

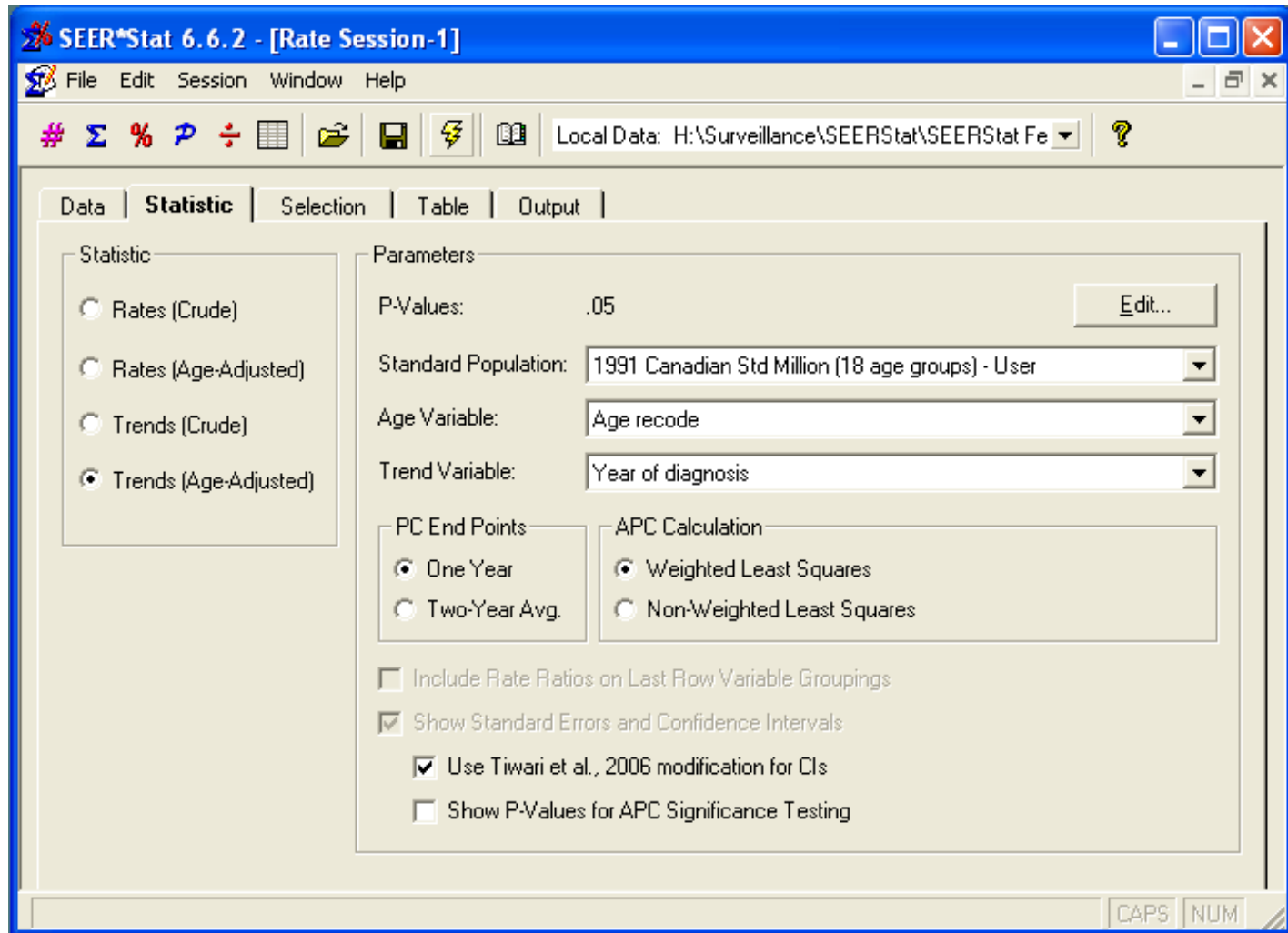
Trend Analysis using SEER*Stat and Joinpoint

Diane Nishri

Senior Research Associate, Surveillance

February, 2011

Trend Analysis in SEER*Stat



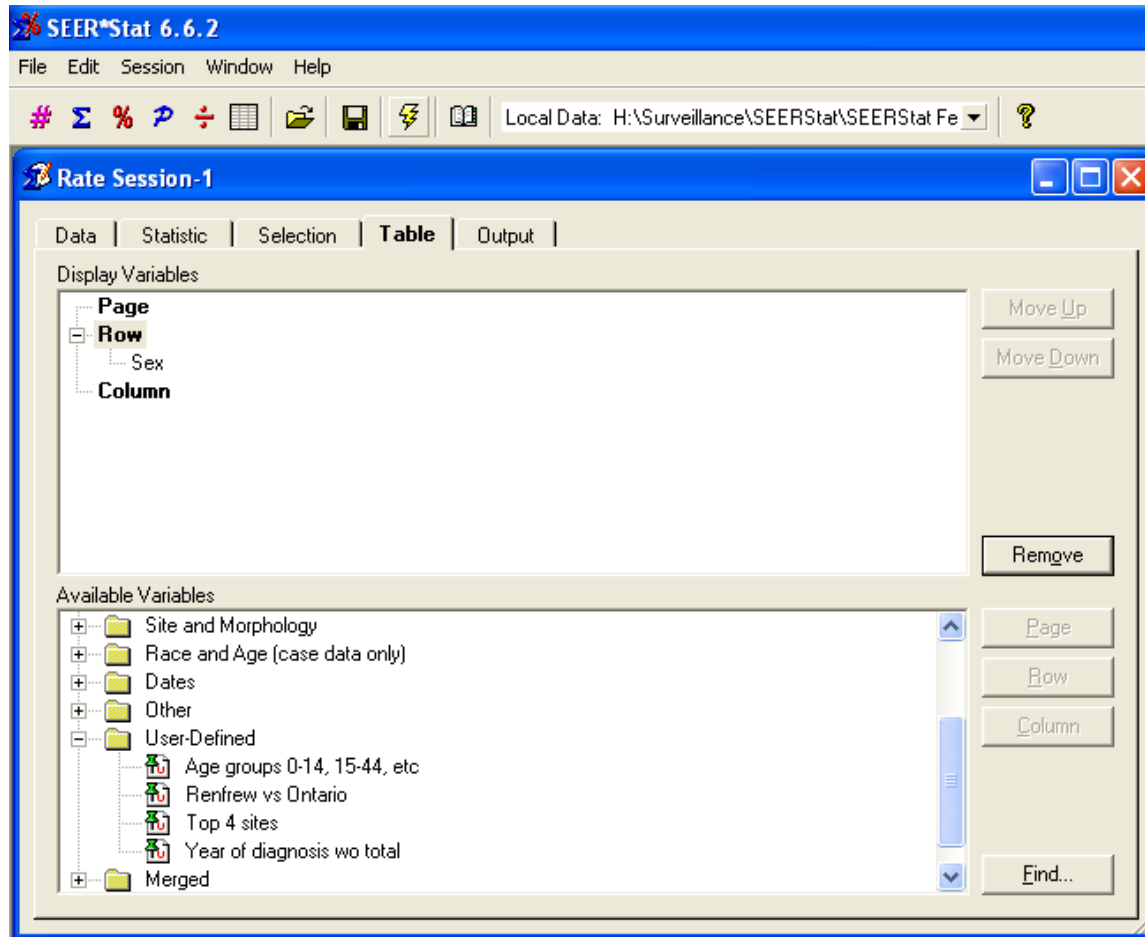
Trends (Crude or Age-adjusted)

- Trends in crude or age-adjusted rates are expressed in two forms: the percent change (PC) and the annual percent change (APC). Note that age adjustment minimizes the effect of a difference in age distributions when analyzing trends.
 - $PC = ((\text{end rate} - \text{initial rate}) / \text{initial rate}) * 100$. Either a one-year rate or the unweighted average of two one-year rates can be used for the initial and end rates.
 - The APC is calculated by fitting a regression line to the natural logarithm of the rates (r) using the calendar year (x) as a regressor variable, i.e., $y = mx + b$ where $y = \ln(r)$. You may also utilize the standard errors of the rates to fit to a weighted least squares regression line. (*recommended*)

Trend example

- Examine the trend in age-standardized female lung cancer incidence rates, Ontario, 1986-2007. Use Canada 1991 standard population.
- Plug for *Cancer Facts*:
 - “International trends in lung cancer death rates reflect smoking patterns (Jan. 2011)”

Table tab



Rate matrix

SEER*Stat 6.6.2
File Edit Matrix Window Help

Σ % P ÷ [grid icon] [save icon] [print icon] [back icon] [forward icon]

Rate Matrix-18

Lung and bronchus, by sex, Ontario

| | | Rate/Trend | SE/P-Value | Lower CI | Upper CI |
|------------------|------------------|------------|------------|----------|----------|
| Female | Total PC | 30.63 | | | |
| | Total APC | 1.24* | 0.00 | 0.99 | 1.48 |
| | 1986 Rate | 33.33 | 0.81 | 31.76 | 34.96 |
| | 1987 Rate | 34.08 | 0.81 | 32.51 | 35.71 |
| | 1988 Rate | 36.61 | 0.83 | 35.01 | 38.27 |
| | 1989 Rate | 34.97 | 0.80 | 33.42 | 36.57 |
| | 1990 Rate | 36.61 | 0.81 | 35.04 | 38.23 |
| | 1991 Rate | 38.26 | 0.82 | 36.67 | 39.89 |
| | 1992 Rate | 39.11 | 0.82 | 37.52 | 40.75 |
| | 1993 Rate | 39.12 | 0.81 | 37.55 | 40.74 |
| | 1994 Rate | 39.40 | 0.80 | 37.84 | 41.01 |
| | 1995 Rate | 40.35 | 0.80 | 38.79 | 41.96 |
| | 1996 Rate | 41.41 | 0.81 | 39.84 | 43.02 |
| | 1997 Rate | 41.46 | 0.80 | 39.91 | 43.06 |
| | 1998 Rate | 43.54 | 0.81 | 41.96 | 45.16 |
| | 1999 Rate | 43.27 | 0.80 | 41.71 | 44.86 |
| | 2000 Rate | 44.04 | 0.80 | 42.48 | 45.64 |
| | 2001 Rate | 43.75 | 0.79 | 42.22 | 45.33 |
| | 2002 Rate | 42.82 | 0.77 | 41.32 | 44.37 |
| | 2003 Rate | 42.11 | 0.76 | 40.64 | 43.62 |
| 2004 Rate | 43.74 | 0.76 | 42.26 | 45.27 | |
| 2005 Rate | 45.30 | 0.77 | 43.81 | 46.84 | |
| 2006 Rate | 44.12 | 0.75 | 42.67 | 45.62 | |
| 2007 Rate | 43.93 | 0.74 | 42.50 | 45.41 | |

Rates are per 100,000 and age-adjusted to the 1991 Canadian Std Million (18 age groups) - User stand.
 Percent changes were calculated using 2 years for each end point; APCs were calculated using weigh
 ^
 Statistic not displayed due to fewer than 6 cases.
 *
 The APC is significantly different from zero (p<0.05). See P-Value in the APC row.

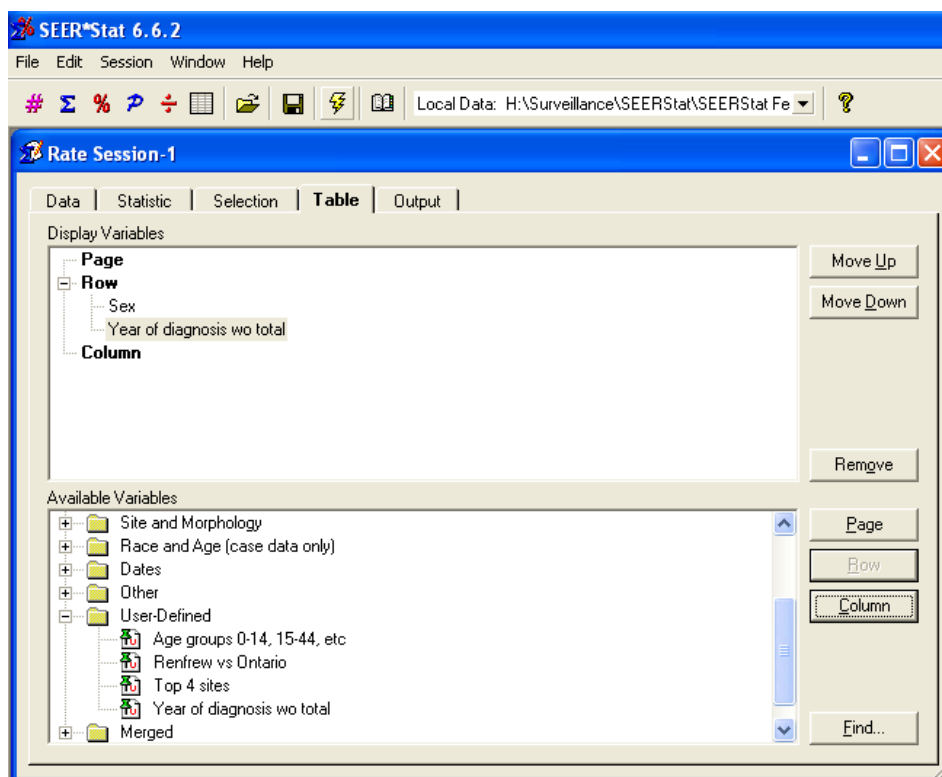
What is Joinpoint?

- A Windows-based statistical software package that analyzes joinpoint models. The software enables the user to test whether or not an apparent change in trend is statistically significant
- Joinpoint fits cancer rates into the simplest model that the data allow, where several different lines are connected together at the “joinpoints”
- Created and maintained by NCI & IMS, the same folks responsible for SEER*Stat

Using Joinpoint

- There are four steps involved in generating any Joinpoint trend analysis:
 - **Step 1:** Create an input data file for Joinpoint
 - The Joinpoint input file must be an ASCII text file!
 - Include standard errors
 - **Step 2:** Set parameters in the Joinpoint program
 - **Step 3:** Execute the Joinpoint Regression Program
 - **Step 4:** View & interpret the Joinpoint Results

Table Tab for Joinpoint data



- The year variable must be the last row variable, and must not include the total
- On Output tab, select maximum number of decimal places

Data for Joinpoint

Rate Matrix-15 [-] [] [X]

Lung and bronchus, by sex, Ontario

| | | Rate | SE | Lower CI | Upper CI | Count | Pop |
|---------------|--------------|--------------|--------------|--------------|--------------|-----------|-----------|
| Female | 1986 | 33.330623978 | 0.812955387 | 31.756123679 | 34.963682870 | 1,696 | 4,786,739 |
| | 1987 | 34.082055594 | 0.809602757 | 32.513556541 | 35.707432798 | 1,785 | 4,893,193 |
| | 1988 | 36.612355831 | 0.829013380 | 35.005371227 | 38.274863985 | 1,962 | 4,995,629 |
| | 1989 | 34.970634764 | 0.799368079 | 33.421279797 | 36.574034016 | 1,926 | 5,132,087 |
| | 1990 | 36.605750709 | 0.809223727 | 35.036712436 | 38.227696772 | 2,059 | 5,230,077 |
| | 1991 | 38.256252403 | 0.817259318 | 36.671053428 | 39.893291468 | 2,206 | 5,273,902 |
| | 1992 | 39.111654799 | 0.818141212 | 37.524398066 | 40.749778963 | 2,304 | 5,349,233 |
| | 1993 | 39.118616295 | 0.807775243 | 37.551263218 | 40.735757725 | 2,369 | 5,413,065 |
| | 1994 | 39.398199296 | 0.803188596 | 37.839544331 | 41.005801135 | 2,436 | 5,485,622 |
| | 1995 | 40.348789169 | 0.804954682 | 38.786371750 | 41.959298719 | 2,547 | 5,558,443 |
| | 1996 | 41.406647283 | 0.807860483 | 39.838251017 | 43.022347384 | 2,668 | 5,623,709 |
| | 1997 | 41.463089978 | 0.800523809 | 39.908781911 | 43.063904172 | 2,729 | 5,690,612 |
| | 1998 | 43.537113961 | 0.812098736 | 41.959825420 | 45.160108872 | 2,931 | 5,760,598 |
| | 1999 | 43.273330664 | 0.802749910 | 41.714121448 | 44.877569425 | 2,972 | 5,828,925 |
| | 2000 | 44.037251434 | 0.801464057 | 42.480271583 | 45.638445394 | 3,099 | 5,914,733 |
| | 2001 | 43.751519805 | 0.789788753 | 42.217109809 | 45.329238094 | 3,159 | 6,017,724 |
| | 2002 | 42.820455017 | 0.774488818 | 41.315795888 | 44.367683594 | 3,158 | 6,116,076 |
| | 2003 | 42.110155487 | 0.757074654 | 40.639249924 | 43.622682974 | 3,201 | 6,193,534 |
| 2004 | 43.743695525 | 0.763361649 | 42.260191147 | 45.268079433 | 3,405 | 6,270,190 | |
| 2005 | 45.304495955 | 0.767116021 | 43.813316497 | 46.835765208 | 3,621 | 6,340,269 | |
| 2006 | 44.123346375 | 0.749545245 | 42.666361633 | 45.619775522 | 3,615 | 6,410,980 | |
| 2007 | 43.934711552 | 0.738592632 | 42.498892626 | 45.409211518 | 3,692 | 6,480,888 | |

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

Matrix -> Export -> Text file...

Export

Data File (*.txt) or Compressed Data File (*.gz)
C:\NONlungincbysex8607.txt Browse...

Export Dictionary File (*.dic)
C:\NONlungincbysex8607.dic Browse...

Generate SAS Code to Read Data (SAS File: *.sas)
Browse...

Output Variables as
 Numeric Representation
 Labels Enclosed in Quotes
 Labels Without Quotes

Line Delimiter
 DOS/Windows (CR/LF)
 UNIX (LF)

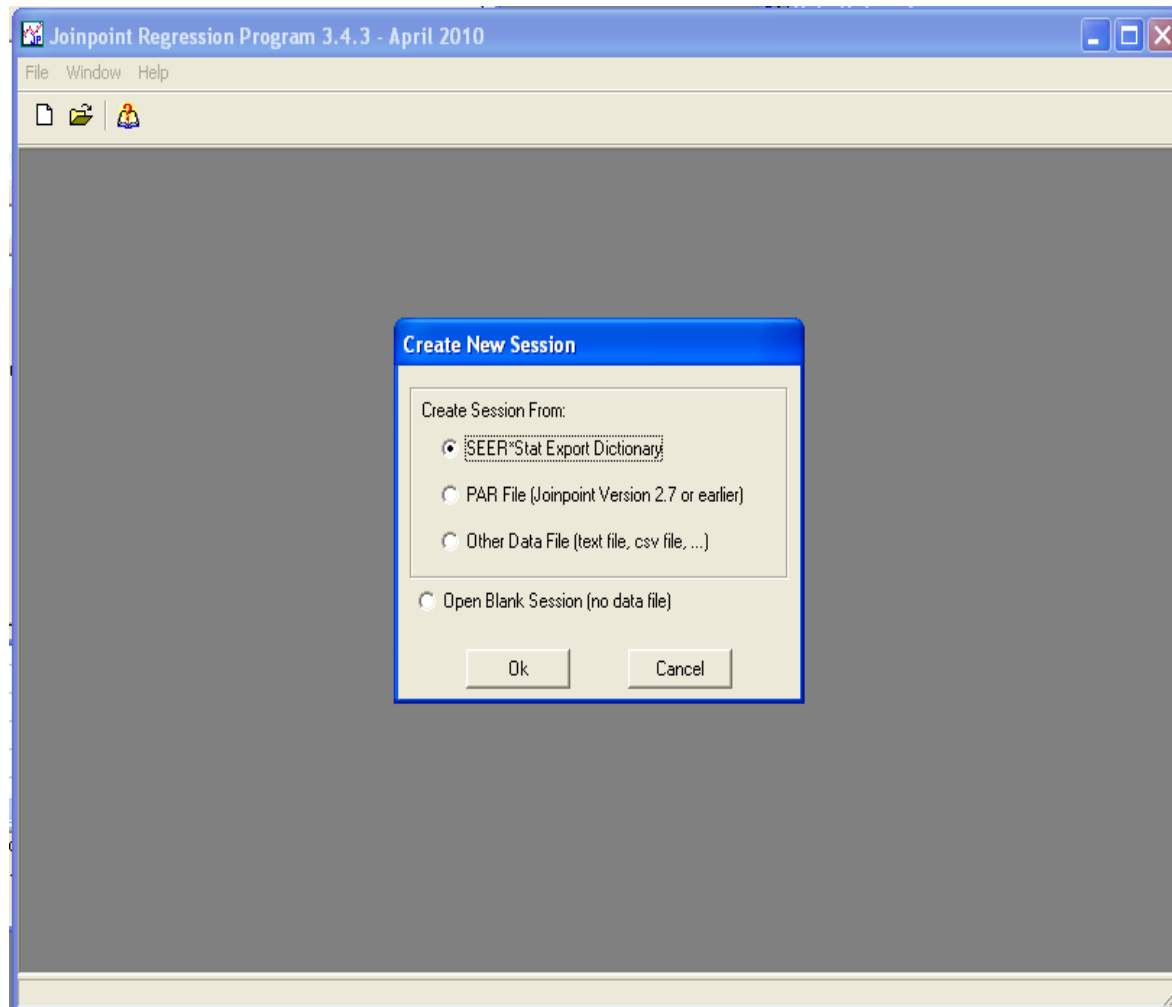
Missing Character
 Space
 Period

Field Delimiter
 Tab
 Space
 Comma
 Semi-Colon

Enclose Fields Containing Delimiter in Quotes
 Remove All Thousands Separators (Commas)
 Remove Flags (Footnote), Prefix and Suffix Characters
 Output Variable Names Before Data
 Preserve Matrix Columns and Rename Fields
 Include full path for input file in SAS code

Defaults Set Default OK Cancel Help

Starting a Joinpoint Session



Specifications

Joinpoint Session - 1

Specifications | **Advanced**

Input Data File: C:\DN\lungincbysex8607.txt Browse...

Model
 $y = xb$ $\ln(y) = xb$

Dependent Variable
Column 3 Define...

Independent Variable
The independent variable is defined by name, position and format.
Column 2 Define...

Shift data points by: 0 [What's this?](#)

Heteroscedastic Errors Option
 Constant Variance
 Input Standard Error of Dependent Variable
Standard Error: Column 4 Define...
 Poisson Model using Rate
Population: Define...
 Poisson Model using Count

By Variables
Variable Name
Column 1
Add... Delete... Define...

Number of Joinpoints
Minimum: 0 Maximum: 3

Test for pairwise differences between by-groups
 Parallel
 Coincident
[What's this?](#)

Input Standard Error of Dependent Variable

- Assumes that the random errors are heteroscedastic (have non-constant variance)
- Estimates the regression coefficients by weighted least squares
 - For model $\ln(y) = xb$, $w = (y^2)/v$, where y^2 is the square of the response for that point and v is the square of the std dev that has been input for that point

Shift Data Points by

- Allows all the values for the independent variable to be shifted up by a fixed value
- If the independent variable is years (1975, 1976,...), but you would like these points to be represented on the graph at the midpoint of the years (1975.5, 1976.5, ...), then enter the value 0.5 for this option
- *Shifting the data points will change the location of the joinpoints and the intercepts but will not change the slopes or APCs.*
- This is especially important if joinpoints are allowed to occur at places other than the data points (either in continuous time using Hudson's algorithm, or using a grid search where grid points are allowed between data points).

Advanced Parameters

The screenshot shows a software window titled "Joinpoint Session - 2" with a blue title bar and standard window controls. The main content area is divided into several sections:

- Specifications** (tabbed view, currently on "Advanced")
- Method**: Radio buttons for "Grid Search" (selected) and "Hudson's".
- Autocorrelated Errors Options**: Radio buttons for "Fit an uncorrelated errors model" (selected) and "Fit an autocorrelated errors model with parameter = []".
- Number of observations**:
 - Minimum number of observations from a joinpoint to either end of the data: [3] (including the first or last joinpoint if it falls on an observation)
 - Minimum number of observations between two joinpoints: [4] (including any joinpoint that falls on an observation)
 - Number of points to place between adjacent observed x values in the grid search: [0]
- Model Selection Method**: Radio buttons for "Permutation Test" (selected) and "BIC (Bayesian Information Criterion)".
- Permutation Test Options**:
 - Overall significance level for the permutation tests (see [help](#) for details): [0.05]
 - Number of randomly permuted data sets for permutation test: [4499]
- Early Stopping Options**: Radio buttons for "B-Value", "Curtailed", and "Fixed" (selected).

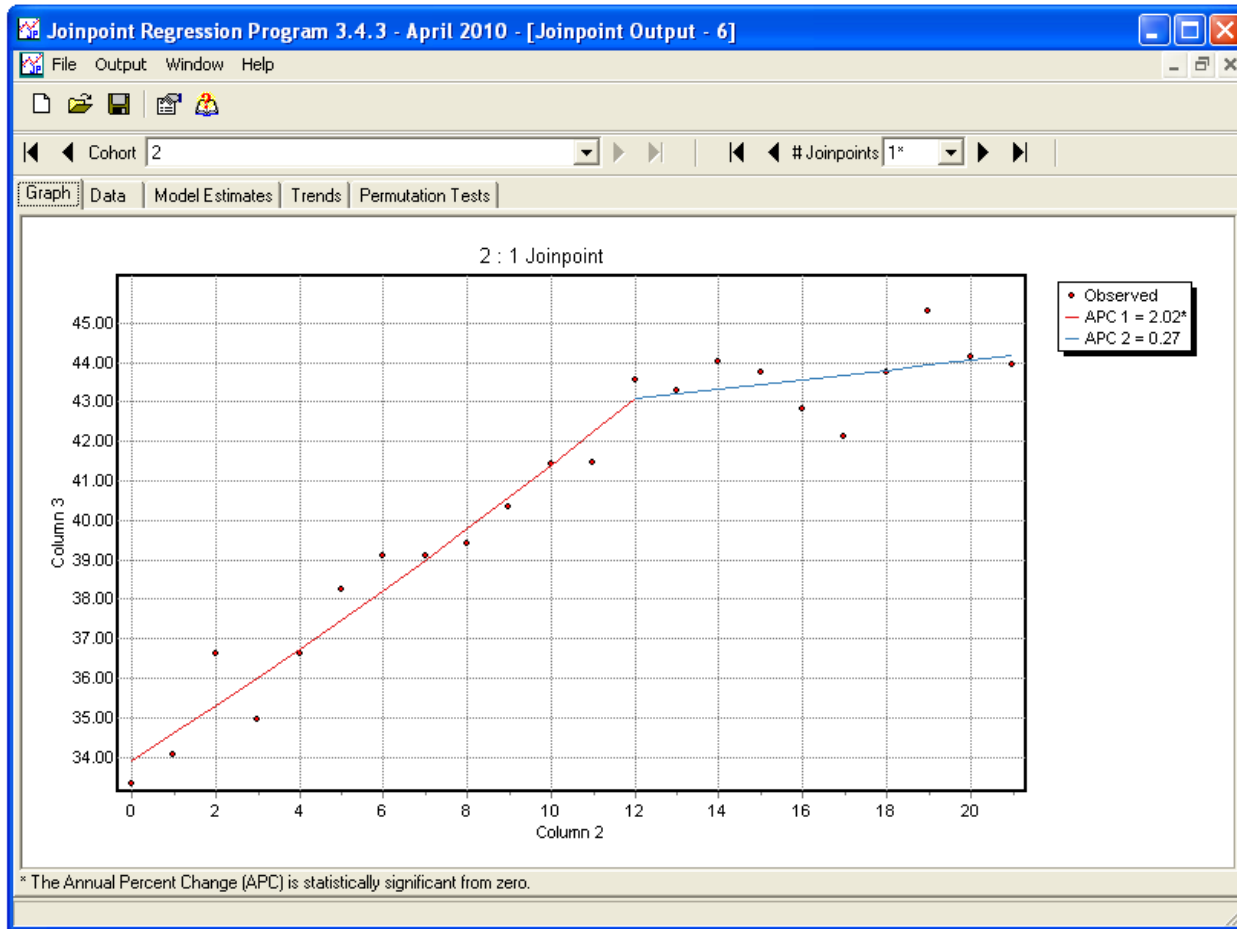
Grid Search

- With the default settings, joinpoints must occur exactly at an observation. This does not, however, always find the best fit.
- A better fit can be achieved by using a finer grid – by changing the setting for “Number of points to place between adjacent observed x values in the grid search” to something larger than the default of zero.
- With lower values for “Number of points to place between...”, this method is computationally more efficient.

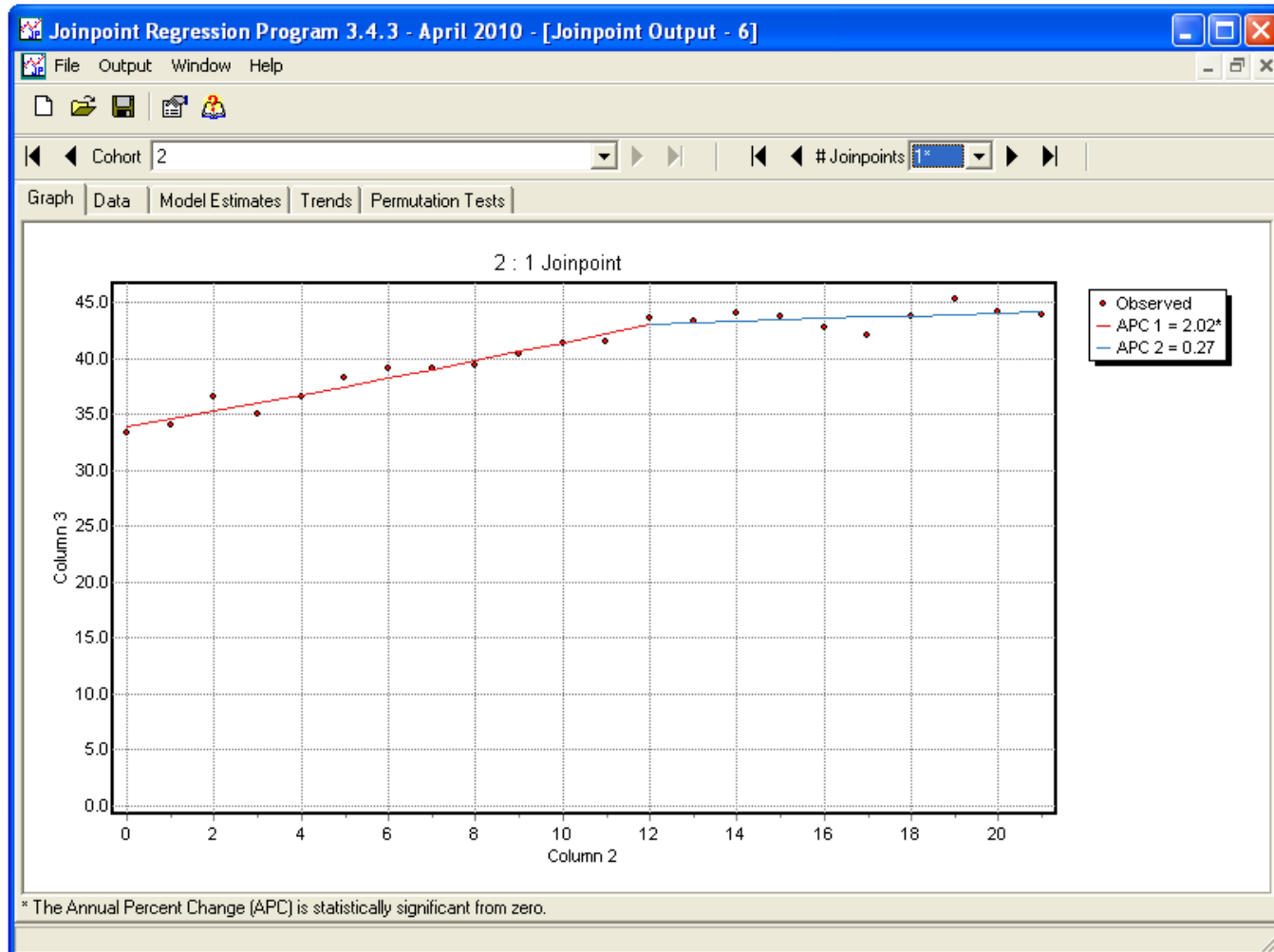
Hudson's Method

- Does a continuous testing between observed x values to find the best model, so the fit will be better than even a fine grid of 9
- *Much* more computationally intensive than the Grid Search with zero "points between", but it is faster than a very fine grid while also achieving a better fit
- Note, since the fit is better for each model and the SSEs are lower, this can impact which joinpoint model is selected as the best one

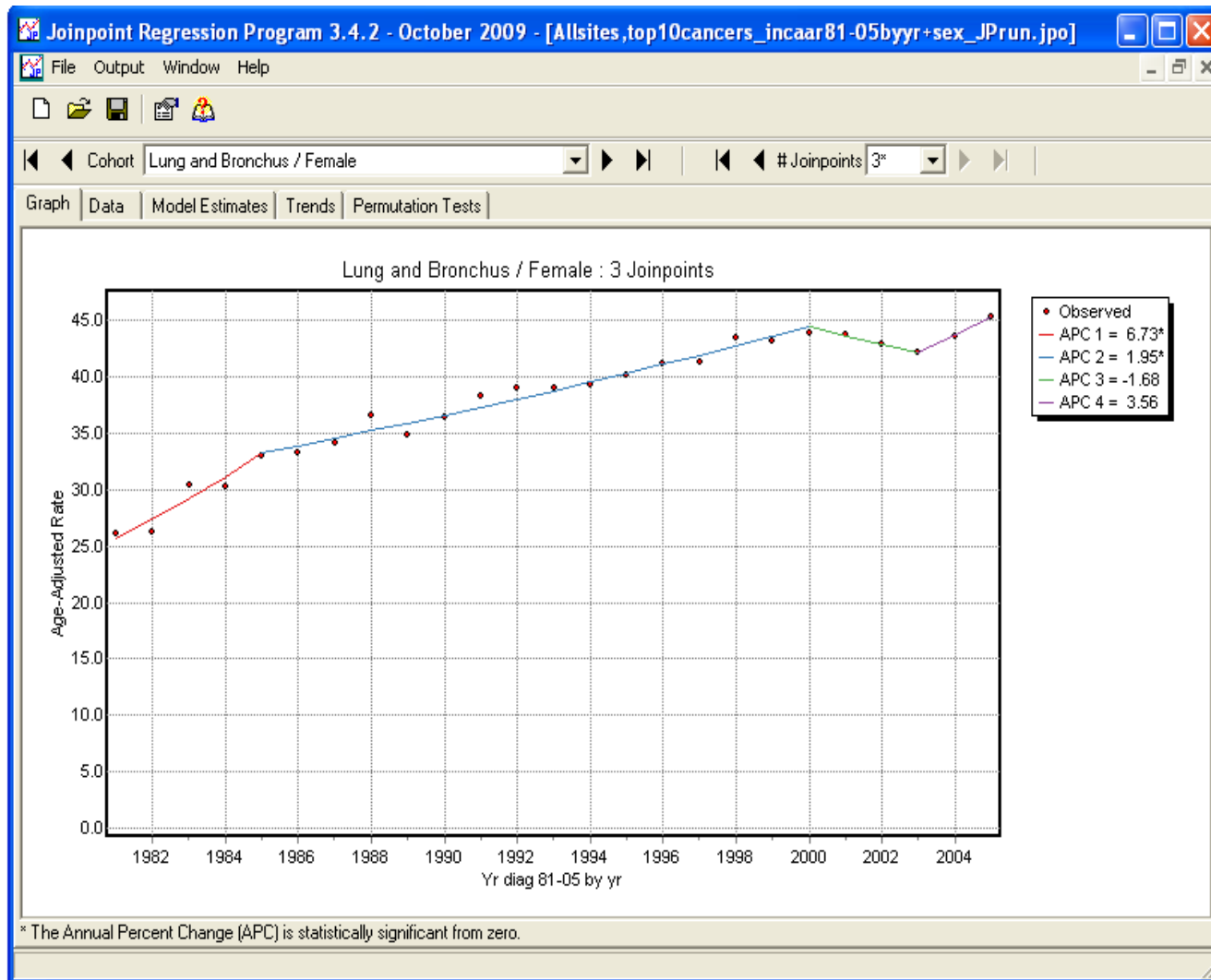
Joinpoint default output



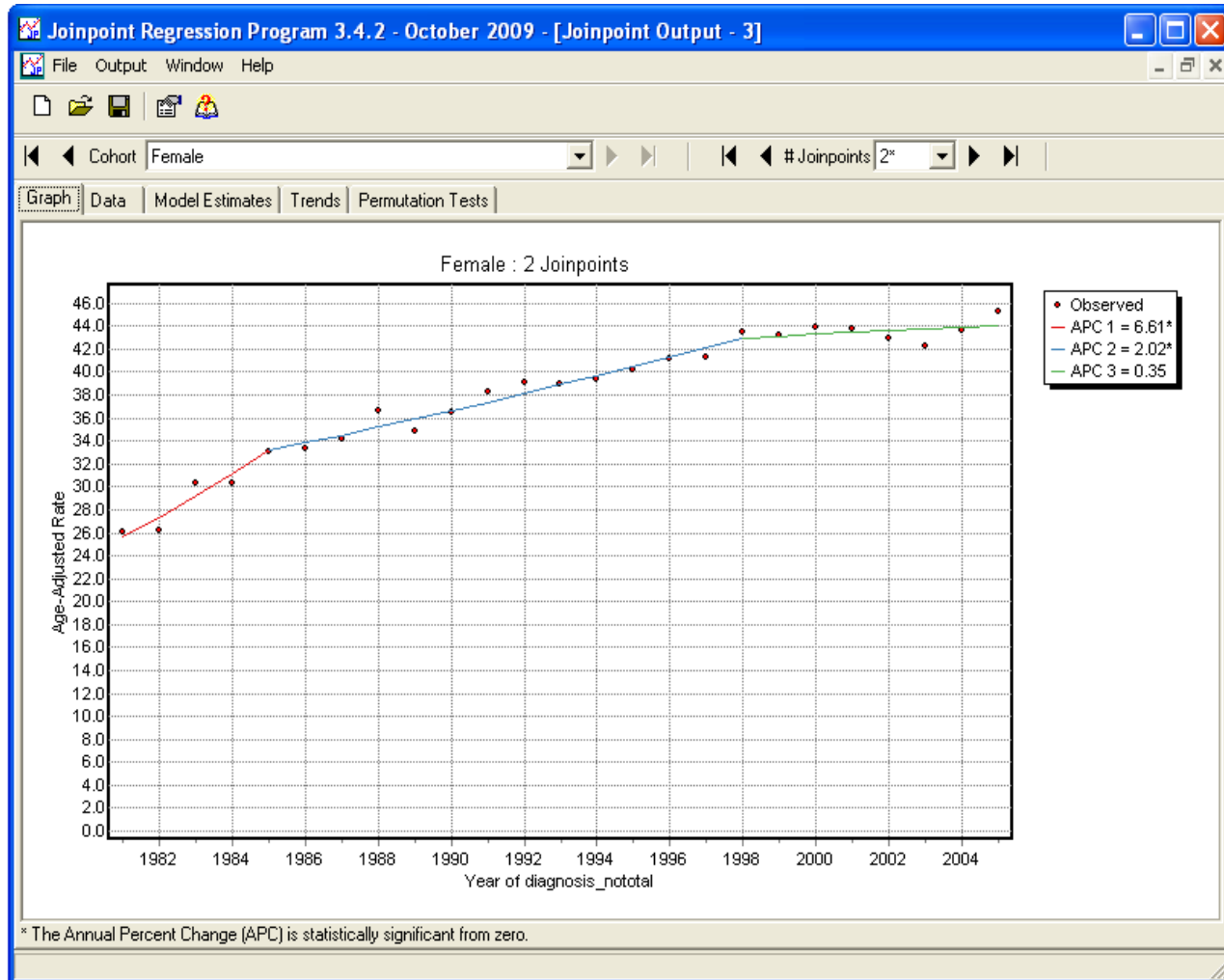
Output -> Options... to fix y-axis!



Female lung cancer incidence, 1981-2005

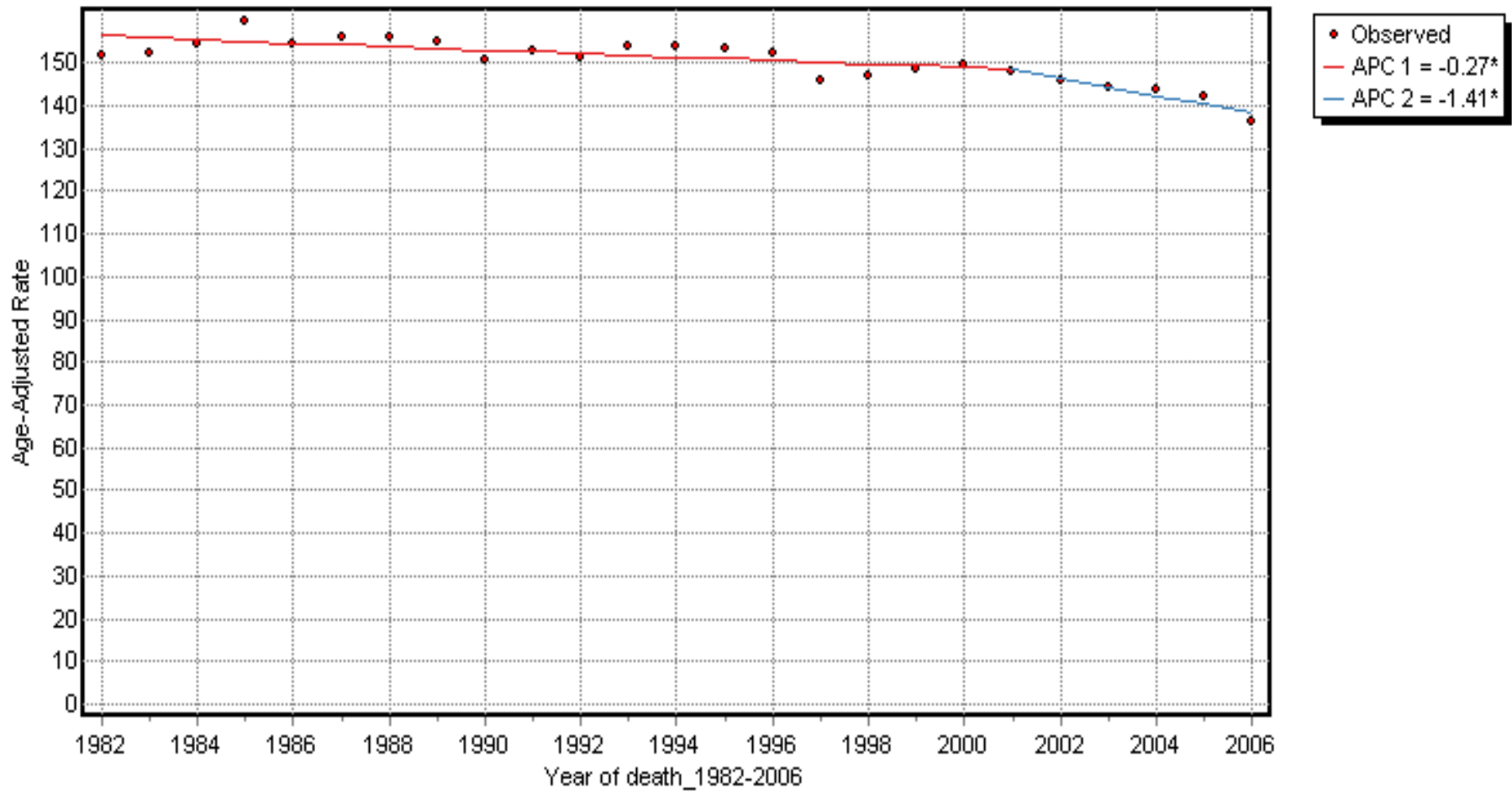


Change min obs from end to 5



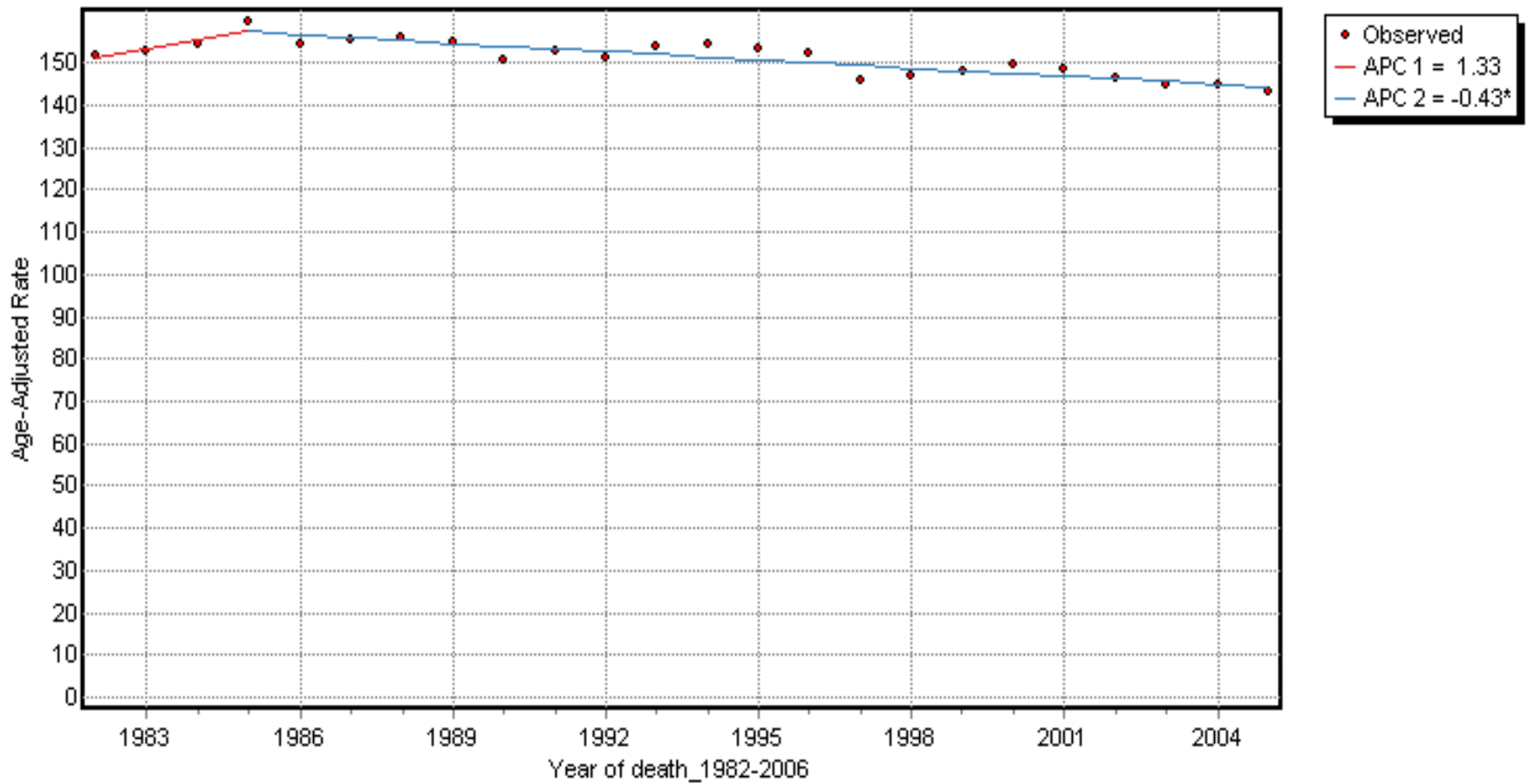
All cancer mortality

All Malignant Cancers / Female : 1 Joinpoint

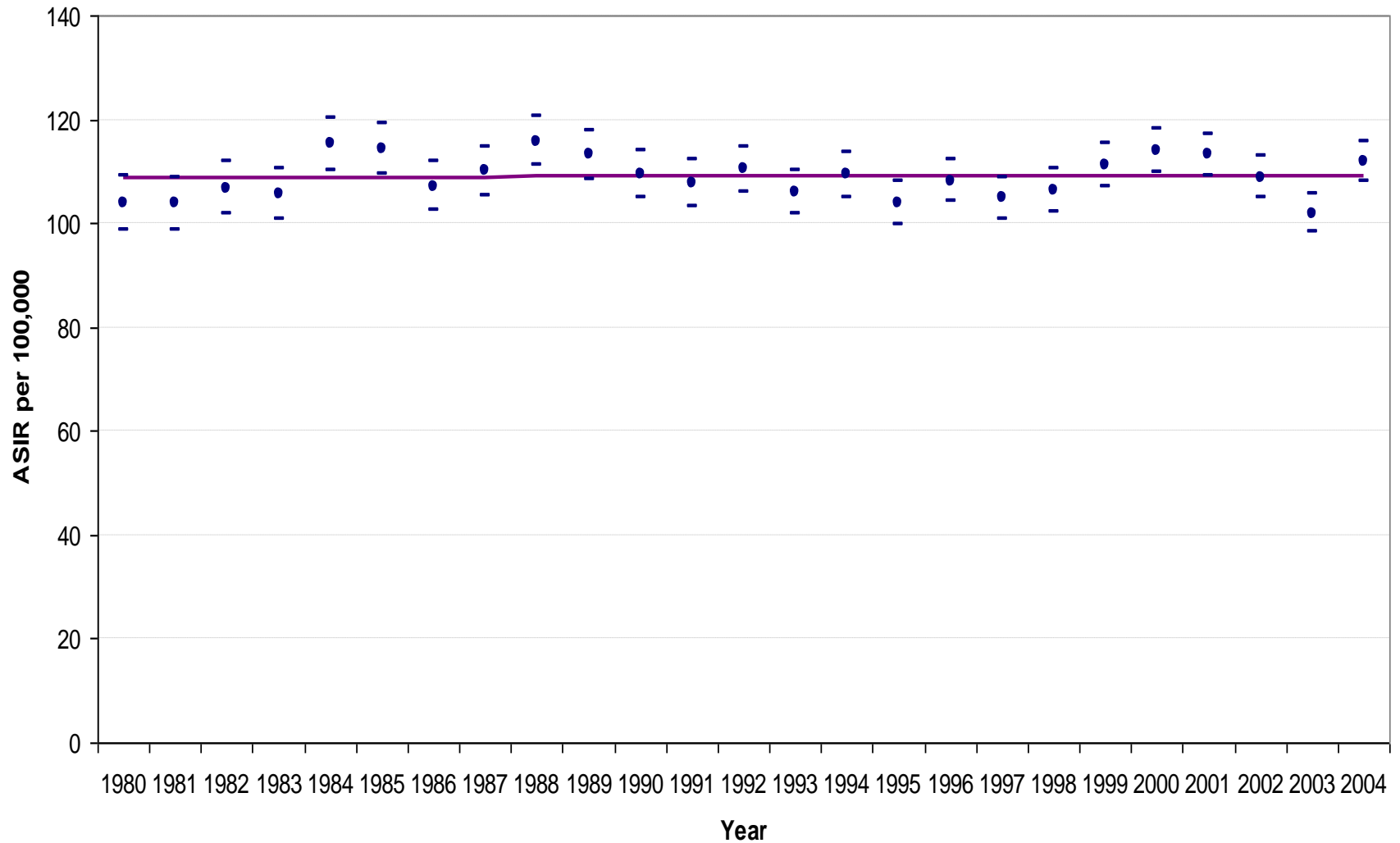


All cancer mortality, removing 2006

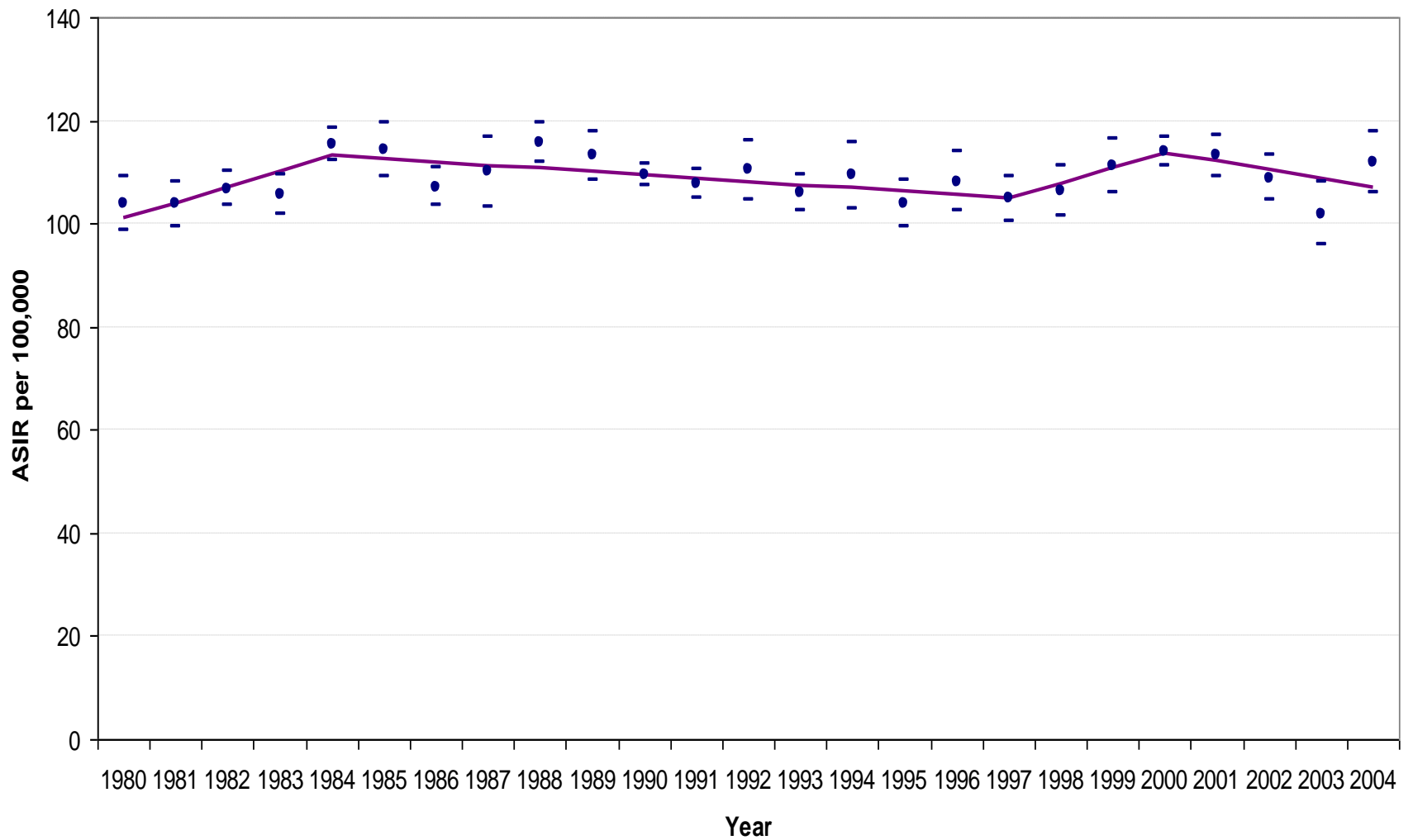
All malignant cancers : Female : 1 Joinpoint



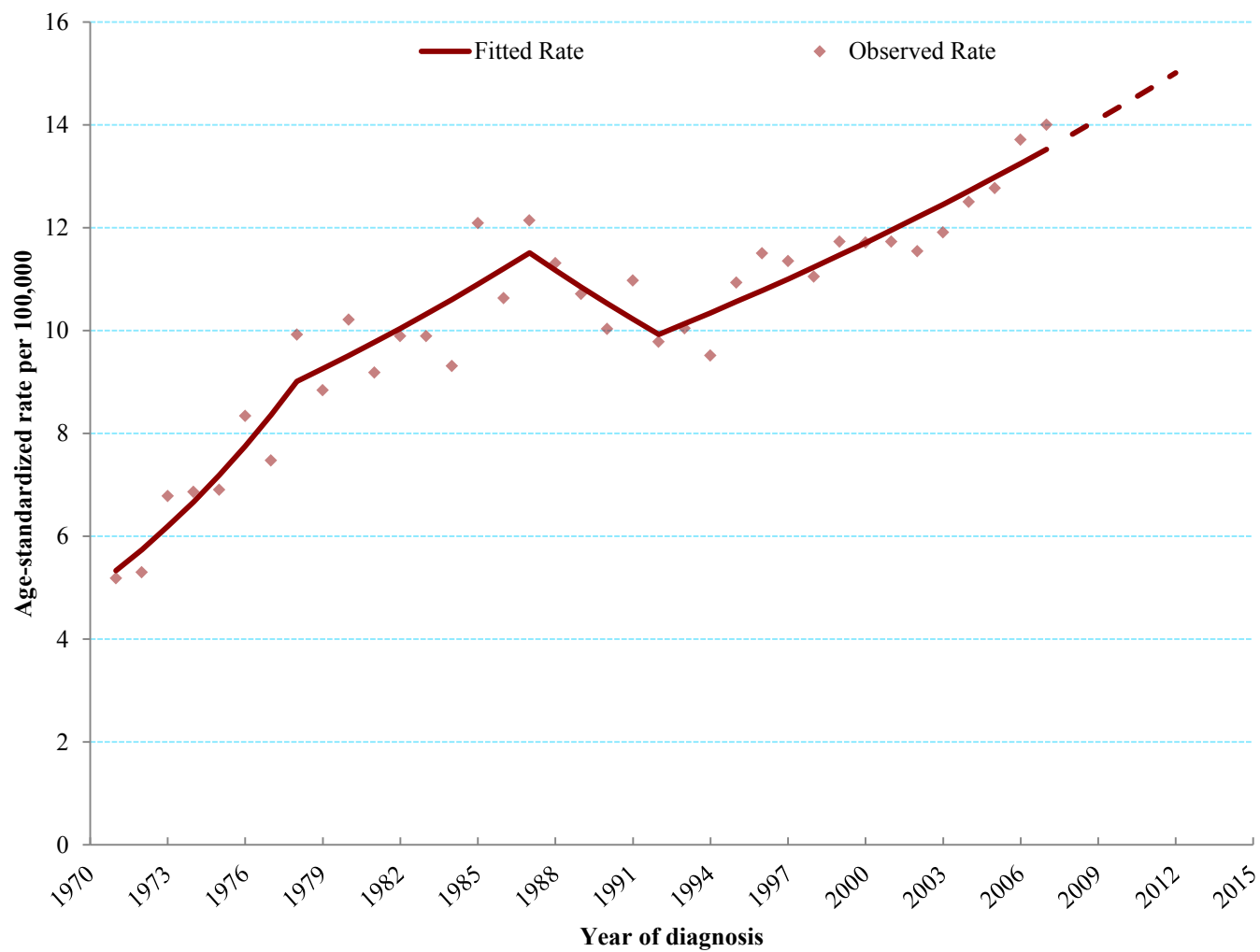
Male colorectal incidence, age group?



Same data, different errors



Incidence rates for melanoma, Ontario, females



Source: Cancer Care Ontario (Ontario Cancer Registry, 2010)

Final Musing

“The only value in a Cancer Registry is in its use.”

Dr. Calum Muir

Deputy Director, IARC

Please use our data

Please tell us when you find problems with the data

Please tell us when you are publishing your results

Exercise #3: Joinpoint

- Examine the incidence trends for one or more of the following cancer definitions:
 - All cancers, by sex
 - Top 4 cancers, sexes combined
 - Top 4 cancers, by sex
- Try both SEER*Stat and Joinpoint trend analyses, if time permits; otherwise choose one method
- Do any of your results differ from Ontario's trend?