

Imputing Missing Estrogen Receptor Status from Population-based SEER Cancer Registries

Nadia Howlader
Mathematical Statistician, MS
Surveillance Research Program, NCI

NAACCR Cancer Surveillance Webinar Series

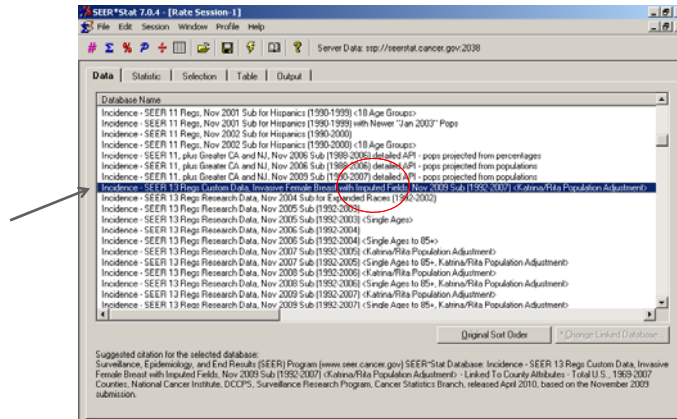


May 20, 2016

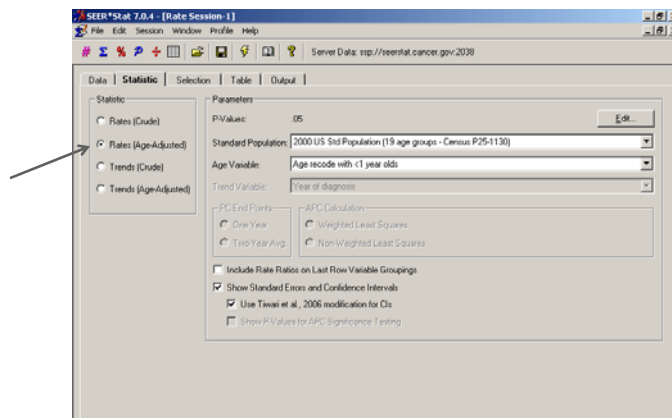
Variables Available in Female Breast Cancer Imputed Dataset: SEER-13, 1992-2012

Demographic Variables	Clinical Variables
Age at diagnosis	Node positive status
Year of diagnosis	Metastasis at diagnosis
Registry	ER status
Race	PR Status
Ethnicity	Histology
Poverty (county level)	Tumor Grade
Urbanacity (county level)	Tumor Size

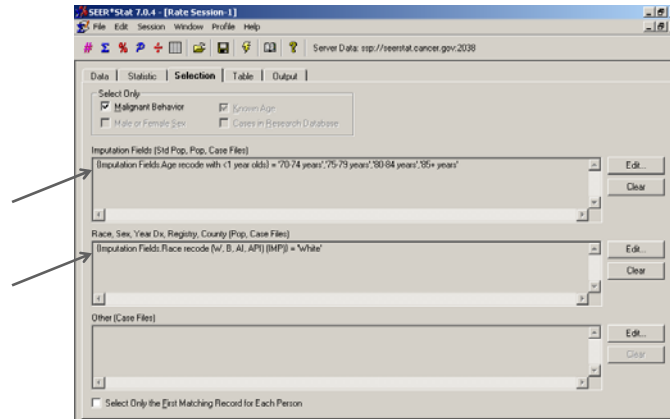
SEER*Stat: Select Database



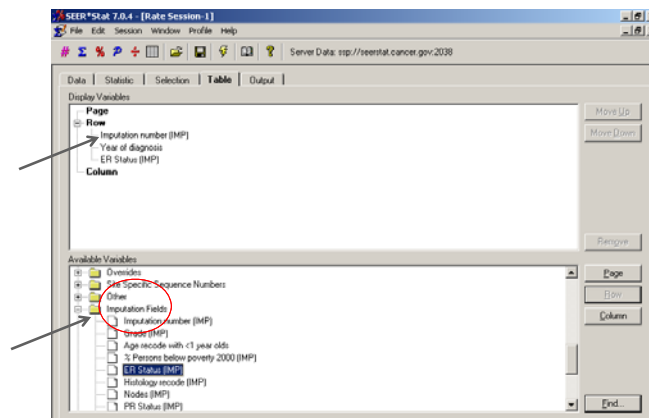
SEER*Stat: Choose Statistics to Display



SEER*Stat: Selection of Cohort



SEER*Stat: Table Variables



SEER*Stat: Output

		Rate	SE	Lower CI	Upper CI	Count	Pop	
Imputation 1	2004	Positive	364.7	4.9	355.1	374.4	5,981	1,565,135
		Negative	69.0	2.2	64.9	73.4	1,055	1,565,135
		Unknown	0.0	-	0.0	0.2	0	1,565,135
	2005	Positive	374.6	5.0	364.8	384.5	5,791	1,569,010
		Negative	65.7	2.1	61.7	69.9	1,015	1,569,010
		Unknown	0.0	-	0.0	0.2	0	1,569,010
	2006	Positive	359.3	4.9	349.8	369.0	5,503	1,572,032
		Negative	62.0	2.1	59.8	66.9	972	1,572,032
		Unknown	0.0	-	0.0	0.2	0	1,572,032
2007	Positive	362.6	4.9	353.0	372.4	5,617	1,578,713	
	Negative	62.7	2.1	58.7	68.9	970	1,578,713	
	Unknown	0.0	-	0.0	0.2	0	1,578,713	
Imputation 2	1992-2007	Positive	398.2	1.3	393.7	398.7	96,818	24,621,318
		Negative	72.3	0.5	71.3	73.4	17,637	24,621,318
		Unknown	0.0	-	0.0	0.0	0	24,621,318
	1992	Positive	404.0	5.3	393.6	414.5	5,805	1,440,173
		Negative	75.1	2.3	70.6	79.7	1,070	1,440,173
		Unknown	0.0	-	0.0	0.3	0	1,440,173
	1993	Positive	391.4	5.2	381.3	401.7	5,726	1,465,670
		Negative	62.7	2.4	70.1	67.5	1,206	1,465,670
		Unknown	0.0	-	0.0	0.2	0	1,465,670
	1994	Positive	389.8	5.1	379.8	400.0	5,774	1,485,982
		Negative	77.5	2.3	73.1	82.1	1,146	1,485,982
		Unknown	0.0	-	0.0	0.3	0	1,485,982
	1995	Positive	406.3	5.2	396.2	416.6	6,107	1,508,983
		Negative						
		Unknown						

Rates are per 100,000 and age-adjusted to the 2000 US Std Population (19 age groups - Census P25-1130) standard. Confidence intervals (Tiwari mod) are 95%.
Statistic could not be calculated.

SAS Code

```

/* Sort data by stratification variables used in BY statement*/
proc sort data=ratedat2;
  by year_of_diagnosis er_status_imp;
run;

proc mianalyze data=ratedat2;
  by year_of_diagnosis er_status_imp;
  model effects age_adjusted_rate;
  stderr standard_error;
run;

```

SAS Output

Yea:

The MIANALYZE Procedure
Verify that 5 imputations were used in the calculation

Model Information

Data Set WORK.RATEDAT2
Number of Imputations 5

Parameter Estimates

Parameter	Estimate	Std Error	95% Confidence Limits	
age_adjusted_rate	396.51	1.32	393.92	399.10

Verify that 5 imputations were used in the calculation

Imputed age-adjusted incidence rate and standard error



**NATIONAL
CANCER
INSTITUTE**