

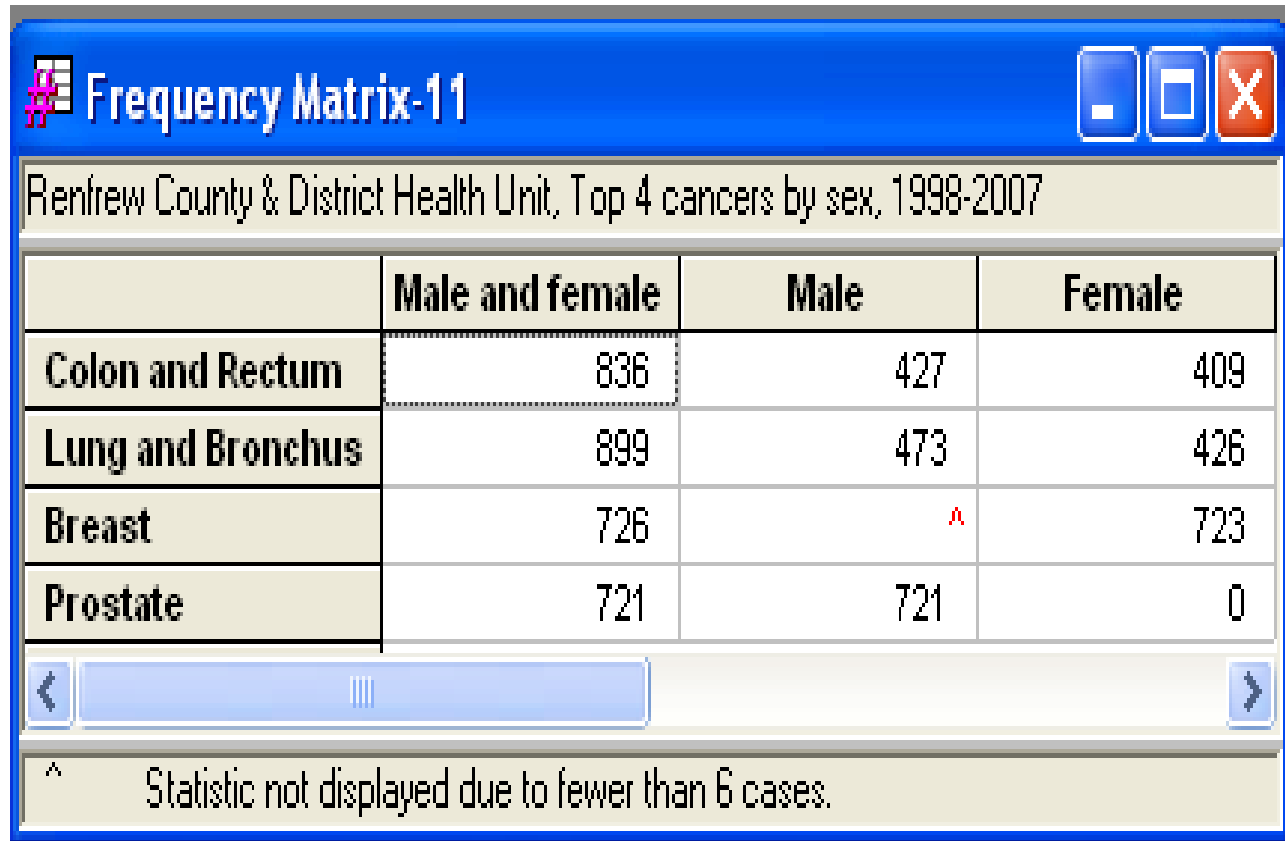
# Merged variables and SEER\*Stat rates

Diane Nishri

Senior Research Associate, Surveillance

February, 2011

# Why create a merged variable?



Frequency Matrix-11

Renfrew County & District Health Unit, Top 4 cancers by sex, 1998-2007

	Male and female	Male	Female
Colon and Rectum	836	427	409
Lung and Bronchus	899	473	426
Breast	726	^	723
Prostate	721	721	0

^ Statistic not displayed due to fewer than 6 cases.

# Defining cancer sites

- All incidence data has been converted to ICD-O-3
- *SEER site recode* variables define the major cancer sites that are commonly used in the reporting of cancer incidence data
  - ‘*with Kaposi and mesothelioma*’ means these cancers have separate categories – this is the recommended variable to use
  - Morphologies 9590-9989 (lymphomas and leukemias) are excluded from the topographically defined sites (eg Ovary C56.9)
  - Malignant melanoma is defined by both site (C44 Skin) and morphology (8720-8790)
  - Ignore ‘nodal’ & ‘extranodal’ for NHL and Hodgkin
  - Analyses of leukemia subtypes is strongly discouraged

# File -> Dictionary -> Merge...

The image shows two overlapping dialog boxes from a software application. The left dialog, titled "Edit Merged Variable", has a "Name:" field, a "Description:" field with an "Edit..." button, and a "Groupings" list. Below the list are buttons for "Add...", "Add All...", "Edit...", "Delete", "Rename", "Move Up", and "Move Dn". At the bottom, there is a "Grouping Definition" area with a list box and checkboxes for "Include 'Other' Grouping" and "Save to Dictionary". The right dialog, titled "New Merged Grouping", features a "Variable" list with folders like "Age at Diagnosis", "Race, Sex, Year Dx, State, County", "Site and Morphology", "Race and Age (case data only)", "Dates", "Other", and "User-Defined". It also has an "Operator" section with "is = to" and "is not = to" options, a "Values" field, a "Find..." button, a "Conjunction:" dropdown, and a "New Line" button. A "Selection Statement" text area is at the bottom with "Up", "Down", "Delete", "Copy", and "Paste" buttons. Both dialogs have "Add (...)", "Delete (...)", "Del All (...)", "OK", "Cancel", and "Help" buttons at the bottom.

**Edit Merged Variable**

Name:

Description:

Edit...

Groupings

Add...

Add All...

Edit...

Delete

Rename

Move Up

Move Dn

Grouping Definition

Include "Other" Grouping

Save to Dictionary

OK Cancel Help

**New Merged Grouping**

Variable

- Age at Diagnosis
- Race, Sex, Year Dx, State, County
- Site and Morphology
- Race and Age (case data only)
- Dates
- Other
- User-Defined

Operator

is = to

is not = to

Values

Find...

Modify Conjunction:

New Line

Selection Statement

Up

Down

Delete

Copy

Paste

Add (...) Delete (...) Del All (...)

OK Cancel Help

# New merged variable

**Edit Merged Variable**

Name:

Description:

Groupings

Colon & Rectum	<input type="button" value="Add..."/>
Lung & Bronchus	<input type="button" value="Add All..."/>
Prostate	<input type="button" value="Edit..."/>
<input type="text" value="Female Breast"/>	<input type="button" value="Delete"/>
	<input type="button" value="Rename"/>
	<input type="button" value="Move Up"/>
	<input type="button" value="Move Dn"/>

Grouping Definition

{Site and Morphology.Site rec with Kaposi and mesothelioma} = ' Breast'  
AND {Sex, Year Dx, Geography.Sex} = ' Female'

Include "Other" Grouping  
 Save to Dictionary

# Create a new variable for calculating SIRs

**Edit Variable - Public Health Unit**

Name:  Category:

Description:

---

Groupings

Ontario excluding unknown  
Renfrew County and District HU  
All other PHUs

Values (0 Selected)

3556 Porcupine HU  
3557 Renfrew County and District HU  
3558 Eastern Ontario HU  
3560 Simcoe Muskoka District HU  
3561 Sudbury and District HU  
3562 Thunder Bay District HU  
3563 Timiskaming HU  
3565 Region of Waterloo HU  
3566 Wellington-Dufferin-Guelph HU  
3568 Windsor-Essex County HU  
3570 York Region HU  
3595 Toronto HU  
9999 Unknown

Save to Dictionary

## Exercise #2a: New variables

- Begin a Rates session and select the PHU incidence file
- Create a merged variable for the following cancers with sexes combined:
  - Female breast, cervical, colorectal, lung, malignant melanoma, oral, prostate
- Create a user-defined variable based on PHU with the following categories in this specific order:
  - Ontario excluding unknown residence
  - Your PHU
  - All other PHUs

## Crude incidence rate

- The number of new cases occurring in a specified population during the same time period, usually expressed as the number of cases per 100,000 population at risk.

35,860 colon & rectum cases diagnosed  
in Ontario, 2002-2006

---

61,917,727 residents at risk 2002-2006

**= 57.9 per 100,000**



# Crude rates example

- Create a table with crude rates for the Top 4 cancers, sexes combined, for the years 2003-2007, for Renfrew
- Use the 1991 Canadian standard population
- Include confidence intervals

# Statistic Tab

The screenshot shows a software window titled "Rate Session-2" with a blue title bar and standard window controls. The main area has a tabbed interface with "Statistic" selected. The "Statistic" tab contains a list of options on the left and a "Parameters" section on the right. The "Parameters" section includes a "P-Values" field set to ".05" with an "Edit..." button, three dropdown menus for "Standard Population" (1991 Canadian Std Million (18 age groups) - User), "Age Variable" (Age recode), and "Trend Variable" (Year of diagnosis). Below these are two sub-sections: "PC End Points" with radio buttons for "One Year" and "Two-Year Avg.", and "APC Calculation" with radio buttons for "Weighted Least Squares" and "Non-Weighted Least Squares". At the bottom, there are four checkboxes: "Include Rate Ratios on Last Row Variable Groupings" (unchecked), "Show Standard Errors and Confidence Intervals" (checked), "Use Tiwari et al., 2006 modification for CIs" (unchecked), and "Show P-Values for APC Significance Testing" (unchecked).

**Rate Session-2**

Data | **Statistic** | Selection | Table | Output

**Statistic**

- Rates (Crude)
- Rates (Age-Adjusted)
- Trends (Crude)
- Trends (Age-Adjusted)

**Parameters**

P-Values: .05

Standard Population: 1991 Canadian Std Million (18 age groups) - User

Age Variable: Age recode

Trend Variable: Year of diagnosis

**PC End Points**

- One Year
- Two-Year Avg.

**APC Calculation**

- Weighted Least Squares
- Non-Weighted Least Squares

Include Rate Ratios on Last Row Variable Groupings

Show Standard Errors and Confidence Intervals

- Use Tiwari et al., 2006 modification for CIs
- Show P-Values for APC Significance Testing

# Selection Tab

**Rate Session-2**

Data | **Statistic** | **Selection** | Table | Output

Select Only

Malignant Behavior       Known Age  
 Male or Female Sex       Cases in Limited-Use Database

Age at Diagnosis (Std Pop, Pop, Case Files)

Age at Diagnosis (Std Pop, Pop, Case Files)

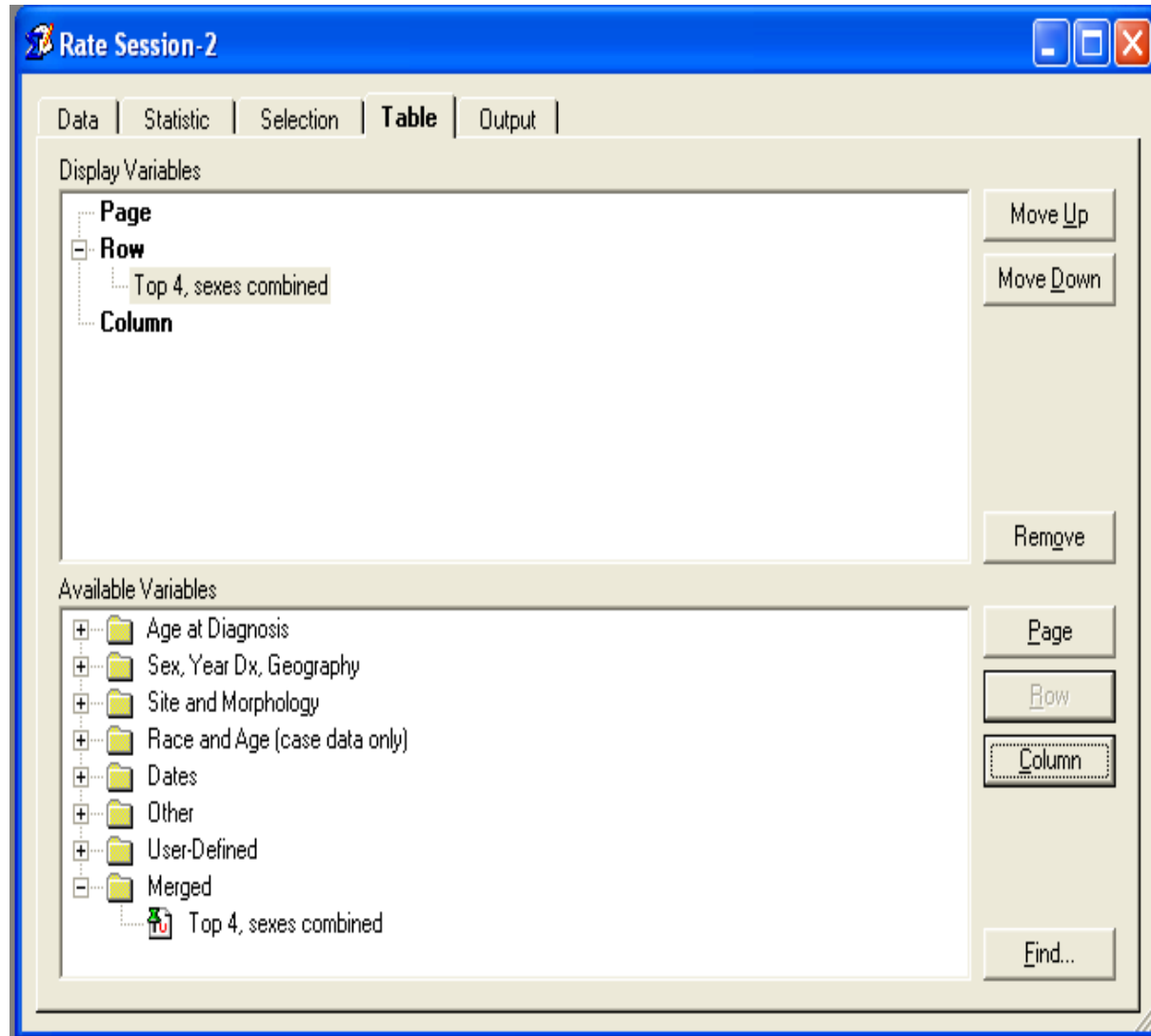
Sex, Year Dx, Geography (Pop, Case Files)

{Sex, Year Dx, Geography.Public Health Unit} = '3557 Renfrew County and District HU'  
AND {Sex, Year Dx, Geography.Year of diagnosis} = '2003','2004','2005','2006','2007'

Other (Case Files)

Select Only the First Matching Record for Each Person

# Table Tab



# Rate Matrix

Rate Matrix-5

Top 4 cancers, sexes combined, crude incidence, 2003-2007, Renfrew County & District HU

	Rate	SE	Lower CI	Upper CI	Count	Pop
<b>Colon &amp; Rectum</b>	92.5	4.3	84.4	101.3	472	510,086
<b>Lung &amp; Bronchus</b>	91.4	4.2	83.2	100.0	466	510,086
<b>Prostate</b>	140.0	7.4	125.8	155.3	357	255,025
<b>Female Breast</b>	151.3	7.7	136.6	167.2	386	255,061

^ Rates are per 100,000; Confidence intervals are 95% for rates.  
Statistic not displayed due to fewer than 6 cases.

## Exercise #2b: Crude rates

- Choose an appropriate time period for your PHU
- Select those years and your PHU
- Calculate crude rates for the seven cancers of interest, sexes combined

# Age-specific incidence rate

- Because differences in the age composition of the population influences crude incidence rates, it is preferable to use age-specific rates in comparing the incidence experience in different geographic areas, population groups or time periods.
- A crude rate calculated for a specific age group. Both the numerator and denominator refer to the same age group.
- The 18 age groups used for cancer epidemiology are usually 0-4, 5-9, 10-14, ..., 80-84, 85+

## Age-specific incidence rate

2,310 colon & rectum cases diagnosed in males  
aged 60-64, Ontario, 2002-2006

---

1,332,967 males aged 60-64 living in Ontario  
2002-2006

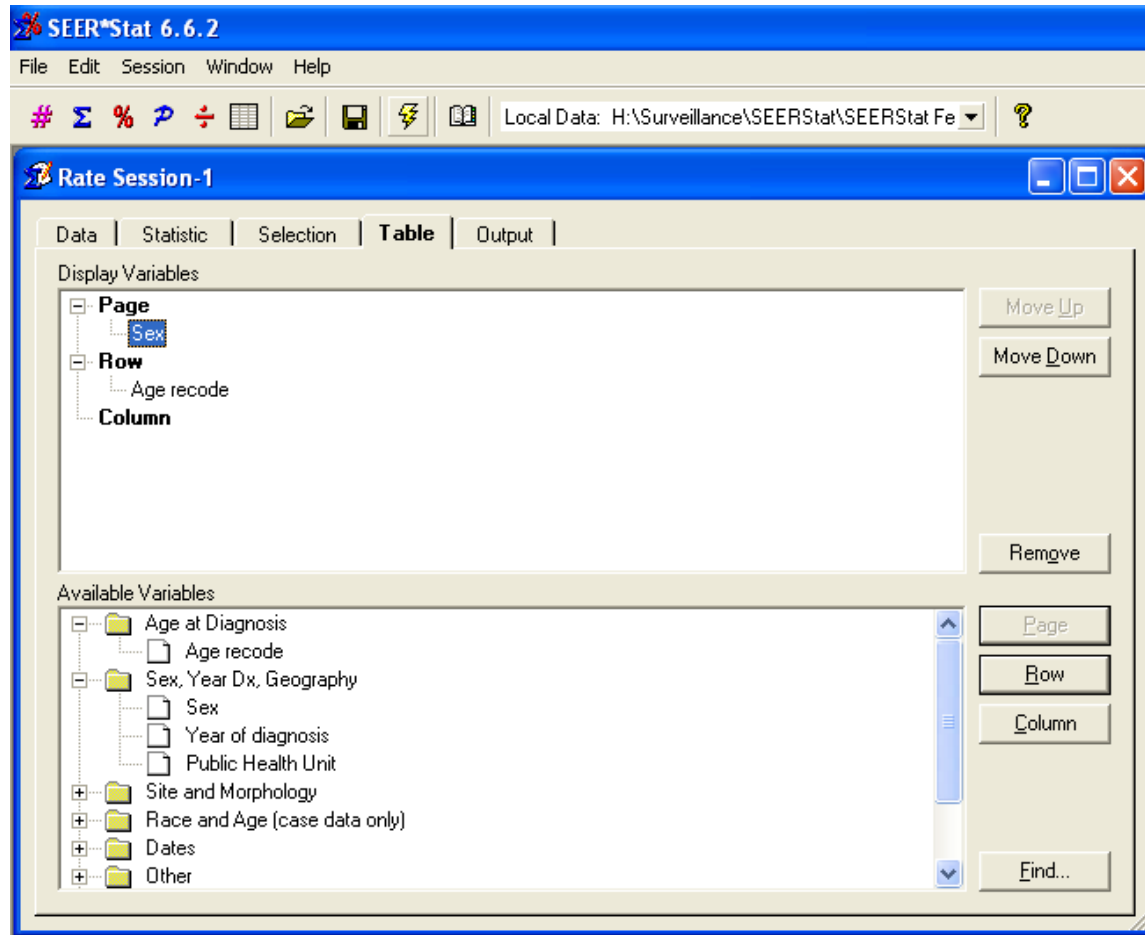
**= 173.3 per 100,000**



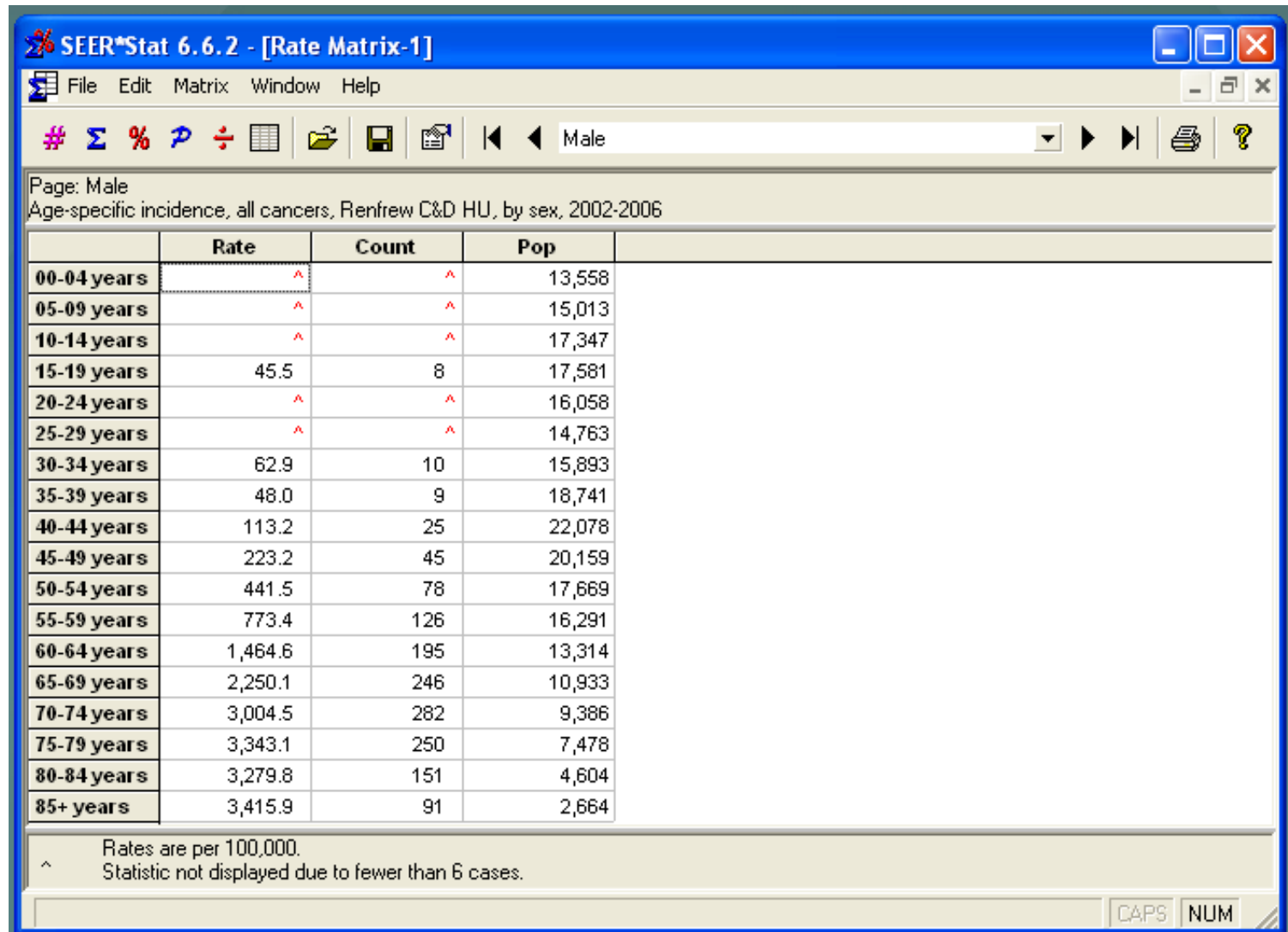
# Age-specific rates example

- Calculate age-specific incidence rates for all cancers by sex, for the 18 age groups 0-4, 5-9, ..., 85+ for Renfrew County & District Health Unit
- Years are 2002-2006, to match Cancer in Ontario report
- Graph the age-specific incidence curves

# Table tab



# Rate Matrix



SEER\*Stat 6.6.2 - [Rate Matrix-1]

File Edit Matrix Window Help

# Σ % P ÷ [Grid] [Folder] [Save] [Print] [Refresh] [Left Arrow] [Right Arrow] Male [Dropdown] [Print] [Help]

Page: Male  
Age-specific incidence, all cancers, Renfrew C&D HU, by sex, 2002-2006

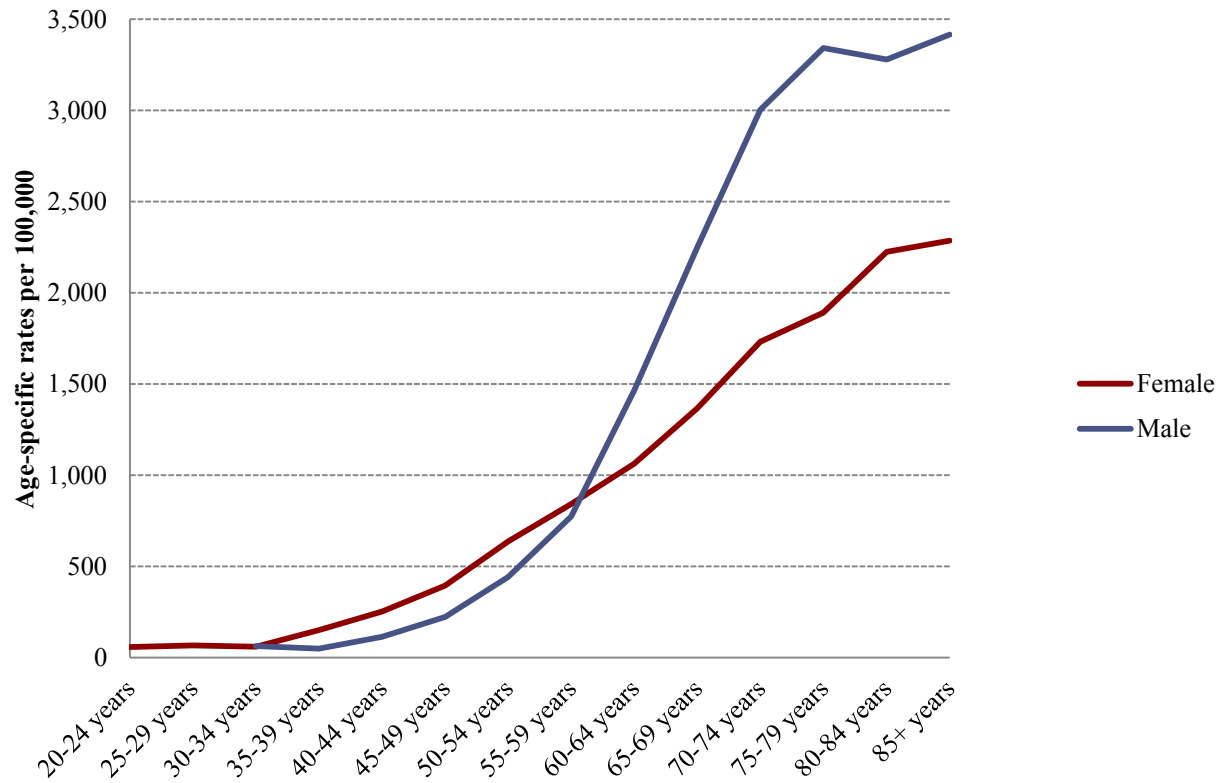
	Rate	Count	Pop
00-04 years	^	^	13,558
05-09 years	^	^	15,013
10-14 years	^	^	17,347
15-19 years	45.5	8	17,581
20-24 years	^	^	16,058
25-29 years	^	^	14,763
30-34 years	62.9	10	15,893
35-39 years	48.0	9	18,741
40-44 years	113.2	25	22,078
45-49 years	223.2	45	20,159
50-54 years	441.5	78	17,669
55-59 years	773.4	126	16,291
60-64 years	1,464.6	195	13,314
65-69 years	2,250.1	246	10,933
70-74 years	3,004.5	282	9,386
75-79 years	3,343.1	250	7,478
80-84 years	3,279.8	151	4,604
85+ years	3,415.9	91	2,664

^ Rates are per 100,000.  
^ Statistic not displayed due to fewer than 6 cases.

CAPS NUM

# Results

## Age-specific incidence rates, all cancers Renfrew C&D HU, 2002-2006



## Exercise #2c: Age-specific rates

- Using the same selection of years and PHU, calculate age-specific rates for the seven cancers of interest, sexes combined

# Age-standardized rates

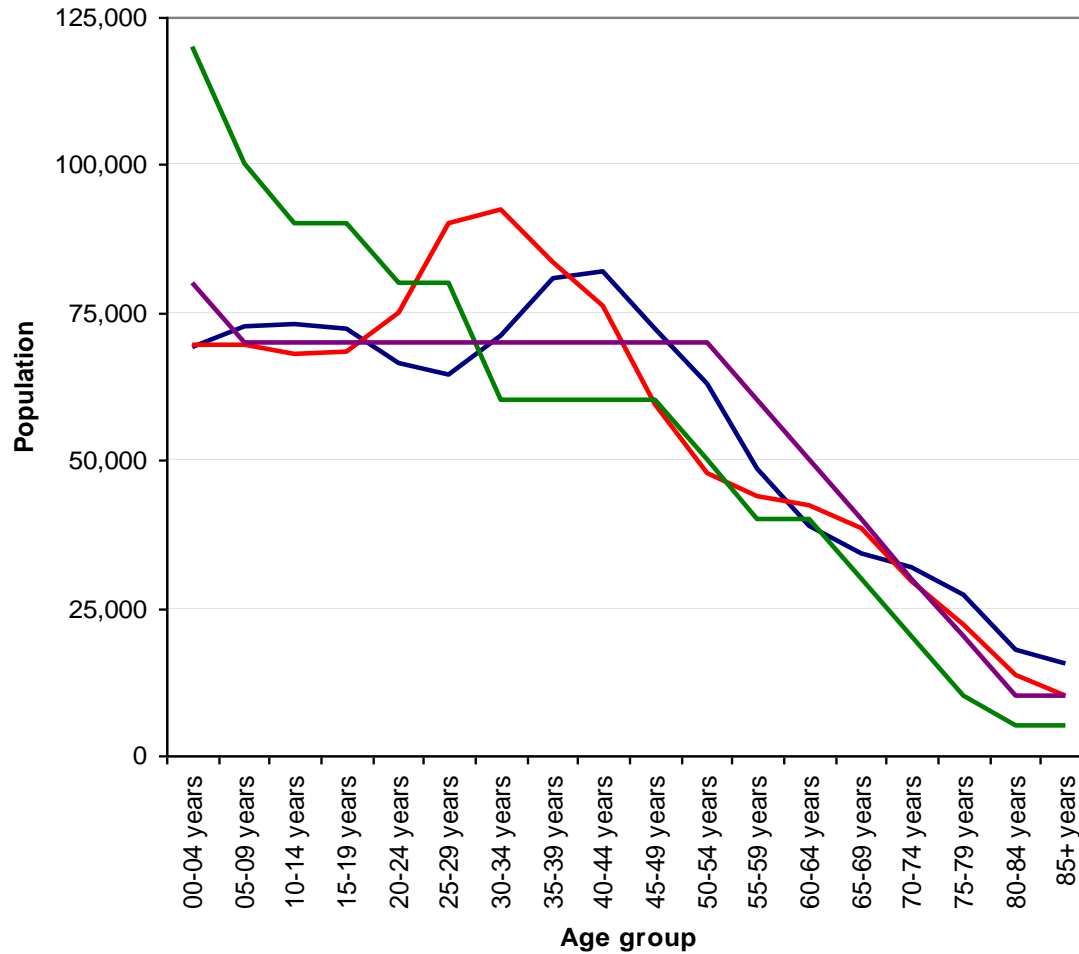
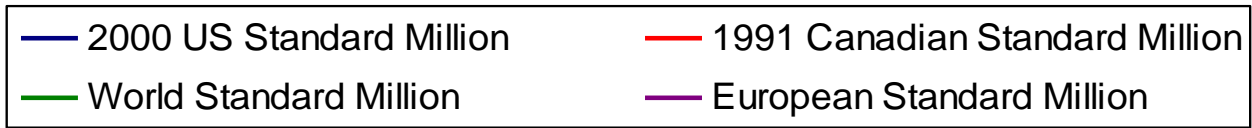
- While it is important to examine age-specific rates, it is very useful to have a summary statistic which takes into account the differences in the age distribution of the population
- Failing to account for differences in population age structure can result in inaccurate comparisons between populations or within a population across time
- Age has a powerful influence on the risk of cancer

# Age-standardized rates

- The Direct Method of age adjustment is the one most commonly used by epidemiologists
  - This requires the selection of a 'standard population' and the availability of age-specific rates for the same age groups

$$\sum_{\text{age groups}} \left( \text{crude rate for age group} \times \frac{\text{standard population for age group}}{\sum_{\text{age groups}} \text{standard population}} \right)$$

# Popular standard populations





# US 2000 Standard Population

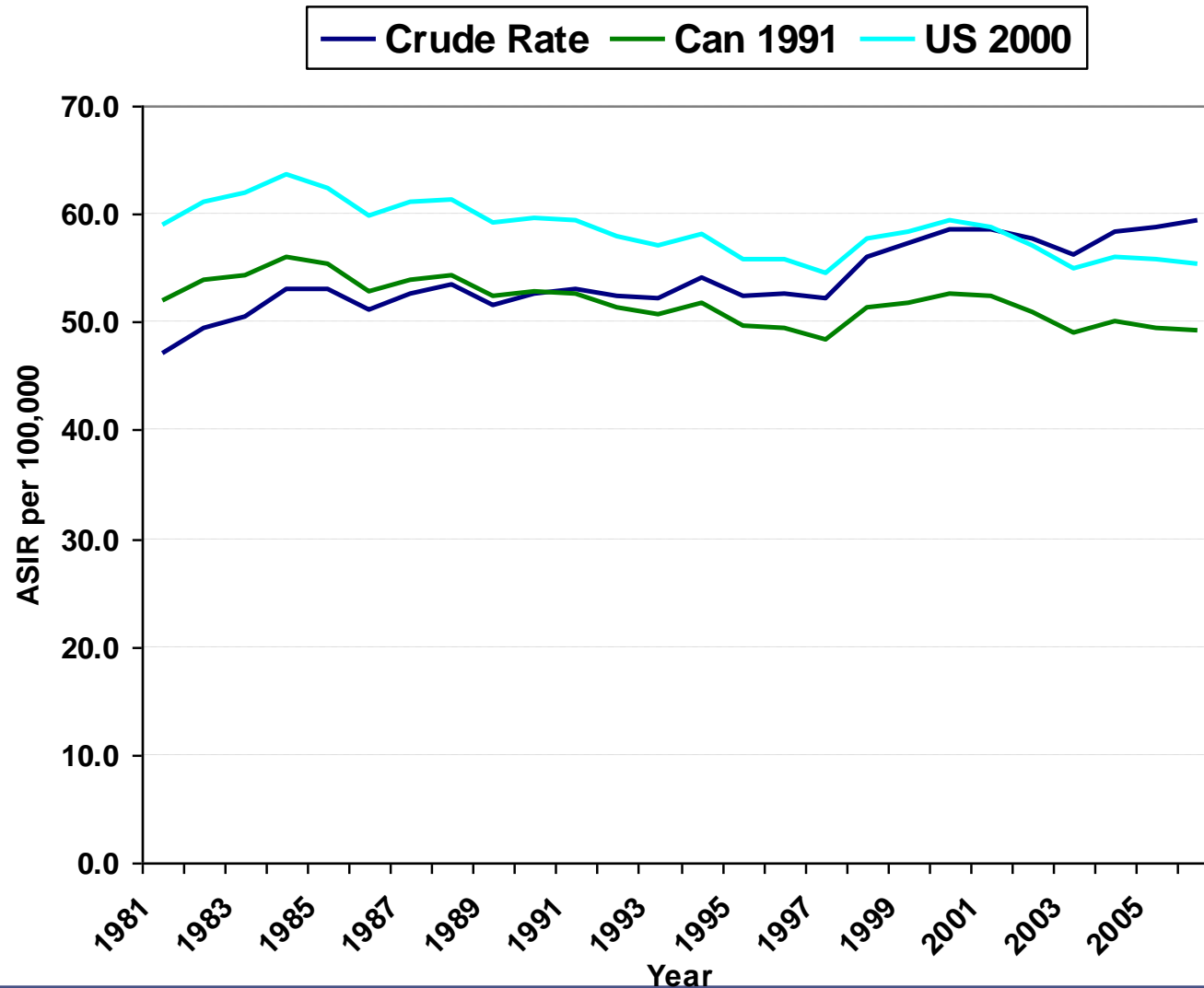
- In August 1998, the Secretary of Health and Human Services issued a policy memorandum directing all agencies and programs of the department to begin adjusting death rates by using the Census Bureau–projected US population for 2000
  - The use of different standards by different agencies was causing confusion
  - It took nearly a decade of review to come to this decision!

# Why Canada 1991?

- Pro:
  - Confusion associated with the introduction of a new standard
- Con:
  - Changing the standard for age adjustment would reflect the current age structure of the population
  - Using different standard populations only produces different results when relative differences are not consistent across age groups

# The impact of different standard pop'ns

## Ontario colorectal cancer incidence



# Age-standardized rates example

- Create a table with age-standardized rates for the Top 4 cancers, sexes combined, for the years 2003-2007, for Renfrew
- Use the 1991 Canadian standard population
- Include confidence intervals

# Statistic tab

**Rate Session-2**

Data | **Statistic** | Selection | Table | Output

Statistic

- Rates (Crude)
- Rates (Age-Adjusted)
- Trends (Crude)
- Trends (Age-Adjusted)

Parameters

P-Values: .05

Standard Population: 1991 Canadian Std Million (18 age groups) - User

Age Variable: Age recode

Trend Variable: Year of diagnosis

PC End Points

- One Year
- Two-Year Avg.

APC Calculation

- Weighted Least Squares
- Non-Weighted Least Squares

Include Rate Ratios on Last Row Variable Groupings

Show Standard Errors and Confidence Intervals

- Use Tiwari et al., 2006 modification for CIs
- Show P-Values for APC Significance Testing

## Tiwari et al modification

- Tiwari RC, Clegg LX, Zou Z. (2006) Efficient interval estimation for age-adjusted cancer rates. *Stat Methods Med Res* 6:547-69
- The modifications of the gamma (rates) and F intervals (rate ratios) in the Tiwari method are more efficient in the sense that they have empirical coverage probabilities less than or equal to their counterparts and also retain the nominal level.

# Rate Matrix window

Rate Matrix-6

Top 4 cancers, sexes combined, 2003-2007, Renfrew County & District HU

	Rate	SE	Lower CI	Upper CI	Count	Pop
<b>Colon &amp; Rectum</b>	64.9	3.0	59.1	71.1	472	510,086
<b>Lung &amp; Bronchus</b>	65.6	3.1	59.7	72.0	466	510,086
<b>Prostate</b>	110.1	5.9	98.9	122.3	357	255,025
<b>Female Breast</b>	104.9	5.5	94.4	116.3	386	255,061

^ Rates are per 100,000 and age-adjusted to the 1991 Canadian Std Million (18 age groups) - User standard; Confidence Statistic not displayed due to fewer than 6 cases.

## Exercise #2d: Age-standardized rates

- Using the same selection of years and PHU, calculate age-standardized rates for the seven cancers of interest, sexes combined
- Include confidence intervals
- How do these results compare to the crude rates?



# Truncated age-adjusted rates

North East LHIN colorectal incidence, 2001-2006, sexes combined

	Age-specific rate per 100,000	Cases	Population	Wrong weights*	Correct weights*
<b>30-34 years</b>	1.97	4	203,042	60,000	60,000
<b>35-39 years</b>	4.13	10	241,958	60,000	60,000
<b>40-44 years</b>	14.79	43	290,766	60,000	60,000
<b>45-49 years</b>	39.11	110	281,255	60,000	60,000
<b>Totals</b>		167	1,017,021	1,000,000	240,000
<b>Rates</b>	<b>16.42</b>			<b>3.60</b>	<b>15.55</b>

\*using the World standard population

# How to calculate truncated age-adjusted rates

- Create a user-defined variable with the appropriate age groups, ie 0-19, 20-69, 70+ for cervical screening
- On the Statistic tab
  - Select age-adjusted rates
  - Leave 'Age recode' as the age variable parameter
- Use your user-defined variable in your Table definition
- If you use the user-defined variable instead of 'Age recode', then age-adjustment will be done *between* age groups, but not *within*

# Standardized incidence ratio (SIR)

- “The ratio of observed new cancer cases to the number expected if the population had the same age-specific incidence rates as Ontario.” (APHEO)
- SEER\*Stat can not do this calculation directly; you would have to use Ontario’s rates and your PHU’s population to calculate expected numbers of cases.
- Instead, SEER\*Stat offers rate ratios, where you can compare your PHU’s incidence rate to Ontario’s incidence rate.

## Include Rate Ratios...

- The first grouping of the last row variable in the analysis is used as the reference grouping (the denominator of the ratio), and the rate for each subsequent grouping is compared (the numerator of the ratio) to the reference grouping's rate
- Four additional columns, containing the rate ratio and related statistics, will be displayed in your Rate results matrix

## Rate ratio example

- Calculate rate ratios for the Top 4 cancers, sexes combined, 2003-2007, comparing Renfrew C&D HU to Ontario.
- Graph the results with confidence intervals
- If we use the default PHU variable as the row variable, all other PHUs will be compared to Algoma, because it is first.
- NB – remove the PHU restriction from the Selection tab

# Statistic tab for rate ratios

**Rate Session-1**

Data | **Statistic** | Selection | Table | Output

Statistic

- Rates (Crude)
- Rates (Age-Adjusted)
- Trends (Crude)
- Trends (Age-Adjusted)

Parameters

P-Values: .05

Standard Population: 1991 Canadian Std Million (18 age groups) - User

Age Variable: Age recode

Trend Variable: Year of diagnosis

PC End Points

- One Year
- Two-Year Avg.

APC Calculation

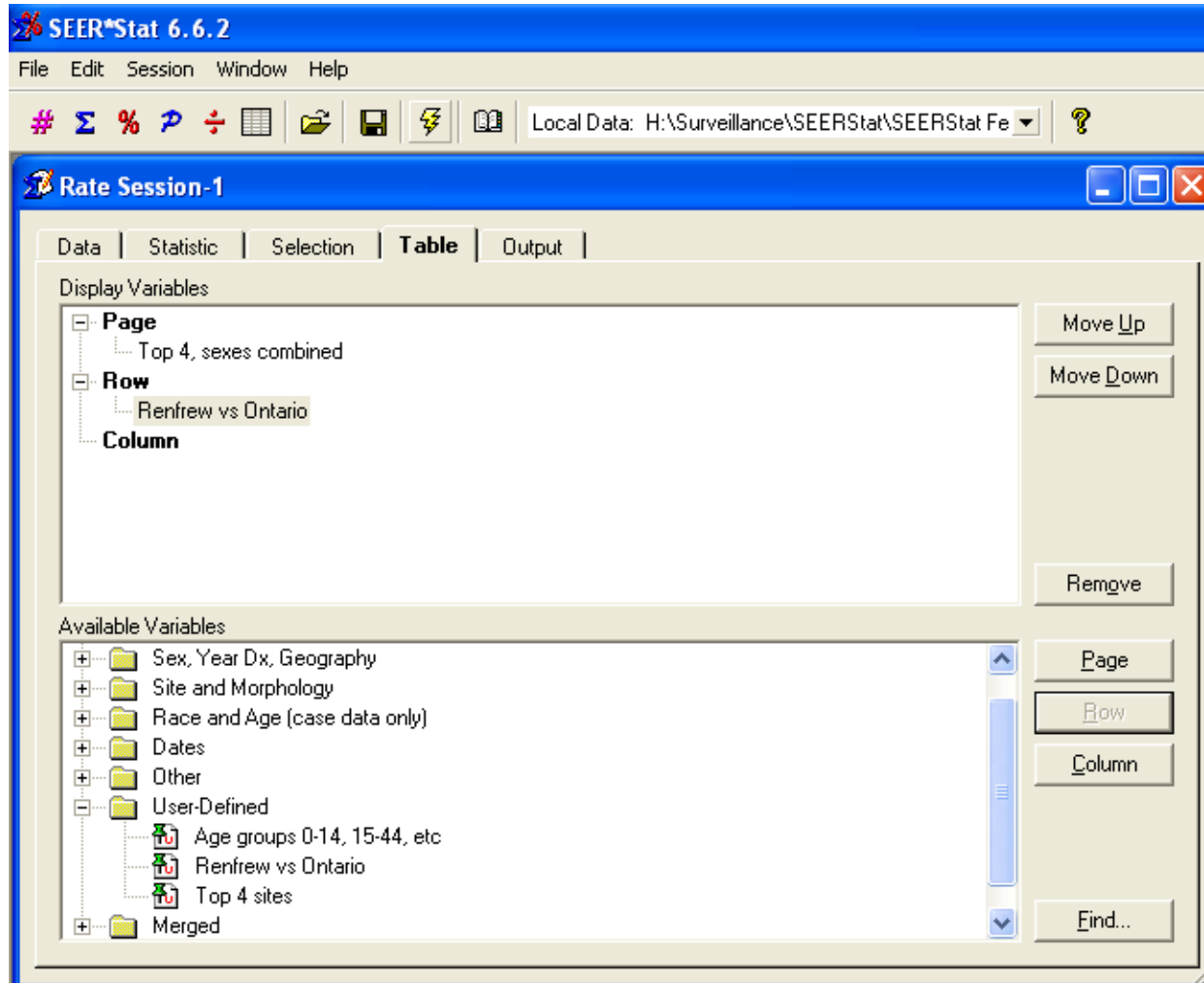
- Weighted Least Squares
- Non-Weighted Least Squares

Include Rate Ratios on Last Row Variable Groupings

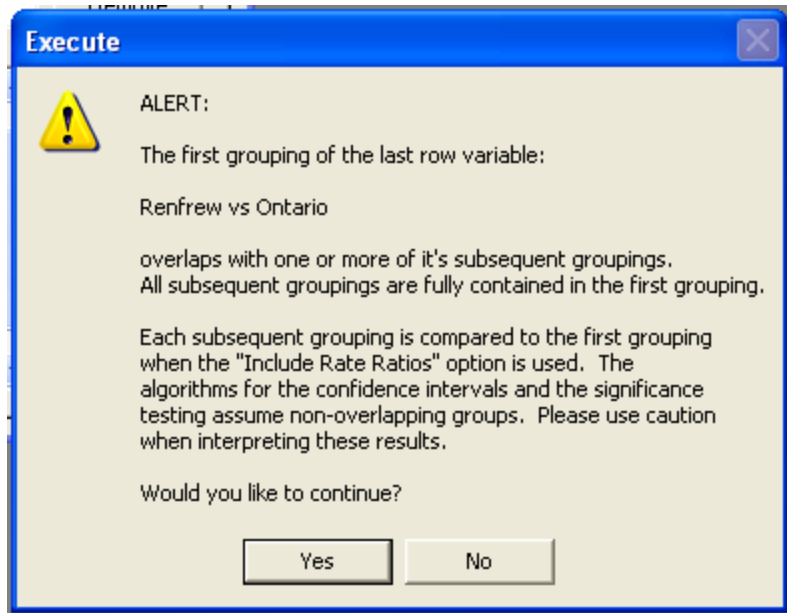
Show Standard Errors and Confidence Intervals

- Use Tiwari et al., 2006 modification for CIs
- Show P-Values for APC Significance Testing

# Table tab



# Danger, Will Robinson!





# Rate Matrix

SEER\*Stat 6.6.2 - [Rate Matrix-4]

File Edit Matrix Window Help

# Σ % P ÷ [grid icon] [print icon] [save icon] [refresh icon] Colon & Rectum [print icon] [help icon]

Page: Colon & Rectum  
Top 4 cancers, sexes combined, by PHU, 2003-2007

	Rate	SE	Lower CI	Upper CI	Count	Pop	Rate Ratio	Ratio Lower CI	Ratio Upper CI	Ratio P-Value
<b>Ontario excluding unknow</b>	49.0	0.3	48.5	49.5	36,449	62,621,387				
<b>Renfrew County and Distri</b>	64.9#	3.0	59.1	71.1	472	510,086	1.3230	1.2042	1.4516	0.0000
<b>All other PHUs</b>	48.9	0.3	48.4	49.4	35,977	62,111,301	0.9967	0.9821	1.0116	0.6667

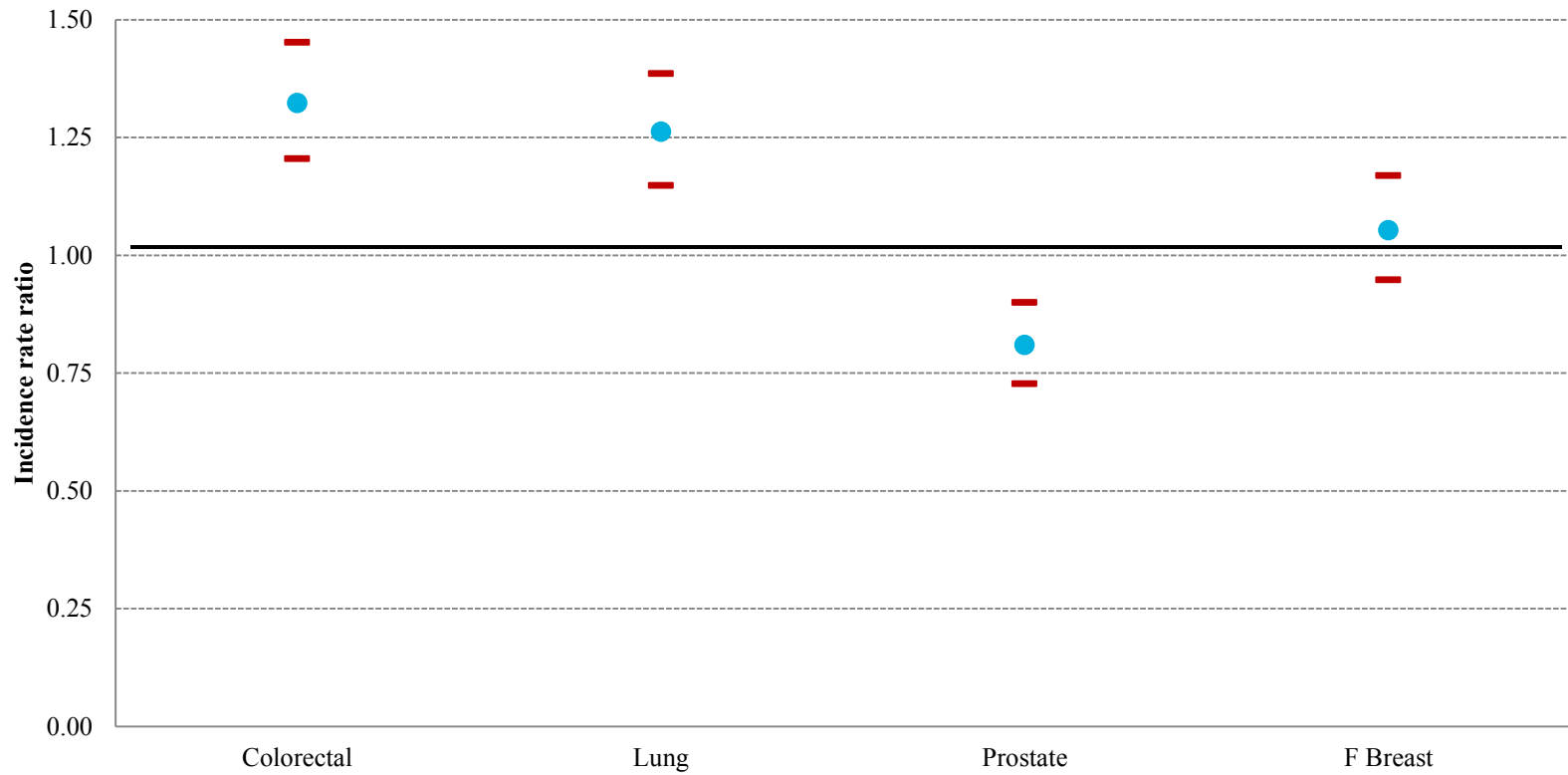
^ Rates are per 100,000 and age-adjusted to the 1991 Canadian Std Million (18 age groups) - User standard; Confidence intervals (Tiwari mod) are 95% for rates and ratios.  
Statistic not displayed due to fewer than 6 cases.

# The rate ratio indicates that the rate is significantly different than the rate for Ontario excluding unknown (p<0.05).  
Warning: Use caution when interpreting ratios and related statistics as the ratio variable contains overlapping groupings.

CAPS NUM

# Graphed results

**Rate ratios for Renfrew vs Ontario  
Top 4 cancers, sexes combined, 2003-2007**



## Exercise #2e: Rate ratios

- Using the same selection of years, calculate rate ratios for the seven cancers of interest, sexes combined, comparing your PHU to Ontario
- Include confidence intervals
- Any interesting results you'd like to share?