen*

Step 3: Identify matches between data sets

Use TDMHSAS identifier to match cases



- ▶ Apply rules to create basic client information record
 - *Rule: Prioritize trustworthy sources.*
 - *Rule: Remove cases with same identifier but different demographic information.*

Lessons learned

- ▶ Matching improves with clean databases.

Data reviews and quality protocols for each data source are important: give feedback to data sources

- ▶ Variables for linkage should provide unique information about each consumer.
- ▶ Record linking and de-duplication with SPSS takes time and requires strict protocols for data management.

TDMHSAS contacts

- ▶ Karen Edwards, Ph.D.
Director, Office of Research
Tennessee Department of Mental Health and Substance Abuse Services
karen.edwards@tn.gov
Phone: 615-532-3640

- ▶ Lymari Benítez, Ph.D.
Statistical Research Specialist, Office of Research
Tennessee Department of Mental Health and Substance Abuse Services
Lymari.benitez@tn.gov
Phone: 615-741-0628