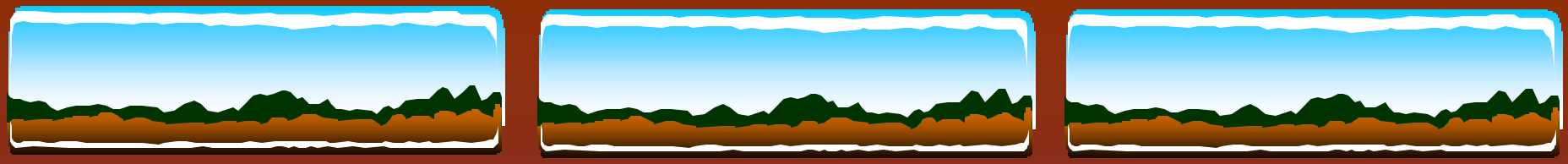


The Northwest Tribal Registry Project

Emily Puukka, MS
Northwest Tribal Epidemiology Center
Northwest Portland Area Indian Health Board

September 18, 2003

Overview

- History and goals of the Registry Project
- Racial Misclassification
- The Northwest Tribal Registry
- Record linkage Methodology
- Cancer Data Linkage Project

History of the Registry Project

- Downsizing of IHS has limited surveillance capacity
- Tribal leaders recognized a need for more complete and accurate health status data
    - Leading causes of death
    - Tribe and region-specific data
    - Small numbers make meaningful analysis difficult
- The Northwest Tribal Registry was created by the Northwest Tribal Epidemiology Center in 1999 with seed funding from IHS and NCI

Goals of the Registry Project

- Ascertain rates of racial misclassification in disease registries
- Obtain accurate estimates of disease burden for Northwest AI/AN
- Use linked case information to generate health status information for tribes and health programs
- Repeat linkages at regular intervals to build health trend data over time

Racial Misclassification

- Racial misclassification is the incorrect coding of an individuals' race or ethnicity
- For example, if an American Indian individual is coded as white in a disease registry
- Racial misclassification generally occurs in one "direction"
- Racial misclassification leads to an underestimation of the burden of disease in AI/AN populations

What Causes Racial Misclassification?

- Data on race or ethnicity come from different sources and are coded in different ways
- Staff may be hesitant to ask an individual about race – make assumptions based on appearance, name, residence, etc.
- Generally, race is more accurately recorded when self-reported

The Northwest Tribal Registry

- Demographic database comprised of AI/AN individuals who have received services from any Indian health care facility (except Seattle Indian Health Board) in the Northwest (OR, WA, or ID) since mid-1980's
- The Registry was created and is updated regularly from the Portland Area IHS *Area Patient File*
- Currently contains 152,620 cleaned and unduplicated records (AI/AN only)
- Records include personally identifying information for matching purposes, but do not include any health status or diagnostic data

The Northwest Tribal Registry

- Current Tribal Registry has been thoroughly cleaned and evaluated to ensure high accuracy
- Duplicate records are retained for matching process to increase probability of a true match
- The Registry is intended to be dynamic and responsive to tribal needs

Record Linkage Methodology

- There are two general types of matching – deterministic and probabilistic
- Deterministic matching works only when records from the two data files to be linked contain identical information with NO errors
- Probabilistic record linkage allows for matching between files even if:
  - Data are missing from some fields
  - There is variation in the way data are recorded for common fields

Record Linkage Methodology

- Employs a probabilistic linking software program called INTEGRITY©
- Produces very high accuracy in identifying AI/AN individuals who exist in both databases - even if race is misclassified or records do not match exactly
- Can process millions of possible match combinations in just a few minutes
- The linkage is conducted by running a series of “passes” – each allowing for errors in some fields, but not in others
- Assigns “weights” to suspected matched pairs based on fields defined by the user

Record Linkage Methodology

- The program runs frequencies of variables to determine likelihood of “chance” match
- For Example – Smith is more likely to appear in the two files than Puukka, thus the program sets a higher probability weight for Puukka matches than for Smith matches
- Allows for individual review of suspected matches
- Allows for recall of linked cases for epidemiological analysis
- No confidential or identifying data are permanently exchanged

Linkage Example

Blocking Variable (must match exactly in the two data files)

- SSN

Matching Variables (contribute to the probability weight assigned by the program based on user-specifications)

- Last name
- First name
- Middle initial
- Date of birth
- Sex

Considered a match:

PROB. WEIGHT	SSN	LAST
35.50	544-13-2323	Clark
35.50	544-13-2323	Clark

FIRST	MIDDLE	DOB	Sex
Jim	L	03/18/1950	M
James	L	08/18/1950	M

Considered a clerical review:

PROB. WEIGHT	SSN	LAST
15.50	544-13-2323	Clark
15.50	544-13-2323	Clarke

FIRST	MIDDLE	DOB	Sex
Jim		03/18/1950	M
James	L	08/18/1950	F

Cancer Data Linkage Project

- Goal was to link cancer patients in the state cancer registries with AI/AN individuals in the Northwest Tribal Registry
- Cancer registry data from 1996 – 1999 were obtained from the state registries
- Conducted a seven-pass run to identify matches in the two data files
- Clericals were reviewed manually to determine if they represented a true match

Linkage Results

Northwest Tribal Registry	136,635
AI/ANs listed in state registries 1996-1999 (prior to linkage)	738
State cancer registries and Tribal Registry matches*	895
Coded AI/AN in state registries but not in Tribal Registry	266

*These records comprised the analytic data file used in subsequent data analyses.

Linkage Results

- Of the 895 state cancer registries and Tribal Registry matches –
    - 472 were correctly coded as AI/AN in the state registries
    - 423 were misclassified as non-AI/AN
- The overall rate of misclassification was 36.4%
- All non-matched records were deleted for subsequent analyses – including individuals classified as AI/AN in the state registries, but not listed in the NTR
- We used the matched cases as numerators and the NTR as population denominators for calculating Northwest AI/AN cancer incidence rates

Leading Cancer Sites for NW AI/AN, by Number of Cases

Pre-Linkage	Post-Linkage
1. Breast	1. Breast
2. Tied: Lung/Bronchus	2. Lung/Bronchus
Tied: Colorectal	3. Colorectal
4. Leukemia/Lymphoma	4. Prostate
5. Prostate	5.Leukemia/Lymphoma

Percent Increase in Total Number of Cases for Leading Cancer Sites

Site	Percent Increase
Breast	78%
Lung/Bronchus	58%
Colorectal	52%
Prostate	149%
Leukemia/Lymphoma	55%

Cancer Rates* for Males

Cancer Site	Age-adjusted Rate
Prostate	77.9 (45.5, 110.3)
Lung/bronchus	42.9 (19.4, 66.3)
Colon-rectum	39.1 (16.6, 61.5)
All	278.5 (219.5, 337.6)

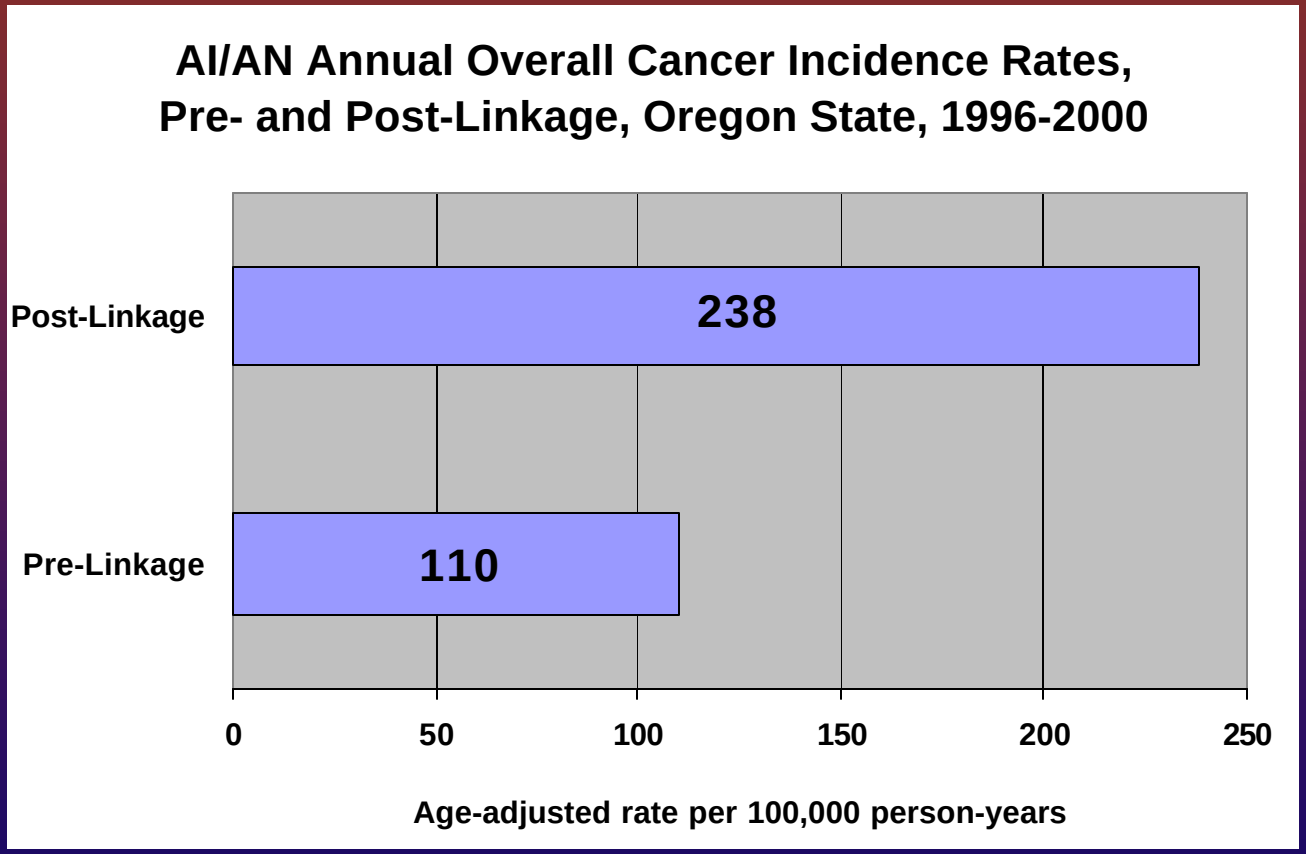
*Age-adjusted rate per 100,000 person-years (95% confidence interval)

Cancer Rates* for Females

Cancer Site	Age-adjusted Rate
Breast	87.4 (59.4, 115.5)
Lung/bronchus	34.2 (15.5, 52.9)
Colon-rectum	30.0 (13.0, 47.0)
All	264.7 (215.4, 314.0)

*Age-adjusted rate per 100,000 person-years (95% confidence interval)

The Effect of Linkage on Rates



Survival Analyses

- Idaho data only, Oregon and Washington do not currently collect survival data
- Mean survival (unadjusted) for all cancer sites – 6.8 years
- 5-year Survival (unadjusted) for all cancer sites – 45.4%

Key Findings

- The linkage revealed 423 additional cases of cancer among AI/ANs that the state registries had misclassified as non-AI/AN (36.4%)
- The overall annual cancer incidence rates for Northwest AI/AN approximately doubled the rates estimated using state registries' data alone (i.e., prior to the linkage)
- Survival rates are low for NW AI/AN

Limitations

- Only four years of cancer incidence data were available
- The NTR does not include all AI/ANs in the Northwest—most notably, it does not include AI/ANs receiving care in Seattle area
- For specific cancer sites, adequate data are not available for a thorough analysis

Summary

- Data linkage is an efficient and effective way to improve the accuracy of cancer rate estimates for AI/ANs
- Numbers of AI/AN cancer cases, and AI/AN cancer rates increased substantially after the linkage identified additional cases
- Annual linkages allow us to follow trends in cancer occurrence
- Intervention programs can be built around the results of improved data

Thank You!

Any Questions?