

Industry and Occupation Coding of Cancer Records The Good, The Bad and The Ugly

Nancy Weiss, PhD, Sharon Cooper, PhD,
Anne Hakenewerth, PhD,

Christina Socias, DrPH(c), Ron Weiss,
Melanie Williams, PhD

Weiss Consulting, UT School of Public
Health, Texas Cancer Registry

Background

- Usual industry (I) & occupation (O) are National Program of Cancer Registries required reportable data items.
- Information is collected in text form.
- Few cancer registries have examined/ coded these text fields, limiting usefulness for research.
- Newly released software is available for auto-coding I & O text.

Background (continued)

NIOCCS Software

- National Institute for Occupational Safety and Health (NIOSH) Industry and Occupation Computerized Coding System (NIOCCS) - released December 2012
- Capable of automatic or computer-assisted coding; single record or batch records for auto-coding; crosswalk coding
- Has online Census I/O Alphabetical Index lookup capability

Background (continued)

NIOCCS Software

- Web-based
- Codes to 2000 or 2002 US Census Bureau classification scheme
- Three confidence levels for automatic coding assignments :

high	90% or greater accuracy
medium	70% or greater accuracy
low	30% or greater accuracy

Background (continued)

NIOCCS Software

- Records meeting user selected confidence levels automatically coded
- Records falling below selected level made available for manual processing in computer-assisted coding module
- Higher the level setting, higher the accuracy of coding results but lower the percent of records coded



NIOSH Industry and Occupation Computerized Coding System (NIOCCS)



[NIOSH](#) > [NIOCCS HOME](#)

[Coding](#)

[Help](#)

Welcome to NIOCCS

The NIOSH Industry & Occupation Computerized Coding System (NIOCCS) is a web-based system that translates industry and occupation (I&O) text into standardized I&O codes.

System Features Include:

- Automatic and Computer-Assisted coding
- Single Record or Batch File submission
- Crosswalk Coding

NIOCCS is available free of charge. Users must register for a NIOCCS account if they wish to submit a batch of records for coding.

To learn more about the NIOCCS software:

- See [NIOCCS Overview](#) Web Page
- Hover on the NIOCCS [Help](#) menu for support options -OR-
- Click [here](#) to contact NIOCCS User Support

NIOCCS User Support hours of operation:
Monday - Friday 8:00am - 5:00pm EST

[Login](#) to NIOCCS

[Register](#) for a NIOCCS Account



News / Updates

NIOCCS Enhancements 3/26/2013

1. Added a new menu item under the **Coding** menu to perform Census Index Searches.
2. Made some special rules corrections.
3. Added flags to the View All screen to indicate if the record was autocoded or not.
4. NAICS codes in the download file have been corrected. The crosswalk developed by Census was used, which can show more than one NAICS codes for a single Census code.

Project Purpose

- Assess the functionality of NIOCCS for coding industry & occupation text in cancer records reported to the Texas Cancer Registry (TCR)
- Make recommendations for implementing NIOCCS coding method into ongoing registry operations
- Project period - August 1, 2012 to August 1, 2013

Methods

- Obtained IRB approval from UTHSC-Houston
- Using WebPlus, downloaded de-identified TCR records for 1995-2011 (N=1,554,163)
- Data elements included unique id, diagnosis year, age, usual industry text , usual occupation text , county at diagnosis, & vital status
- Removed records with both missing I & O
- Split files into <1 MB files (N< 17,000 recs.)

Methods (continued)

- Coded to 2000 US Census Bureau scheme
- Prior to software release - identified & globally coded unknown, retired & common industry/occupation text (e.g. school, teacher; home, housewife)
- After software release, pilot tested 2000 diagnosis year cases (N=8,460) at high, medium & low level runs
- Ran subset of 1995-2009 records (not globally coded; N= 174,937) at high level

Methods (continued)

- Ran all 2010-2011 records (N= 146,993) at high level
- Manually coded records not coded at NIOCCS high level
- To assess coding agreement, compared random sample of 1,000 NIOCCS high-level coded 2010 records with manual coding
- Coded files were returned to the TCR via WebPlus

Results

- A total of 35 NIOCCS runs were made (31 high level, 3 medium level, 1 low level).
- NIOCCS run time varied by confidence level:

high	86 records/minute
medium	14 records/minute
low	54 records/minute

(Data runs completed in Jan/Feb 2013)
- File also may have wait time in run queue.

NIOCCS High Level Coding

Percent of records submitted and I/O coded

	<u>Industry</u>	<u>Occupation</u>
• 1995-1999	35.0	35.8
• 2000-2004	28.3	28.9
• 2005-2009	24.8	29.3
• 2010-2011	34.7	39.6

NIOCCS High Level Coding

Records with both I/O coded

	Number	Percent
• 1995-1999	4,804	1.2
• 2000-2004	6,596	1.4
• 2005-2009	32,326	6.2
• 2010	31,801	30.8
• 2011	15,044	22.8

Staff-Assisted Coding Methods

- Includes use of NIOCCS auto-assist coding module, project staff coding algorithms and manual code assignments:

	Number	Percent
• 1995-1999	388,122	97.2
• 2000-2004	442,735	95.9
• 2005-2009	410,255	78.3
• 2010-2011	Coding still underway	

Note: Records with both I/O coded

NIOCCS High-Level Coding Agreement (2010 Record Sample)

	<u>No.</u>	<u>Percent</u>
• Industry only (N=8)	3	37.5
• Occupation only (N=97)	62	63.9
• Both I/O coded (N=895)		
Both I/O agree	841	94.0
Industry agrees	890	99.4
Occupation agrees	882	98.5

Note:

46% of sample contains “unknown” or “retired” in text

Data Quality

- Large percent of records missing industry & occupation- 64.8% for 1995-2011
- Percent of records with missing text improving

1995-1999	92.4%
2000-2004	85.2%
2005-2009	42.5%
2010-2011	13.2%

The Good



The Good

- NIOCCS available free of charge
- Web-based and user friendly
- Online industry/occupation look-up with Census Alphabetical Index
- Crosswalk for coding schemes
- **NIOSH staff very responsive to feedback & continue to improve NIOCCS (e.g. coding efficiency/accuracy, run times & file handling)**
- Free training in I/O coding & use of NIOCCS

The Bad



The Bad

- Some wrong code assignments by NIOCCS at high level requiring manual correction
- Removal of records with blank O/I text, file splitting and remerging tedious
- Low percentages of auto-coding
- NIOCCS unable to recognize company names
- Manual review of records needed for medium and low -level code assignments

The Ugly



The Ugly (Really Ugly)

- Many records with missing I/O text
- “Unknown”/”NA” also in many records
- Poor quality of text provided - misspellings, abbreviations, acronyms, transpositions of I/O fields
- Poor data quality limited auto-coding
- Labor intensive to manually review/code
- Missing/insufficient data limit research use

Recommendations

- Have designated & trained staff in I/O coding even if only use NIOCCS auto-coding function
- Know the data quality of your existing records
- Know the industries in your state
- Include county of diagnosis to assist with industry searches (via the internet)

Recommendations (continued)

- Don't try coding 17 years of data in one year
- Start with more recent years and assess data quality before coding retrospectively
- Audit cancer reporters to see if I/O info is truly missing/unknown in medical record
- Monitor quality of incoming I/O text
- Provide on-going training of reporters on collection of I/O information

Recommendations (continued)

- With more NIOCCS users and data exposure, the better the software becomes
 - “purification through utilization”
- Go through IRB if release individual cancer records with I/O text for coding

Conclusions

- NIOCCS is a helpful tool for coding I & O text & continues to improve but other registry resources are required.
- Improvement in data quality of reported text information is needed to maximize the efficiency of NIOCCS & improve the availability of coded specific I & O information for occupational cancer research.

Future Thoughts

- For deceased patients, compare Vital Statistics death records' industry/ occupation coding with cancer records'
- Consider sharing coding resources/ expertise across programs
- Gain researcher and other support in improving the collection and reporting of I/O by the medical community

Acknowledgements

- CDC/NPCR & ICF International
- Pam Schumaker - pks1@cdc.gov
- Sue Nowlin- sxn1@cdc.gov
- Jeff Purdin
- Surveillance Branch, NIOSH

Team Contact Information:

NIOCCS@CDC.GOV

NIOCCS website: <http://wwwn.cdc.gov/niosh-nioccs/>

The End Happy Trails

