

Data Linkages for Data Quality Improvement for American Indians and Alaska Natives

David Espey, MD

Melissa Jim, MPH

CDC/IHS



Linkages between:

- ◆ IHS patient registration data and state death records
- ◆ IHS patient registration data and central cancer registries



Assessing and improving the quality of mortality/cancer incidence data for American Indians

- Link records from state vital statistics and cancer registries with IHS patient registration database
- Identify AI/AN records misclassified as non-Native
- Use “improved” data to report mortality patterns and cancer burden of American Indians



Linkage Process

- Proposal drafted & presented to states
- Research determinations (CDC and IHS) & MOA's
- Data management & linkages
- Analysis plan



Data Management

- MOA/Data Use Agreement
- Data sent to IHS Division of Epidemiology in Albuquerque
- Linkage performed
- Send back Match File 3
 - Contains both IHS & state information that was used to determine if the match was a link
- Create analysis files



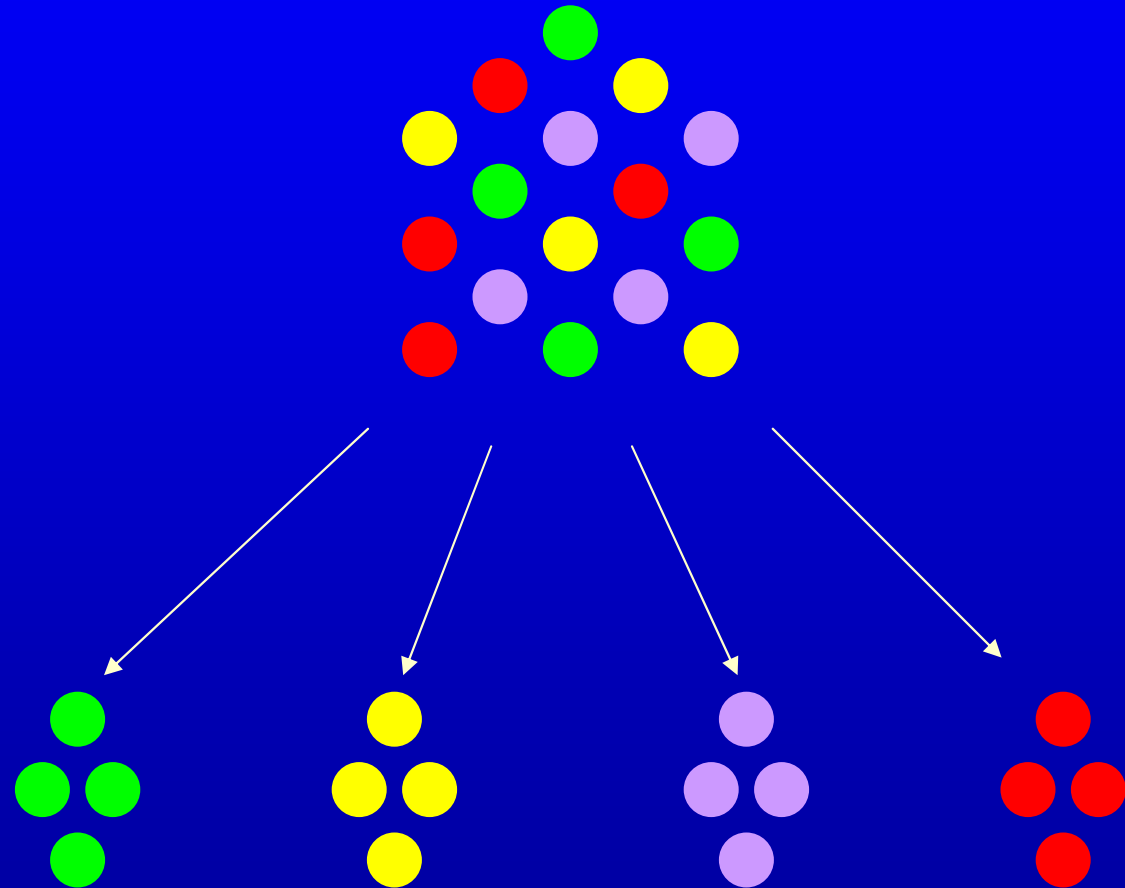
LinkPlus

- Probabilistic record linkage program developed at the Centers for Disease Control and Prevention (CDC)
- Free linkage software available for download at the CDC website
<http://www.cdc.gov/cancer/registryplus/lp.htm>



Blocking Variables

- SSN
- Last Name
- First Name
- Date of Birth



Matching Variables

- SSN
- Date of Birth
- Last Name
- First Name
- Date of Death
- Middle Name
- Sex



Example

Considered a "Match"

Melissa	Ann	Jim	06021976	123456789
Melisa	A	Jim	06021976	123456789

Considered a "Clerical Review"

Melissa	Ann	Jim	06021976	123456789
M	A	Jin	06991976	213456789



Link Plus - [Linkage Configuration]

File Help

File 1 (Cancer Registry Data)

CRS Plus
 NAACCR 9
 Delimited
 Fixed Width

Data File:

File 2 (External Data)

Delimited
 Fixed Width

Data File:

Select blocking variables and phonetic systems

Data Item Name (File 1)	Data Item Name (File 2)	Phonetic System
SSN	SSN	
Last Name	Last Name	
First Name	First Name	

Select ID variables (File 1)

IHS Reckey

Select matching variables and matching methods

Data Item Name (File 1)	Data Item Name (File 2)	Matching Method
SSN	SSN	SSN
Date of Birth	Date of Birth	Date
Last Name	Last Name	Last Name
First Name	First Name	First Name
Date of Death	Date of Death	Date
Middle Name	Middle Name	Middle Name
Sex	Sex	Exact

Select ID variables (File 2)

VS UniqueID

Missing Value (File 1)

Missing Value (File 2)

Matching Methods

Last Name
 First Name
 Middle Name
 SSN
 Date
 Exact
 Generic String
 Value-specific (frequency based on)

Add Remove Add Remove

Please use this Graphical Editor to set the parameters and then click Run.

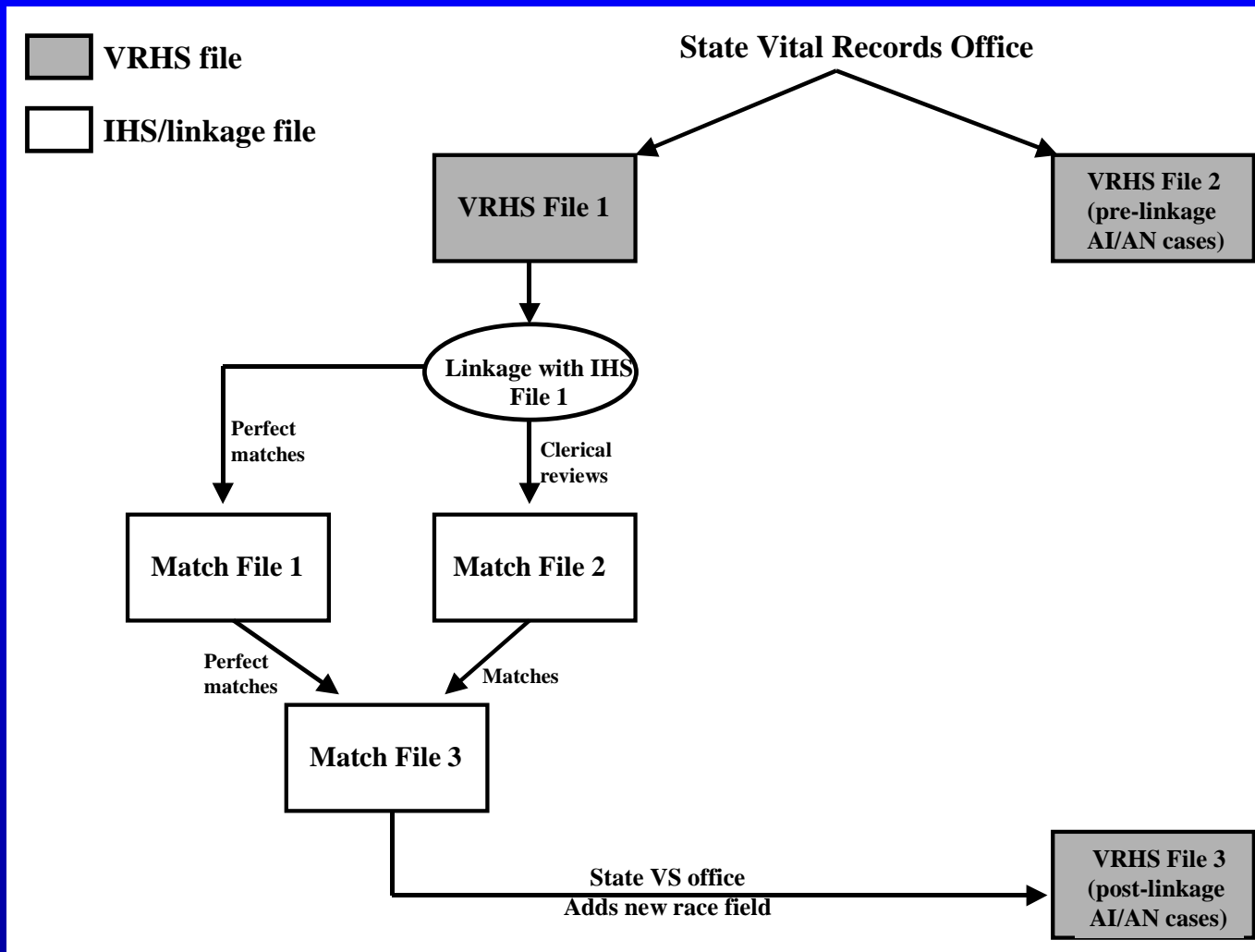
Direct Method

No
 Yes

Cutoff Value:

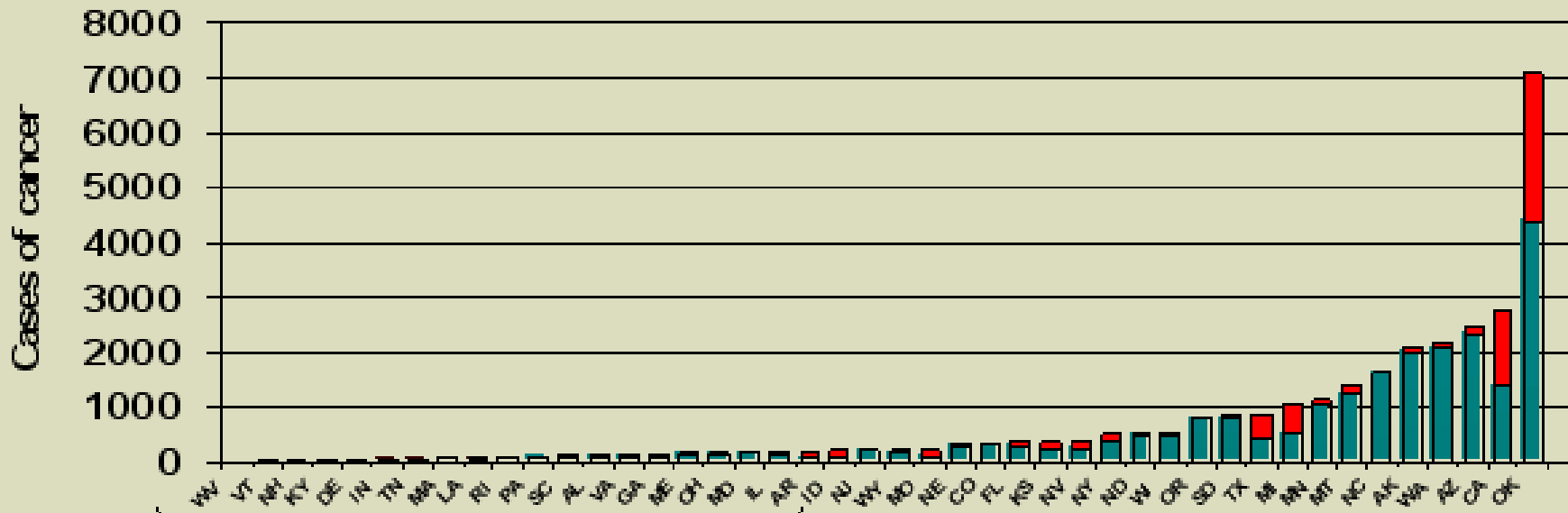


Data Flow

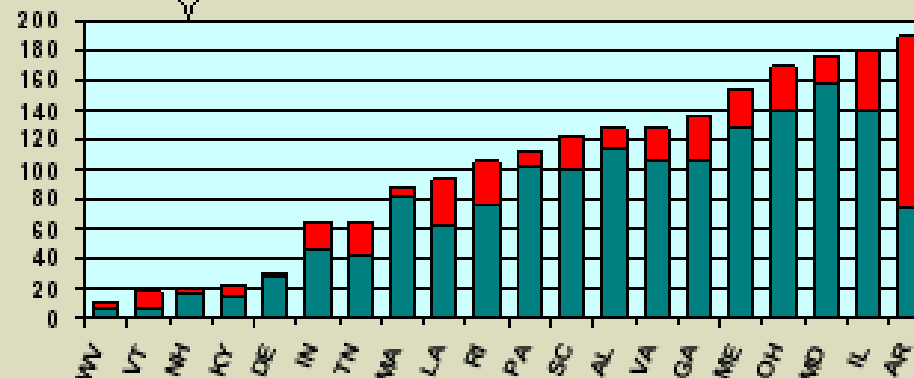


Record Linkage between NPCR Program registries and IHS administrative records

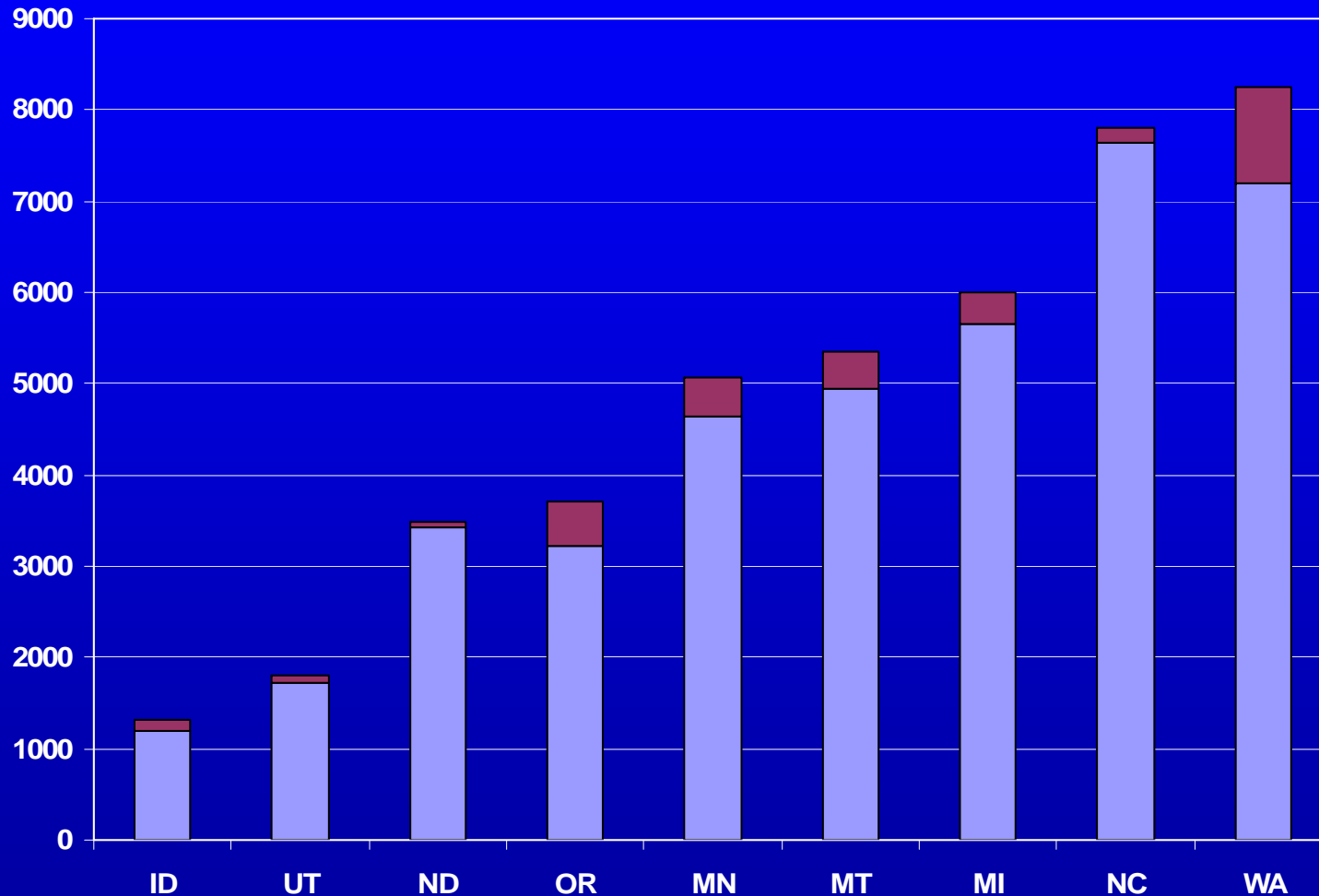
pre-link added by linkage



Graph with smaller scale for lower 20 NPCR Program registries



Preliminary VS-IHS Linkage Results



Other linkage activities in IHS

- IHS patient registration and SSA data
 - quarterly
- IHS patient registration and CMS data
 - quarterly
- IHS patient registration and Census data
 - annually

