

Development of NPCR SAS Tool for Population-Based Cancer Relative Survival Analysis

June 2015

By X Dong, K Zhang,
R Wilson, Y Ren

ACKNOWLEDGEMENT

- The study was supported by CDC under the contract 200-2008-27957

Outline

- Introduction
- Data preparation
- Methods and Processes
- Results and discussions
- Conclusions

Introduction

- **Relative survival is widely used to measure cancer survival for population-based cancer surveillance systems**
 - Observed survival of a group of cancer patients
 - Expected survival of a comparable group from general population
 - Ratio of observed survival and expected survival

Introduction

□ Background of NPCR SAS tools for relative survival analysis

- It is a by-product of NPCR's relative survival analysis production for USCS, an extension of our work in preparing NPCR data for relative survival analysis using SEER*Stat
- The SAS tool turned out to be a good QC tool for validating results produced by SEER*Stat
- It can also benefit other researchers who would like to take the advantage of the SAS flexibility to conduct customized relative survival analysis

Data Preparation

□ November 2012 NPCR submission and major inclusion criteria

- *Diagnosis year 2003-2010 with follow-up to 2010*
- *25 states based on NDI linkage and data quality*
- *Age at diagnosis between 0-99*
- *Malignant cases only with coding rules 2001+ (ICD-O-3)*
- *No death certificate or autopsy only cases*
- *Missing month and day date of diagnosis and date of last contact were imputed with a SEER date imputation SAS program*

Data Preparation

□ NDI linkage and presumed-alive

- NPCR states are not funded for active follow-up
- To obtain more complete death information, CDC funds NPCR states to conduct NDI linkage at least every other year. Not all states participate
- Cases not known to be deceased by the NDI linkages will be assumed alive and censored at the end of study period (presumed-alive)
- Total cases included: $n = 6,383,241$

Data Preparation

□ Life tables for expected survival

- Used NCHS U.S. 2000-2009 life tables by individual year (White and Black), Ages 0-99 (updated in 2014)
- For patients having missing and unknown sex and/or race information, the total race table was used
- Omitted case whose follow-up age was greater than 100 due to life table age limitation

Methods and Processes

☐ Survival estimation

- Observed survival: Actuarial life table method
 - Standard error is given by Greenwood's formula
- Expected survival: Ederer II method
- Relative survival = observed survival/Expected survival
 - Standard error is given as the standard error of observed survival divided by expected survival rate

Methods and Processes

□ Estimation process

- The core survival estimation process: Paul Dickman's SAS Macro
- Reprogramming were made to facilitate analysis
 - Estimate relative survival for multiple cancer sites
 - Estimate relative survival by predefined categorization (e.g., by sex, age groups or combinations)
- Additions to make NPCR tool comparable to SEER*Stat
 - Implemented case selection options similar to those of SEER*Stat (e.g., primary site)
 - Allowed adjustment of abnormal relative survival patterns

Methods and Process

□ NPCR SAS tool

- Three basic modules
 1. Data reprocessing/date imputation for incidence data
 2. Relative survival estimation and reporting
 3. SEER*Stat database creation (life table and incidence data)
- Implemented with SAS 9.2 with batch processing
- For large compressed SEER*Stat databases, 7-Zip command line called by SAS 9.2 is required to create GZIP files

Results and Discussions

□ Validation Analysis

- NPCR SAS tool and SEER*Stat used the same set of NPCR incidence data and NCHS life tables
- Relative survival rates and standard errors were estimated by race, gender and age group of 0-44, 45-54, 55-64, 65-75, and 75-99
- 24 cancer sites as USCS
- All-sites-combined used the first primary only; other sites allowed multiple primaries
- Comparison analysis on RS rate for 6 major sites:
Standard errors on 1 cancer site

Results and Discussions

□ Results: All-sites-combined

DX Age	S1-NPCRSASTool				S2-SEER* Stat				Net S1 and S2			
	Whites		Blacks		Whites		Blacks		Whites		Blacks	
	M	F	M	F	M	F	M	F	M	F	M	F
Ages <45	78.1	85.5	63.5	74.5	78.1	85.5	63.5	74.5	0.0	0.0	0.0	0.0
Ages 45-54	67.5	78.6	60.6	65.3	67.5	78.6	60.6	65.3	0.0	0.0	0.0	0.0
Ages 55-64	68.9	70.5	65.8	59.1	68.9	70.5	65.8	59.1	0.0	0.0	0.0	0.0
Ages 65-74	66.9	61.6	66.3	52.5	66.9	61.6	66.3	52.5	0.0	0.0	0.0	0.0
Ages 75+	55.5	50.2	51.5	41.1	55.5	50.3	51.5	41.2	0.0	-0.1	0.0	-0.1

Results and Discussions

□ Results: Colon and Rectum

DX Age	S1-NPCRSASATool				S2-SEER* Stat				Net S1 and S2			
	Whites		Blacks		Whites		Blacks		Whites		Blacks	
	M	F	M	F	M	F	M	F	M	F	M	F
Ages <45	67.3	71.8	61.0	63.8	67.3	71.8	61.0	63.8	0.0	0.0	0.0	0.0
Ages 45-54	69.8	73.1	61.5	66.7	69.8	73.0	61.5	66.7	0.0	0.1	0.0	0.0
Ages 55-64	68.5	70.1	60.6	64.7	68.5	70.1	60.6	64.7	0.0	0.0	0.0	0.0
Ages 65-74	66.1	68.1	57.0	61.9	66.1	68.0	56.9	61.9	0.0	0.1	0.1	0.0
Ages 75+	59.5	60.1	47.2	50.4	59.5	60.2	47.3	50.4	0.0	-0.1	-0.1	0.0

Results and Discussions

□ Results: Lung and bronchus

DX Age	S1-NPCR SASTool				S2-SEER* Stat				Net S1 and S2			
	Whites		Blacks		Whites		Blacks		Whites		Blacks	
	M	F	M	F	M	F	M	F	M	F	M	F
Ages <45	25.6	33.3	19.2	27.6	25.6	33.3	19.2	27.6	0.0	0.0	0.0	0.0
Ages 45-54	18.7	27.4	15.0	22.2	18.7	27.4	15.0	22.2	0.0	0.0	0.0	0.0
Ages 55-64	18.0	25.6	14.8	21.0	18.0	25.6	14.8	21.1	0.0	0.0	0.0	-0.1
Ages 65-74	17.3	23.3	14.0	19.4	17.3	23.3	14.1	19.4	0.0	0.0	-0.1	0.0
Ages 75+	13.2	16.6	10.1	14.4	13.2	16.7	10.2	14.5	0.0	-0.1	-0.1	-0.1

Results and Discussions

□ Results: Breast (female)

DX Age	S1-NPCRSASTool				S2-SEER* Stat				Net S1 and S2			
	Whites		Blacks		Whites		Blacks		Whites		Blacks	
	M	F	M	F	M	F	M	F	M	F	M	F
Ages <45	+	89.3	+	78.9	+	89.3	+	78.9	+	0.0	+	0.0
Ages 45-54	+	90.9	+	80.4	+	90.9	+	80.3	+	0.0	+	0.1
Ages 55-64	+	90.4	+	80.6	+	90.4	+	80.5	+	0.0	+	0.1
Ages 65-74	+	90.7	+	81.8	+	90.7	+	81.8	+	0.0	+	0.0
Ages 75+	+	88.6	+	77.4	+	88.6	+	77.4	+	0.0	+	0.0

Results and Discussions

□ Results: Prostate

DX Age	S1-NPCR SASTool				S2-SEER* Stat				Net S1 and S2			
	Whites		Blacks		Whites		Blacks		Whites		Blacks	
	M	F	M	F	M	F	M	F	M	F	M	F
Ages <45	95.2	+	96.5	+	95.2	+	96.5	+	0.0	+	0.0	+
Ages 45-54	98.0	+	97.5	+	98.0	+	97.5	+	0.0	+	0.0	+
Ages 55-64	99.3	+	97.9	+	99.3	+	97.9	+	0.0	+	0.0	+
Ages 65-74	99.8	+	97.1	+	99.8	+	97.1	+	0.0	+	0.0	+
Ages 75+	94.4	+	87.3	+	94.5	+	87.3	+	-0.1	+	0.0	+

Results and Discussions

□ Results: Urinary bladder

DX Age	S1-NPCR SASTool				S2-SEER* Stat				Net S1 and S2			
	Whites		Blacks		Whites		Blacks		Whites		Blacks	
	M	F	M	F	M	F	M	F	M	F	M	F
Ages <45	88.5	87.8	78.4	60.3	88.5	87.8	78.4	60.3	0.0	0.0	0.0	0.0
Ages 45-54	84.7	84.4	75.8	60.0	84.7	84.4	75.8	60.0	0.0	0.0	0.0	0.0
Ages 55-64	82.6	82.3	72.0	64.1	82.6	82.4	72.0	64.1	0.0	-0.1	0.0	0.0
Ages 65-74	79.1	77.4	70.2	59.5	79.1	77.4	70.1	59.5	0.0	0.0	0.1	0.0
Ages 75+	71.8	65.7	59.5	49.8	71.8	65.8	59.4	49.9	0.0	-0.1	0.1	-0.1

Results and Discussions

□ Results: Urinary bladder (standard error)

DX Age	S1-NPCR SASTool				S2-SEER* Stat				Net S1 and S2			
	Whites		Blacks		Whites		Blacks		Whites		Blacks	
	M	F	M	F	M	F	M	F	M	F	M	F
Ages <45	0.6	0.9	2.6	4.9	0.6	0.9	2.6	4.9	0.0	0.0	0.0	0.0
Ages 45-54	0.4	0.6	1.6	2.7	0.4	0.6	1.6	2.7	0.0	0.0	0.0	0.0
Ages 55-64	0.3	0.5	1.4	2.0	0.3	0.5	1.4	2.0	0.0	0.0	0.0	0.0
Ages 65-74	0.3	0.5	1.4	1.8	0.3	0.5	1.4	1.8	0.0	0.0	0.0	0.0
Ages 75+	0.4	0.5	1.9	1.8	0.4	0.5	1.8	1.8	0.0	0.0	0.1	0.0

Conclusions

□ Conclusions

- The results from the NPCR SAS tool and SEER*Stat were almost identical in both relative survival rates and standard errors
- The NPCR SAS tool can be used for the state level relative survival analysis if the state level life tables are available
- If SEER*Stat life tables are applied, NPCR SAS tool could be used seamlessly along with SEER*Stat but with more flexibilities and customization for advanced analysis.

Thank You!

Questions?

Contact Information

Xing Dong xdong@cif.com

Reda Wilson dfo8@cdc.gov

NPCR <http://www.cdc.gov/cancer/npcr/index.htm>

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.