Combined T, N, and M From Directly Coded Clinical and Pathologic T, N, and M

Evaluation of a Derivation Algorithm and Opportunities for Registrar Education and Computer Edits

NIH NATIONAL CANCER INSTITUTE

Authors

- *Ward, K.*

  Georgia Cancer Registry, Atlanta GA
- *Ries, L.*

  National Cancer Institute (Contractor), Bethesda, MD
- *Ruhl, J.*

  National Cancer Institute, Bethesda, MD
- *Schussler, N.*

  Information Management Services, Calverton, MD

1. *Background*
2. *Algorithm Development & Modification*
3. *Opportunities for Education & Edits*
4. *Conclusions*

Background

- Since the beginning of national population-based cancer surveillance in the US, registries have been charged with helping to assess cancer patient outcomes at the population level

- Outcome assessment relies on the ability to accurately and completely capture important prognostic information for each patient across multiple data sources

Cancer Stage

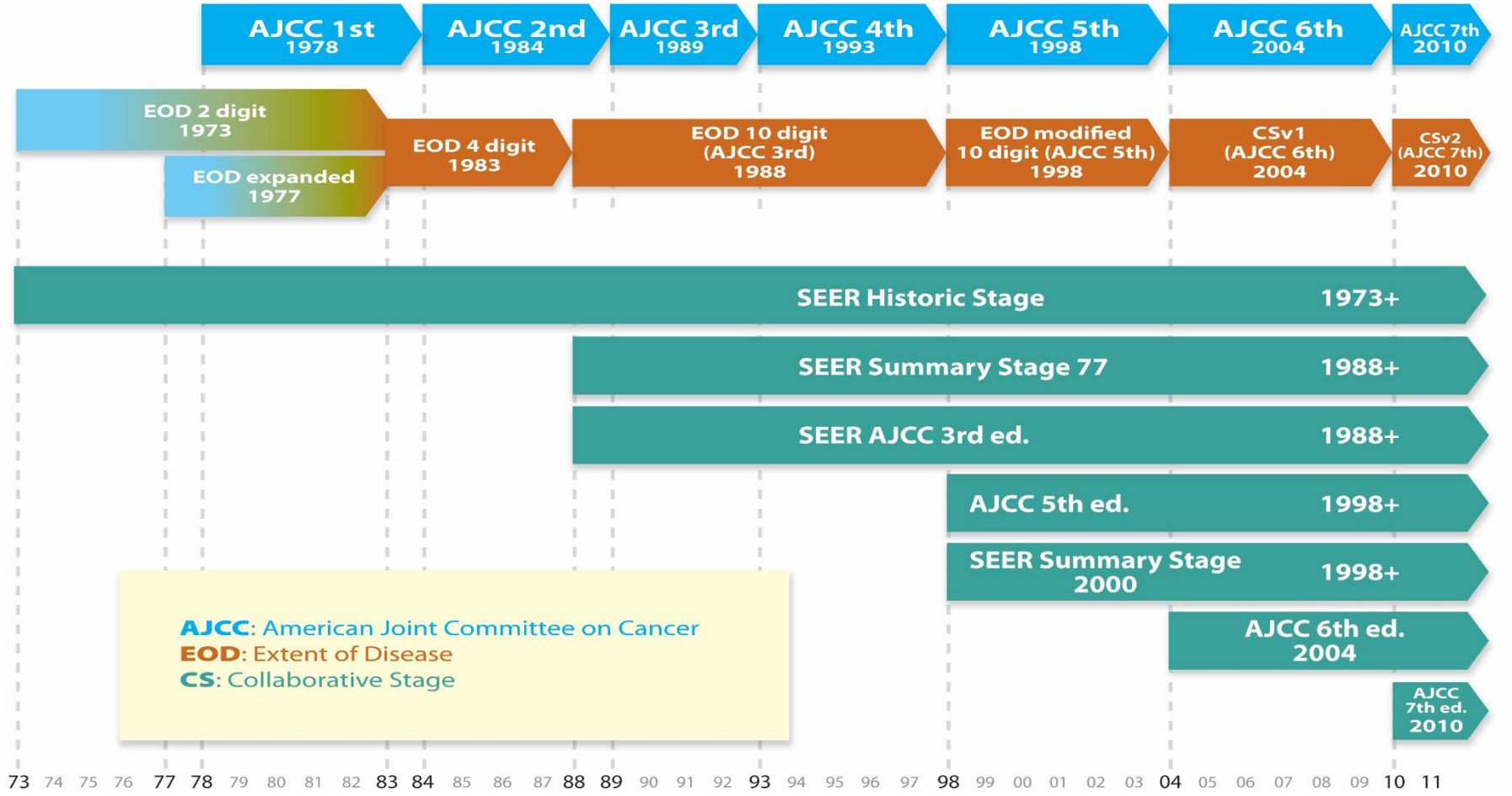
- ***Stage of disease*** is one of the strongest prognostic indicators of patient outcome

- Clinical information supplemented with pathologic data following surgery, where available, provides the ***most accurate assessment*** of disease involvement for each patient and results in the ***best predictor*** available in registries of ***cancer outcome***

Combined Stage

- Registries have historically, since their beginning, collected stage through ***combining clinical and pathologic*** information
  - Summary Stage
  - Extent of Disease
  - Collaborative Stage

Staging Timeline



2016 Data Collection

- Beginning in 2016, population-based cancer registries (PBCR) collect ***separate*** clinical and pathologic stage data (TNM) for each patient

- To ensure seamless utilization of stage data by consumers (researchers and others) through coordination with historical data, there is a need for PBCRs to develop methods to ***combine*** this separate clinical and pathologic information for prognostic purposes

Algorithm Development

- The NCI and Information Management Services (IMS) developed ***algorithms*** for combining the directly coded clinical TNM elements with the pathologic TNM elements

- Rules for combining elements mirrored those from SEER Extent of Disease and Collaborative Stage to the greatest degree possible

Algorithm Development, cont.

Input data included (where available)

- Directly coded clinical T, N & M
- Directly coded pathologic T, N & M
- Site & Histology (for schema determination)
- Behavior
- Regional nodes positive
- Treatment sequence (to determine if neoadjuvant therapy was provided)

Algorithm Development, cont.

- First, address blanks in any individual element

- Blanks converted to X for calculation purposes (with some exceptions)

Inputs		Temporary for Calculations	
Clin T	**Path T**	**tClin T**	**tPath T**
null	null	TX	TX
null	Tis	Path T	*copy*
null	*-any T*	TX	*copy*
-any-	null	*copy*	TX
-any-	*-any-*	*copy*	*copy*

Combined T

Combined T considered

- Schema*
- Neoadjuvant status
- TX and T0 combinations
- TX or T0 with other valid T values (Tis – T4)
- Two values, neither being TX or T0

*Separate tables were created for selected cancers to address specific issues (ex. Breast T4 cases – Inflammatory)

Example: Combined T

Directly coded cT2, pT1c

Without neoadjuvant therapy

Neo Status	Clin T	Path T	Combined T
…			
Non-neo	Any (not TX, T0)	TX, T0	Clin T
Non-neo	TX, T0	Any (not TX, T0)	Path T
Non-neo	Any (not TX, T0)	Any (not TX, T0)	Path T

Combined T = T1c (source = path)

Algorithm Development

Similar algorithms created for N & M (not shown)

Output data included

- Combined T
- Combined N
- Combined M

Algorithm rigorously tested internally to ensure working as desired

Also evaluated in a novel way using existing data

Algorithm Evaluation

Many PBCRs have historically collected directly coded clinical & pathologic T, N & M elements from CoC facilities in addition to Collaborative Stage (CS)

CS contains the elements

- Derived T
- Derived N
- Derived M

Elements were used for evaluating the algorithm acknowledging some differences existed in data collection methods and rules for TNM & CS

Algorithm Evaluation

Several SEER registries provided raw de-identified ***facility-based data*** in a special data submission (dx year 2014)

These data had not been consolidated across facilities

Each record contained

- Clinical & pathologic T, N & M (7th edition)
- CS derived T, N & M

Algorithm Evaluation

Five (5) sites chosen for initial analysis

- breast (5,364 cases)
- colon (1,603 cases)
- rectum (663 cases)
- lung (3,190 cases)
- prostate (1,932 cases)

Agreement & disagreement percentages calculated by comparing the algorithm's combined T, N, & M based on separate clinical & pathologic information to the comparable CS derived fields

Results

Overall agreement was ***very good*** with most categories having greater than 90% agreement

- Combined T – 87.8% lung to 96.1% colon
- Combined N – 84.7% rectum to 96.1% prostate
- Combined M – 89.9% lung to 99.7% breast

Results, cont.

Site		Algorithm Percent (%)	
		Agree	Disagree
Breast	T	95.1	4.9
	N	87.9	12.1
	M	99.7	0.3
Colon	T	96.1	3.9
	N	92.1	7.9
	M	94.4	5.6
Rectum	T	92.3	7.7
	N	84.7	15.3
	M	95.4	4.6

Site		Algorithm Percent (%)	
		Agree	Disagree
Lung	T	87.8	12.2
	N	94.5	5.5
	M	89.9	10.1
Prostate	T	91.3	8.7
	N	96.1	3.9
	M	97.7	2.3

Algorithm Modification

Group examined every case that did not agree to understand why & see if the algorithm could be improved

A ***few minor changes*** were made

Example

For lung N, original algorithm took the pathologic value (if one existed) over the clinical value. In reviewing the data, N3 nodes (contralateral) were usually not removed with the primary. Algorithm was modified to take clinical information on positive N3 nodes over pathologic information.

Opportunities for Education

Missing subgroup in directly coded data

- Breast N discrepancies
  - Path N subgroup missing for N1, N2, or N3 or
  - Only N0 listed, but positive results from IHC/Mol test

- Over 70% of the discrepant cases for lung M were missing the appropriate subgroup

- **Education in conjunction with specific rules and pick lists will help with these issues as well.**

Opportunities for Edits

- Some cases had regional nodes coded positive but the directly coded path N was pN0
- Other cases had regional nodes coded negative but the directly coded path N was pN1
  - **Edits will help with these types of issues**

Inconsistencies

Disagreement due to inconsistent information collected in the directly coded TNM fields compared to what was captured through CS (including missing information on directly coded T N & M)

Example

- Breast clinical & path T both = T1a
- CS derived T = T1b

Will be important to rerun the algorithm on more current data (2015/2016) to investigate these further

Conclusions

- Algorithm seems to be ***working well***
- Registrar education needs to emphasize the importance of coding the explicit ***subcategory*** for the c & p T, N, & M
- Pick lists & edits will improve the data quality & facilitate coordination with historical combined stage data

Thank You!

Questions?

NATIONAL CANCER INSTITUTE

www.cancer.gov **www.cancer.gov/espanol**