

Capture and Coding of Occupation and Industry Measures: Findings from 10 Central Cancer Registries

MaryBeth Freeman, MPH

Loria Pollack, MD, MPH

June 16, 2016

NAACCR Annual Conference

Objectives

- **Describe methods and quality of Occupation and Industry (O/I) collection practices**
 - As part of a special study involving 10 NPCR states
- **Overview of a tool to code O/I data**
- **Explore ability of the tool to assign O/I codes**

Background

- **Estimated annual impact of occupational exposure¹**
 - 40,000 new cancer cases and
 - 20,000 cancer deaths

Occupation(s)	Associated Cancer(s)
Asbestos fibers ²	Mesothelioma
Semi-conductor workers, dry cleaners, hairdressers, and mechanics ³	Acute myeloid leukemia
Firefighting ⁴	Colorectal, lung, non-Hodgkin lymphoma, and leukemia

1. National Institute for Occupational Safety and Health (NIOSH). <http://www.cdc.gov/niosh/programs/crcd>.

2. Roelofs et al. Analysis of Cancer Registry Data 1988-2003.

3. Tsai R, et al. Leuk Lymphoma. 2014.

4. Tsai R, et al. American J Industrial Med. 2015; 58: 715-729.

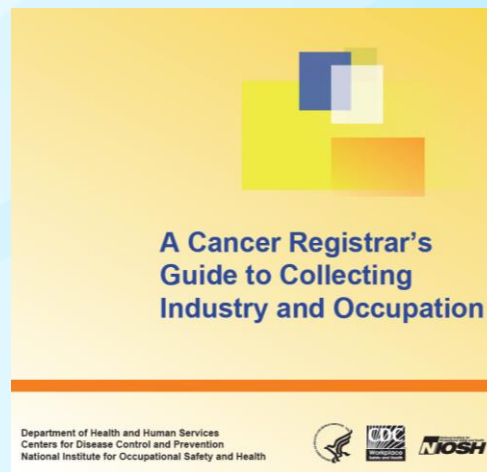


Cancer Registry Amendment Act

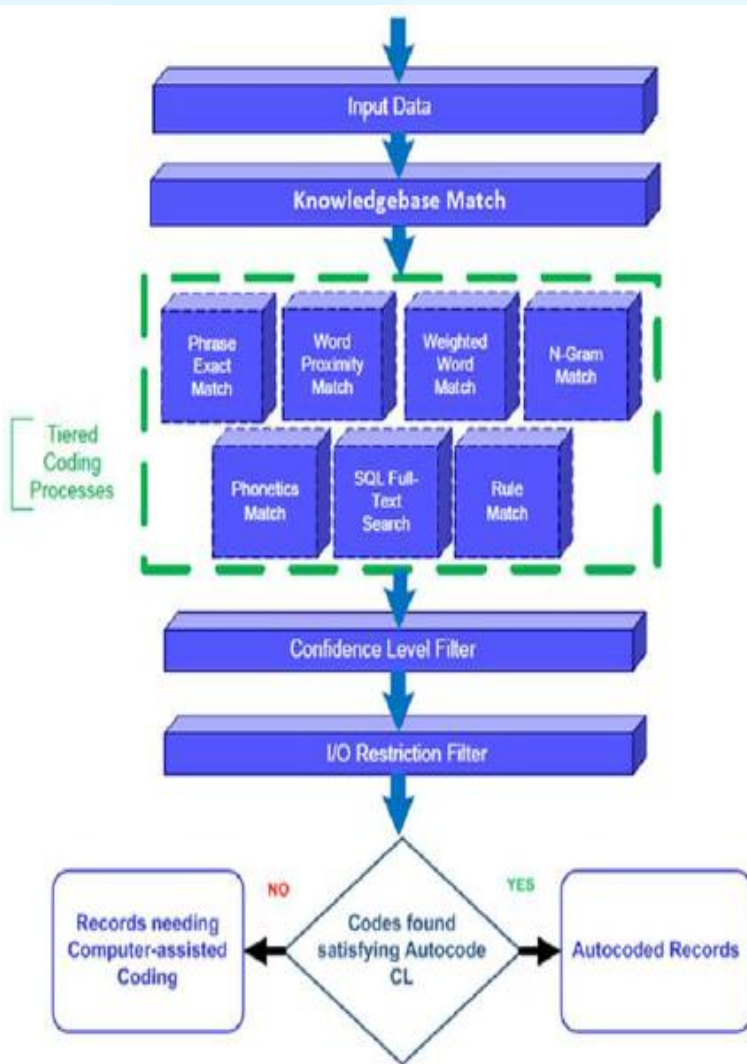
- Established National Program of Cancer Registries in 1992
- “To support the operation of population-based, statewide cancer registries...to collect...data concerning--
 - (1) demographic information about each case of cancer;
 - (2) ...industrial or occupational history of the individuals with the cancers, to the extent such information is available from the same record;**
 - (3) ...date of diagnosis and source of information;
 - (4) pathological data...”

Occupation and Industry Definitions

Patient's Usual Occupation	Patient's Usual Industry
Type of job patient engaged in for the greatest number of working years	Type of business or industry where patient worked in his or her usual occupation
Example: Registered Nurse Tour Guide	Example: Hospital Tourism



NIOSH Industry & Occupation Computerized Coding System (NIOCCS)



- **Web-based system that translates text into standardized O/I codes**
- **Goals**
 - Consistent terminology
 - Increased capture
 - Lower cost of manual coding

Coding of Cancer Data with NIOCCS Tool



- **NAACCR Data Dictionary recommends O/I coding at the CCR**

NPCR Special Study

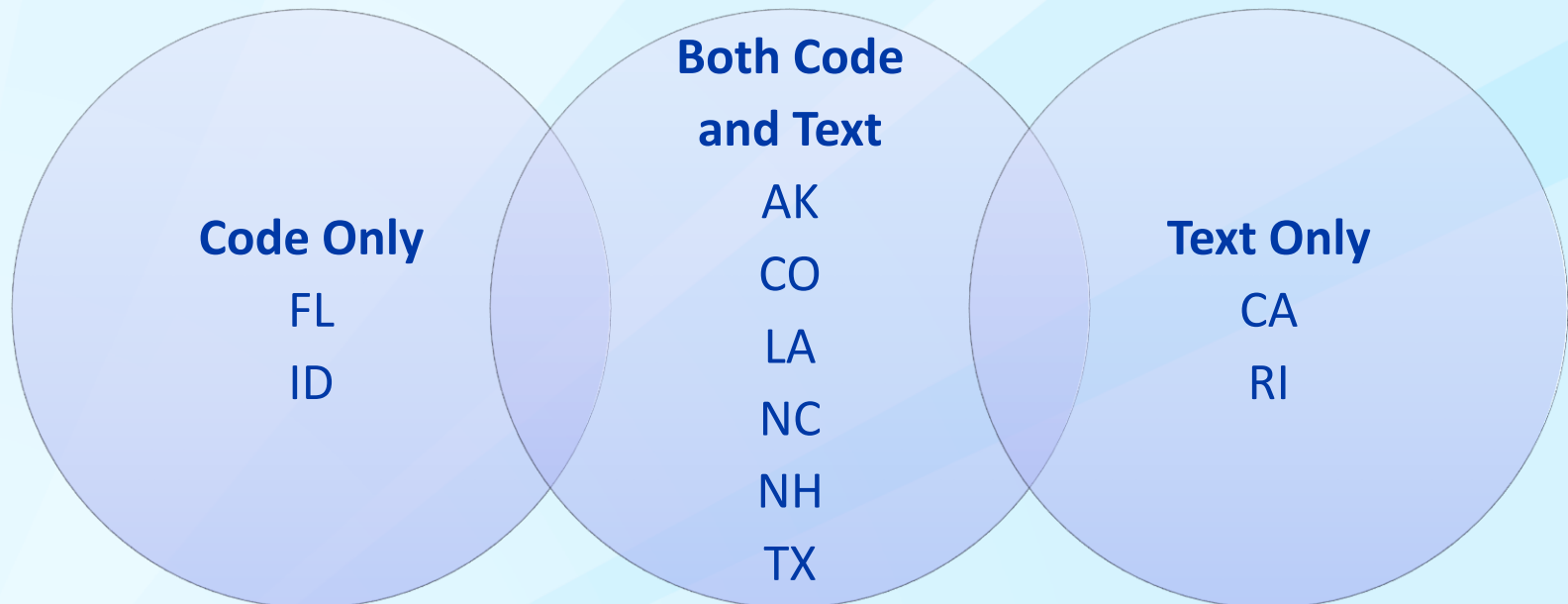
■ Comparative Effectiveness Research

- Supported by American Recovery and Reinvestment Act funds
- Expanded data collection in 10 registries (CCRs)
- 2011 incident cases: breast, colon, rectum, and CML
 - Biomarkers in diagnostic work-up
 - Beyond first course of treatment
 - SES, BMI, smoking
 - **Occupation and Industry**

■ CCRs were encouraged to recode textual O/I into census numeric codes, using NIOCCS

Data Sources

- NPCR CER Project – 8 complete states and 2 specified county groupings within California and Florida



- Census.gov – Census estimates of occupation and industry groups for 2011 were obtained for RI and NH

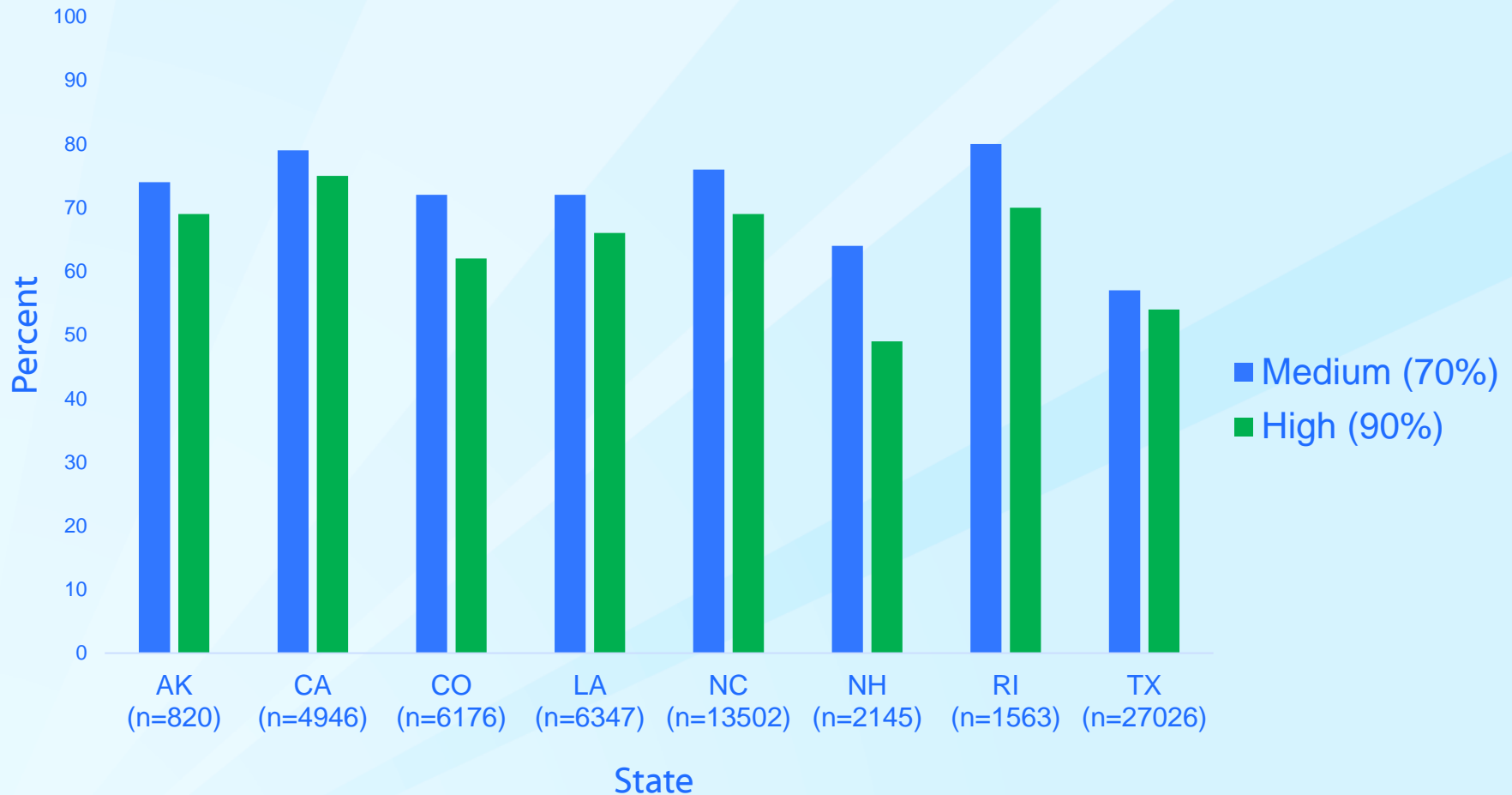
Methods

- **Available text data submitted to the NIOCCS tool (v3) at CDC**
- **Both medium (70%) and high (90%) NIOCCS confidence levels used**
 - Confidence: “only matched candidates where NIOCCS has _% or greater confidence of accuracy will be automatically coded”
- **SAS v. 9.3 (32) used to determine percentage of the NIOCCS auto-coded codes that were sufficiently analyzable O/I codes**
 - Insufficient codes: unknown, retired, never worked, military

Analysis

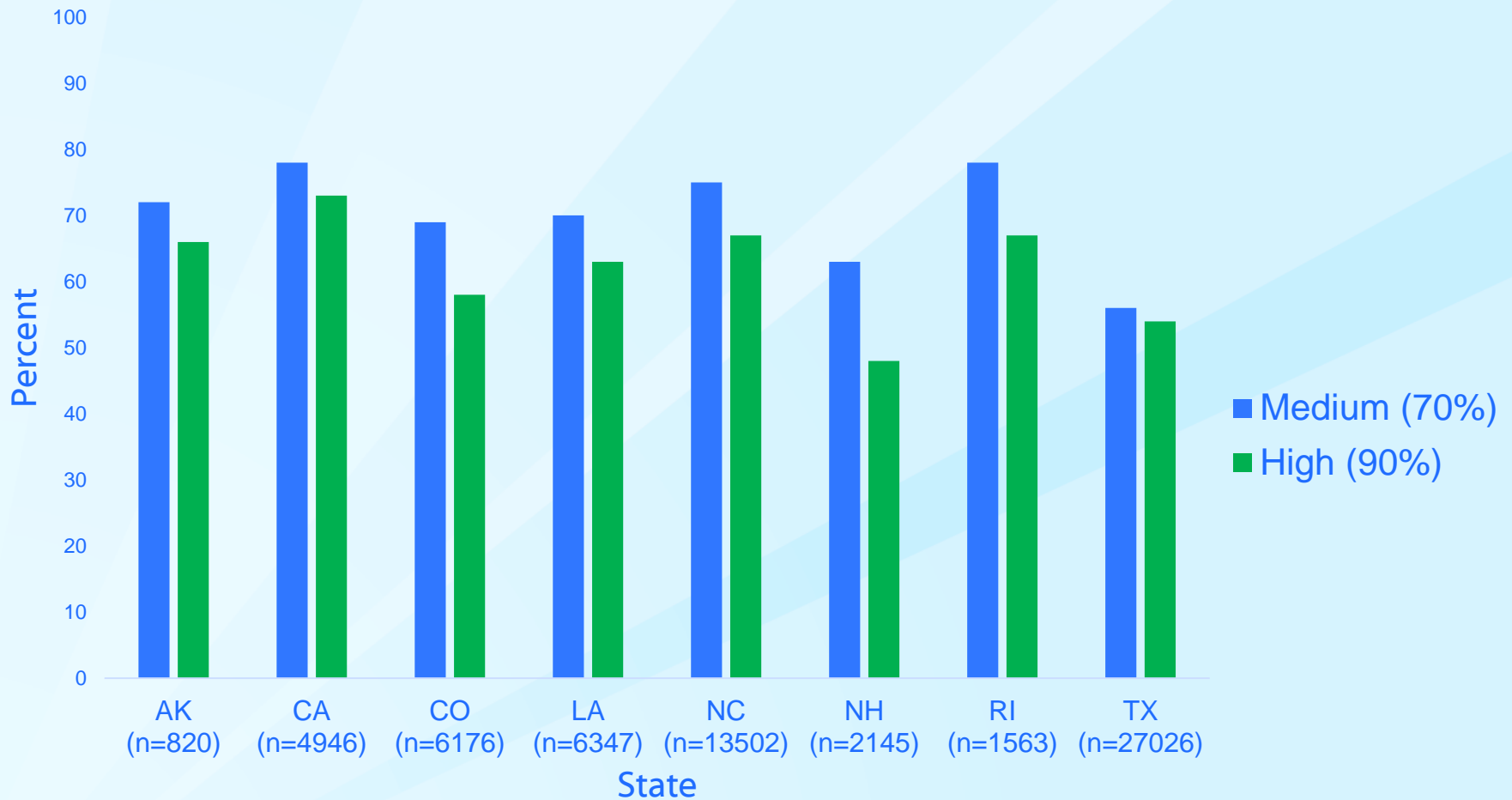
- **Assessment of O/I data assignment**
- **Comparison of auto-coding ability of the NIOCCS tool by confidence level and by state**
- **Examination of the quality of the data that was auto-coded by the NIOCCS tool**
 - Sufficient/Analyzable vs. insufficient (unknown, retired, never worked, military)
- **Comparison of O/I groups among two CER states with Census estimates for the same year**

Occupation text auto-coded by NIOCCS tool



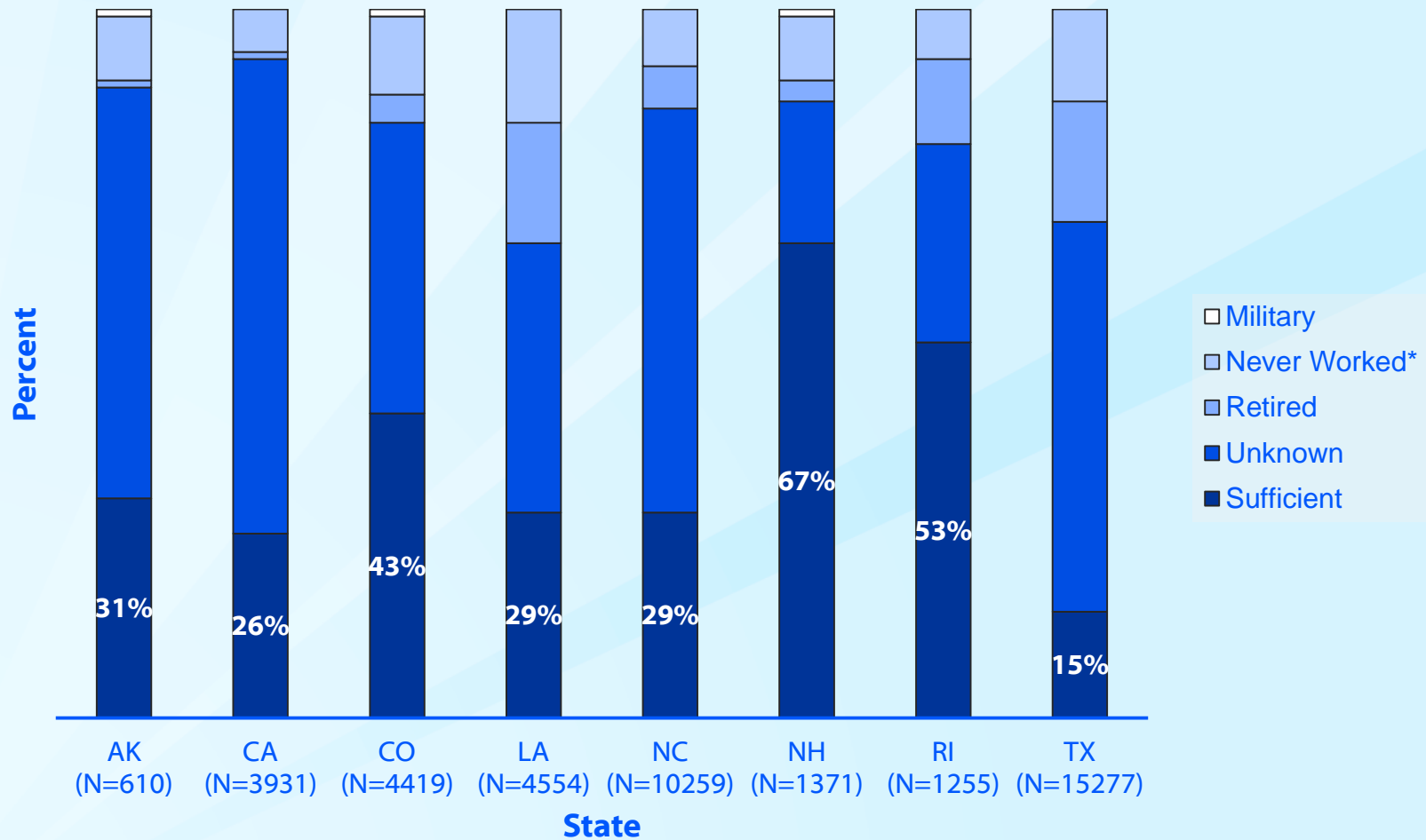
Note: 'Auto-coded' data includes codes for unknown, retired, never worked, military.

Industry text auto-coded by NIOCCS tool



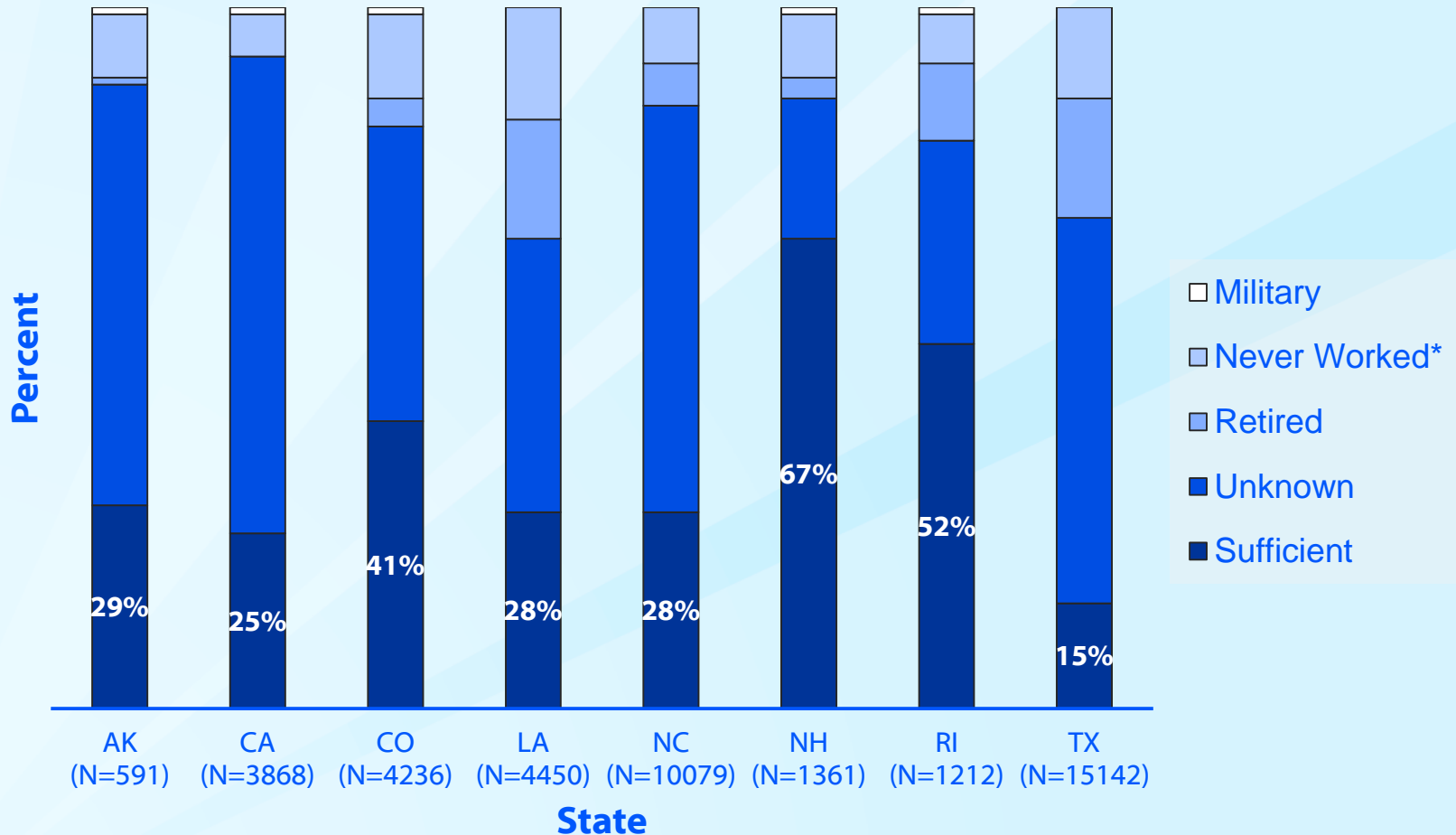
Note: 'Auto-coded' data includes codes for unknown, retired, never worked, military.

Types of Occupation data auto-coded by NIOCCS tool



*Includes homemaker, student, volunteer, never worked.

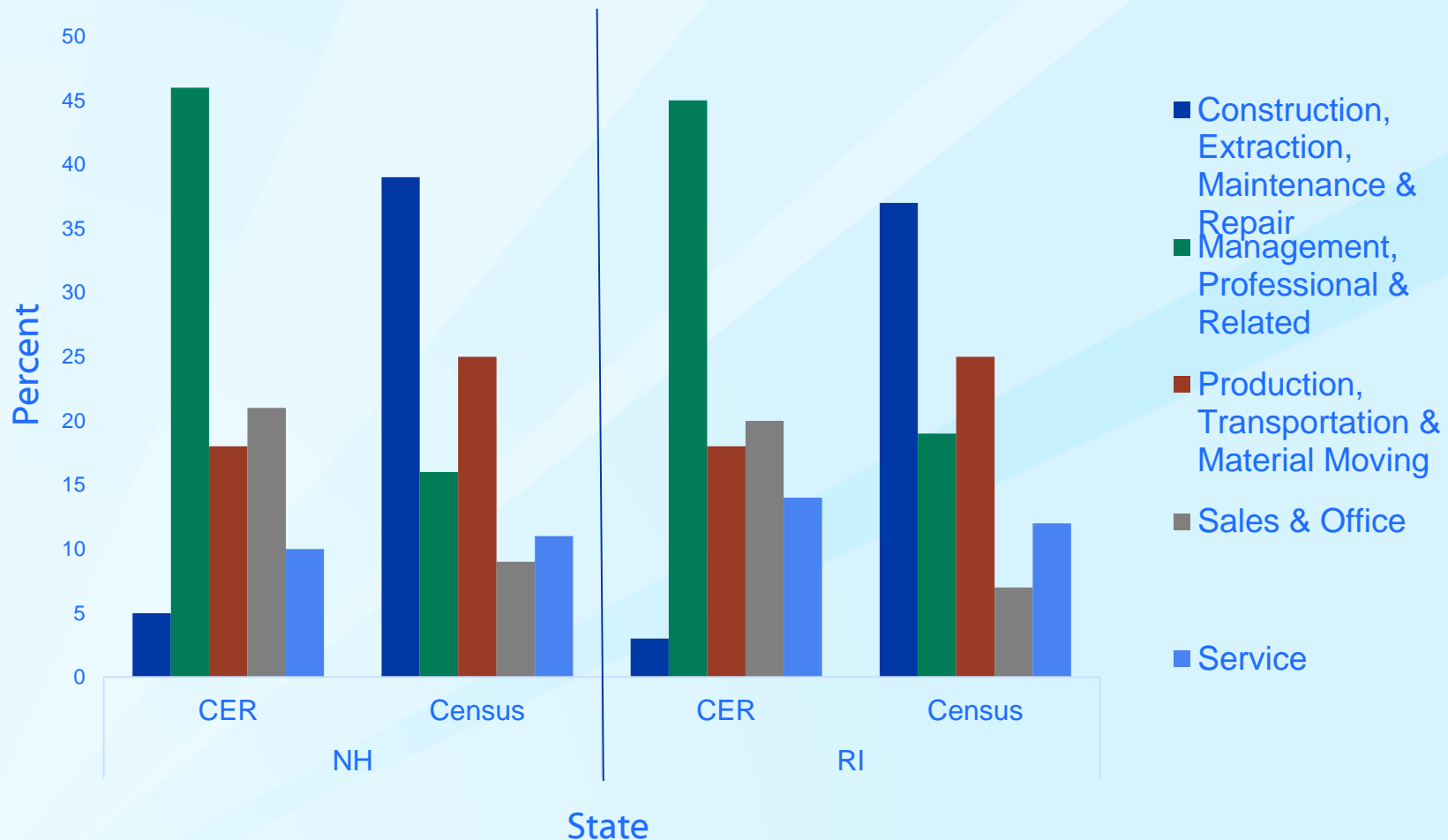
Types of Industry data auto-coded by NIOCCS tool



*Includes homemaker, student, volunteer, never worked.

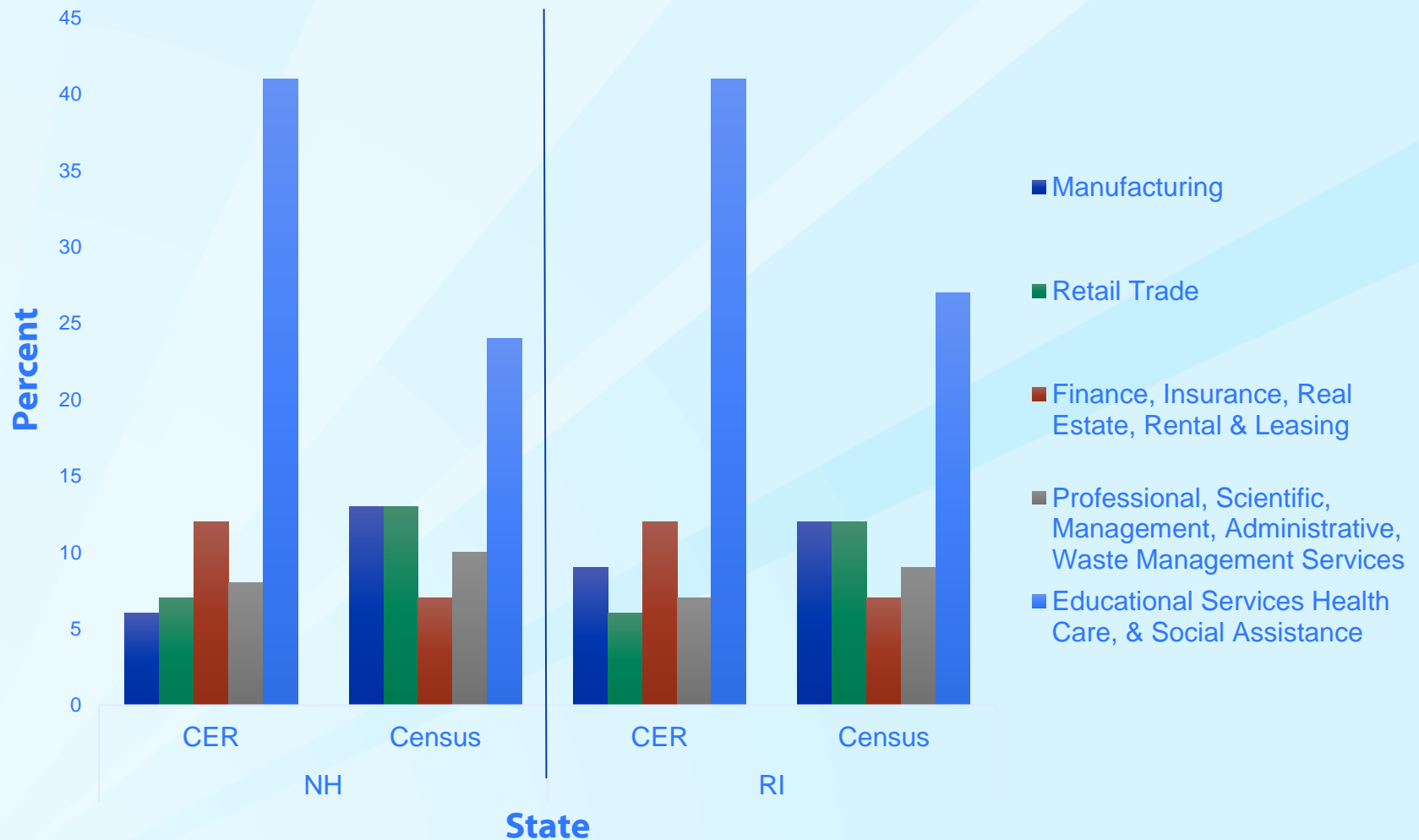
Occupational groups

CER cases compared to Census estimates



Top 6 Industry groups

CER cases compared to Census estimates



Findings

- **O/I for many cases was missing, unknown, or otherwise insufficient for analysis (33-85% of auto-coded cases)**
- **CODING: Occupational results mirror industry results**
- **Census comparison**
 - Mean ages different
 - Women > men in CER study
 - Latency periods between occupational exposures and cancer
 - Limited number of sites examined by the CER project (breast, colon, rectum, CML)

Use of NIOCCS Tool

- **Easy to use, but uploaded files require specific formatting**
- **NIOCCS tool confidence levels**
 - Medium (70%): more cases auto-coded, relatively less reliable codes
 - High (90%): fewer cases auto-coded, relatively more reliable codes
- **Computer-assisted coding available, but requires knowledge base of O/I assignment**

Reflections from Participating States

- **Hospital medical records often have insufficient documentation for O/I fields, especially in elderly patients that are now retired**
- **The coding still requires extensive manual review and manipulation**
- **Ongoing training needed for hospital registrars to collect better quality text information on O/I**

Implications

- **Minimum percentage of sufficient/analyzable O/I codes needed for analysis**
- **Collection of O/I data among central cancer registries is sub-optimal, but can be improved upon**
 - Study in New Hampshire: emphasized statewide training to highlight importance of O/I data, which improved O/I data quality across the state¹

Strengths

- **Examination of the utility and capabilities of the NIOCCS tool**
- **Identification of specific factors for consideration when analyzing O/I with cancer cases**
- **Comparison to Census estimates**

Limitations

- Limited to auto-coded data from the NIOCCS tool, manual coding was too time and labor intensive to complete at this time
- CER project focused on breast, colon, rectum, and CML diagnoses
- Only one year of data analyzed

Future Directions

- **Electronic Health Records – Natural Language Processing**
- **Linkage of Central Cancer Registries to Occupational Registries**

Acknowledgements

- **Participating NPCR Registries: AK, CA, CO, ID, FL, LA, NH, NC, RI, TX**
- **NIOSH: Sarah Luckhaupt, Marie Sweeney, Geoffrey Calvert**
- **CDC: Manxia Wu, Blythe Ryerson, Jane Henley, Reda Wilson**

Questions?

MaryBeth Freeman, MPH

yrj8@cdc.gov

770-488-7878

This project was supported in part by an appointment (MaryBeth Freeman) to the Research Participation Program at the Centers for Disease Control and Prevention administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the Centers for Disease Control and Prevention.

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.