

---

# SAMHSA's IDB Project

## Record Linking

*Combining Files without a Common Identifier*



U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES  
Substance Abuse and Mental Health Services Administration  
Center for Substance Abuse Treatment  
Center for Mental Health Services  
[www.samhsa.gov](http://www.samhsa.gov)

***SAMHSA***

## Project Information

---

- Project: *Integrated Medicaid and State Agency Mental Health and Substance Abuse: Data Analysis and Technical Assistance*
- Jointly Funded: CSAT and CMHS
- Project Officers:
  - Rita Vandivort (CSAT)
  - Jeff Buck (CMHS)



U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES  
Substance Abuse and Mental Health Services Administration  
Center for Substance Abuse Treatment  
Center for Mental Health Services  
[www.samhsa.gov](http://www.samhsa.gov)

***SAMHSA***

# Project Team

---

- Medstat (Prime Contractor)
- RTI International
- NASADAD
- NASMHPD Research Institute



U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES  
Substance Abuse and Mental Health Services Administration  
Center for Substance Abuse Treatment  
Center for Mental Health Services  
[www.samhsa.gov](http://www.samhsa.gov)

***SAMHSA***

## Project Introduction / Background

---

- In 1997, SAMHSA and Medstat built an Integrated Database (IDB) linking client-level mental health (MH) and substance abuse (SA) data with Medicaid claims and eligibility.
- Tracking individuals across multiple data systems, the IDB captures nearly-total service utilization for the first time.



U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES  
Substance Abuse and Mental Health Services Administration  
Center for Substance Abuse Treatment  
Center for Mental Health Services  
[www.samhsa.gov](http://www.samhsa.gov)

***SAMHSA***

## Project Introduction / Background (cont'd)

---

- The IDB includes:
  - Multiple years of services data (1996-1998);
  - Three States: Delaware, Oklahoma, and Washington;
  - Both a broad data warehouse and narrowly-focused analytic files.
- The project surmounted difficult challenges of confidentiality, disparate data structure and person-level linking.



U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES  
Substance Abuse and Mental Health Services Administration  
Center for Substance Abuse Treatment  
Center for Mental Health Services  
[www.samhsa.gov](http://www.samhsa.gov)

***SAMHSA***

## Data Integrated for the IDB

---

- No common file key
- SSN not available for 25 to 35 percent of the records
- Medicaid IDs collected by MH/SA Agencies usually not current/accurate
- Disparities in names
  - Misspellings
  - Nicknames
  - Maiden names



U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES  
Substance Abuse and Mental Health Services Administration  
Center for Substance Abuse Treatment  
Center for Mental Health Services  
[www.samhsa.gov](http://www.samhsa.gov)

***SAMHSA***

# Match-Merging Related Files

---

- Match-merging files from a single authority is usually very accurate
- Agency specific identifiers are often used as file keys
- Identifiers from a single authority are generally reliable – errors are rare



U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES  
Substance Abuse and Mental Health Services Administration  
Center for Substance Abuse Treatment  
Center for Mental Health Services  
[www.samhsa.gov](http://www.samhsa.gov)

***SAMHSA***

# Problems with Match-Merge Methods

---

- Problems when keys are incorrect
- May occur because of omissions and errors
- Two outcomes:
  - Records do not match when they should
  - Records match when they should not



U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES  
Substance Abuse and Mental Health Services Administration  
Center for Substance Abuse Treatment  
Center for Mental Health Services  
[www.samhsa.gov](http://www.samhsa.gov)

***SAMHSA***

# Match-Merging Unrelated Files

---

- There are often problems merging files from separate agencies – even with a common identifier (i.e., SSN)
- Omissions and errors are more prevalent with identifiers that are not specific to an agency
- Manual review of SSN match-merges reveal many errors
  - records that should be linked but are not
  - incorrect links



# Probabilistic and Deterministic Linking

---

- Related techniques – overcome limitations of match merging
- Makes linking possible, even with
  - Missing information
  - Errors in data
- Uses multiple criteria – identifiers
- More work than match-merging



# Probabilistic Linking

---

- Similar to deterministic linking
  - Multiple criteria/comparisons
  - Scores to determine links
- Differences from deterministic linking
  - Points and scoring not known beforehand
  - Commonly uses disagreements as well as agreements
  - More complex



# Useful identifiers

---

- As a general rule, strong identifiers are better than weak ones
- But strong is not the same as good or useful
- A useful identifier is available on all files
- An identifier found on a single file cannot be used for comparisons



U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES  
Substance Abuse and Mental Health Services Administration  
Center for Substance Abuse Treatment  
Center for Mental Health Services  
[www.samhsa.gov](http://www.samhsa.gov)

***SAMHSA***

# IDB Record Linking

---

- Probabilistic record linking
- Personal Identifying Information
  - Name, ID, DOB
  - Demographics (i.e., Gender, Race)
- Flexible approach
  - Links determined by data
  - Weights/Importance of identifiers determined by values and prevalence within the data



# IDB Process

---

- Linking
  - Concatenating Data
  - Scaling Factors
  - Blocking, Joining, and Comparing
  - Initial Links – Deterministic
  - Probabilistic Iterations
- Manual review
- Mapping of IDs



# General Comparison of Linking

---

- Based on analysis of two states
- Relative to overlap from Probabilistic Population Estimate (Pandiani & Banks)
- Links found
  - Probabilistic linking: 80-86%
  - Match merge: 51-72%
  - Deterministic links: 59-76%



U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES  
Substance Abuse and Mental Health Services Administration  
Center for Substance Abuse Treatment  
Center for Mental Health Services  
[www.samhsa.gov](http://www.samhsa.gov)

***SAMHSA***

# Accuracy of Methods

(1)	(2)	(3)	(4)	(5)	(6)
State	Probabilistic Linking Total	Links Not Found with Match Merge	Common Links	False Links from Match Merge	Match Merge Total
Delaware	5,612	1,488	4,124	523	4,647
Oklahoma	38,683	18,655	20,028	6,809	26,837
Washington AOD-Medicaid	14,836	4,766	10,070	588	10,658
Washington MH-Medicaid	56,990	35,135	21,855	995	22,850
Washington AOD-MH	7,830	6,562	1,268	2,412	3,680

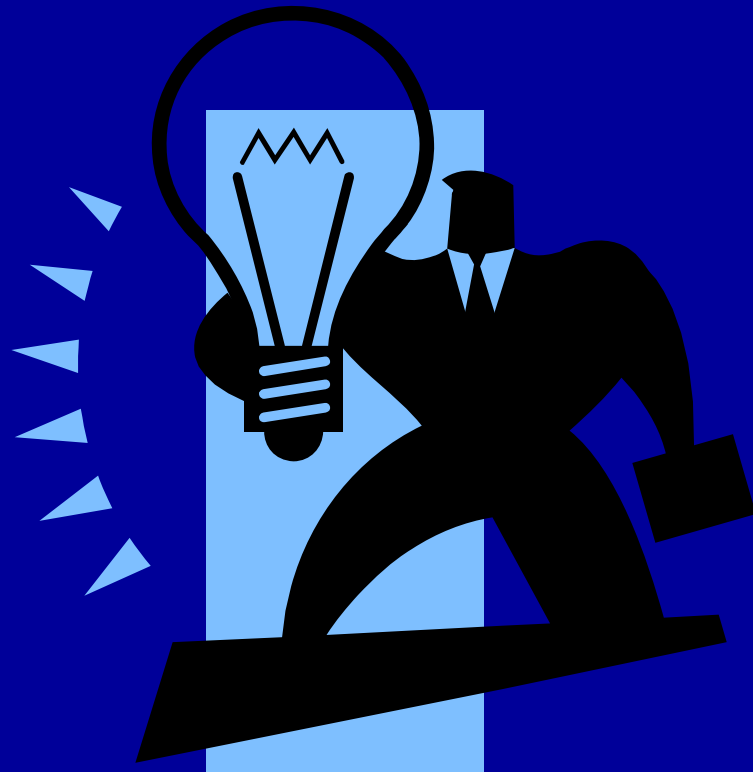


**U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES**  
 Substance Abuse and Mental Health Services Administration  
 Center for Substance Abuse Treatment  
 Center for Mental Health Services  
[www.samhsa.gov](http://www.samhsa.gov)

**SAMHSA**

# Conclusion/Discussion

---



**U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES**  
Substance Abuse and Mental Health Services Administration  
Center for Substance Abuse Treatment  
Center for Mental Health Services  
[www.samhsa.gov](http://www.samhsa.gov)

***SAMHSA***