National Cancer Institute

Cancer Etiology and Natural History: A Web Tool for Age-Period-Cohort Analysis

Philip S. Rosenberg

David P. Check

William F. Anderson

&

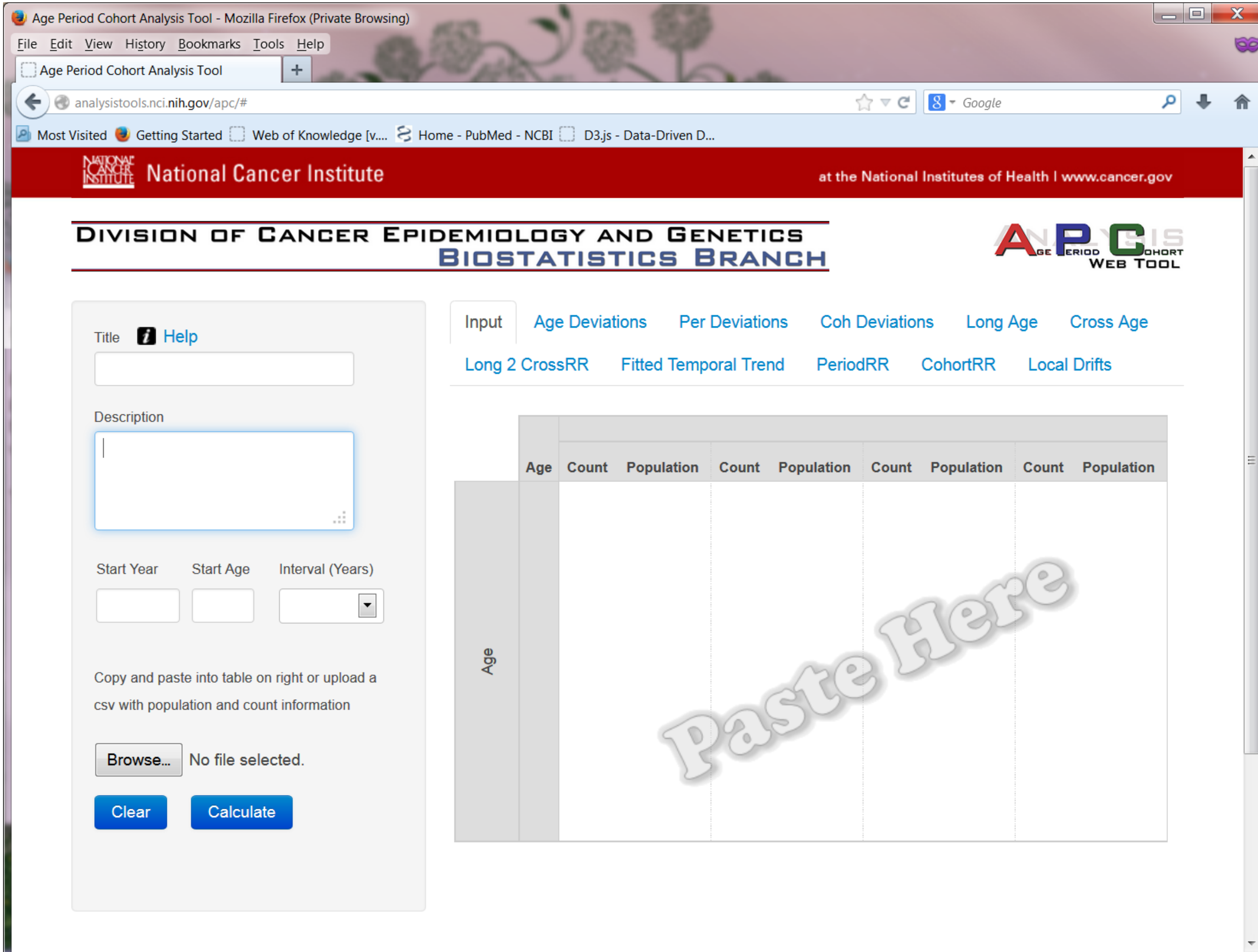
The NCI Center for Biomedical Informatics and Information Technology (CBIIT)

Biostatistics Branch
Division of Cancer Epidemiology and Genetics
National Cancer Institute

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

National Institutes of Health

06 March 2014, DCEG Seminar


Age Period Cohort Analysis Tool - Mozilla Firefox (Private Browsing)
File Edit View History Bookmarks Tools Help
Age Period Cohort Analysis Tool
analysistools.nci.nih.gov/apc/#
Google
Most Visited
Getting Started
Web of Knowledge [v....
Home - PubMed - NCBI
D3.js - Data-Driven D...
National Cancer Institute
at the National Institutes of Health I www.cancer.gov
Division of Cancer Epidemiology and Genetics
Biostatistics Branch
Age Period Cohort
Web Tool
Title
Help
Description
Start Year
Start Age
Interval (Years)
Copy and paste into table on right or upload a csv with population and count information
Browse...
No file selected.
Clear
Calculate
Input
Age Deviations
Per Deviations
Coh Deviations
Long Age
Cross Age
Long 2 CrossRR
Fitted Temporal Trend
PeriodRR
CohortRR
Local Drifts
Age
Count
Population
Count
Population
Count
Population
Count
Population
Age
Paste Here

Age Period Cohort Analysis Web Tool

- *What does it do?*
- *Why is that important?*
- *Who built it, and how?*
- *How do I use it?*

Acknowledgments

CBIIT Center for Biomedical Informatics and Information Technology

- **Robert Shirley**, NCI CBIIT
- **Sue Pan**, NCI CBIIT
- **Larry Brem**, Leidos Biomedical Research, Inc. (NCI CBIIT Dev Team Contractor)
- **Brent Coffey**, Leidos Biomedical Research, Inc. (NCI CBIIT Dev Team Contractor)
- **Shaun Einolf**, Leidos Biomedical Research, Inc. (NCI CBIIT Dev Team Contractor)
- **Sula Rajapakse**, Leidos Biomedical Research, Inc. (NCI CBIIT Dev Team Contractor
- **Cuong Nguyen**, SRA International, Inc. (NCI CBIIT Systems Team Contractor)

Division of Cancer Epidemiology & Genetics

Discovering the causes of cancer and the means of prevention

- **Carl McCabe,** Office of Division Operations and Analysis
- **Sholom Wacholder, Nicolas Wentzensen, Christine Fermo**
  - http://analysistools.nci.nih.gov/meanstorisk/

Outline

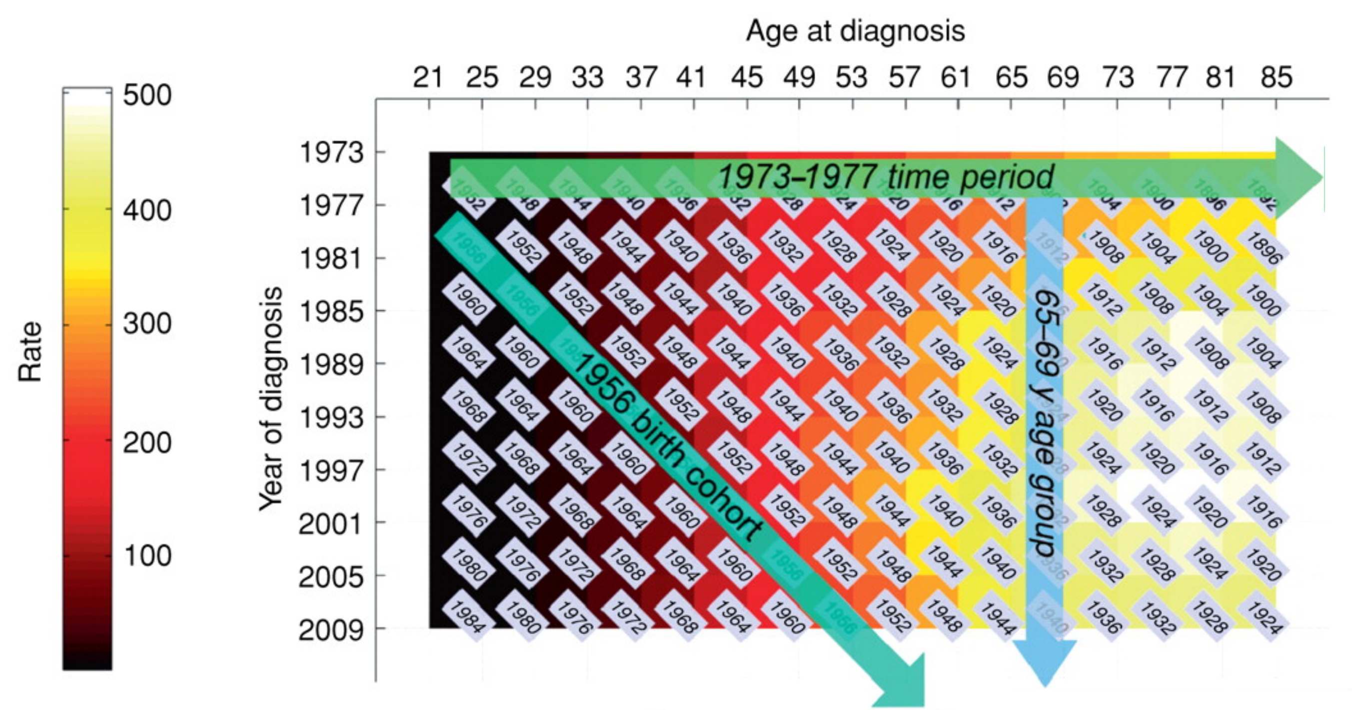
- The **A****P****C** Model
  - *Overview*
  - *Examples from the literature*
- The Web Tool
  - *What's in it (and why)*
  - *How it works*

APC Model: Overview

- Macro-epidemiological model for population-based cancer surveillance data
  - Incidence and Mortality
  - **SEER**, **IARC**, other large-scale open cohorts
- Parametric approach
  - complements traditional descriptive approaches
- Quantification (via ***parameters*** and ***functions***)
  - Burden
  - Trends
  - Natural History
  - Etiology
  - Disparity

APC Model: Data

A registry is a cohort of cohorts . . .



We can study patterns both cross-sectionally (over time) and longitudinally (by cohort).

Rate matrix or Lexis diagram for invasive female breast cancer.

Rosenberg P S , and Anderson W F Cancer Epidemiol Biomarkers Prev 2011;20:1263-1268

Cancer Epidemiology, Biomarkers & Prevention

APC Model: Parameters from Data

Longitudinal Form

$$\rho_{ac} = \mu + (\alpha_L + \pi_L)(a - \bar{a}) + (\pi_L + \gamma_L)(c - \bar{c}) + \tilde{\alpha}_a + \tilde{\pi}_p + \tilde{\gamma}_c$$

Longitudinal Age Trend

Net Drift

Deviations

Cross-sectional Form

$$\rho_{ap} = \mu + (\alpha_L - \gamma_L)(a - \bar{a}) + (\pi_L + \gamma_L)(p - \bar{p}) + \tilde{\alpha}_a + \tilde{\pi}_p + \tilde{\gamma}_c$$

Cross-Sectional Age Trend

APC Model: Putting the pieces together

Through independent and collaborative descriptive studies, we developed a panel of standard and novel **functions**** and corresponding **hypothesis tests** that appear to be effective in identifying *signatures* or *patterns* in disease rates for many types of cancers.

** linear combination of estimable parameters in the APC Model

APC Model: Key Parameters, Functions, and Tests

- **Net Drift is the single most important parameter!**
  - Model analogue of EAPC adjusted for cohort effects
  - Determines ratio of Longitudinal to Cross-sectional Age Curves

- **Age effects** (Longitudinal and Cross-sectional Age Curves)
- **Period effects** (Fitted Temporal Trends, Period RR)
- **Cohort effects** (Cohort RR; Local Drifts = age-specific EAPC)

- **The Significance Test for Local Drifts is the second most important APC statistic!**
  - Tells you if you have important cohort effects

Outline

- The **APC** Model
  - *Overview*
  - ***Examples from the literature***
- The Web Tool
  - *What's in it (and why)*
  - *How it works*

uicc global cancer control
IJC International Journal of Cancer

Trimodal age-specific incidence patterns for Burkitt lymphoma in the United States, 1973–2005

Sam M. Mbulaiteye[1], William F. Anderson[2], Kishor Bhatia[1], Philip S. Rosenberg[2], Martha S. Linet[3] and Susan S. Devesa[2]

Age Effects (Longitudinal Age Curve)

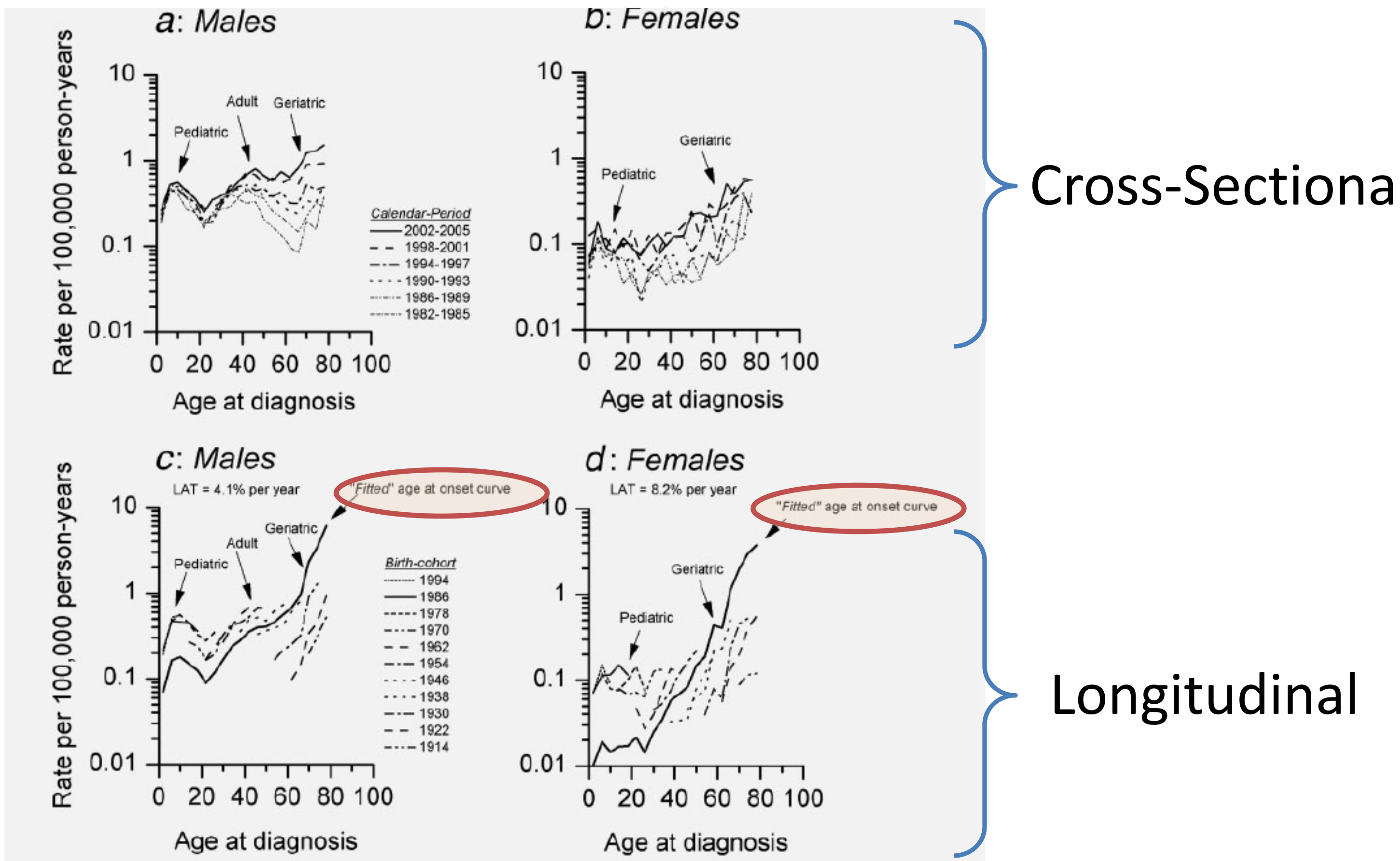


Figure 2. Burkitt lymphoma APC model-based expected period- and cohort-specific age-specific incidence rates by sex, SEER 9, 13 and17 registries, 1982–2005. Panels *c* and *d* included "fitted" age-at-onset curves (see Methods). Cases diagnosed during 1973–1981 were excluded because of sparse numbers. LAT = longitudinal age trend.

Period Effects

(Deviations, Period RR)

VOLUME 31 · NUMBER 17 · JUNE 10 2013

JOURNAL OF CLINICAL ONCOLOGY | ORIGINAL REPORT

Ovarian Cancer Incidence Trends in Relation to Changing Patterns of Menopausal Hormone Therapy Use in the United States

Hannah P. Yang, William F. Anderson, Philip S. Rosenberg, Britton Trabert, Gretchen L. Gierach, Nicolas Wentzensen, Kathleen A. Cronin, and Mark E. Sherman

All authors: National Cancer Institute, Bethesda, MD.

Published online ahead of print at www.jco.org on May 6, 2013.

Supported entirely by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Insti-

A B S T R A C T

Purpose
After a report from the Women's Health Initiative (WHI) in 2002, a precipitous decline in menopausal hormonal therapy (MHT) use in the United States was linked to a decline in breast cancer incidence rates. Given that MHT use is also associated with increased ovarian cancer risk, we tested whether ovarian cancer incidence rates changed after 2002.

Ovarian Cancer Incidence and Menopausal Hormone Therapy Use Pattern

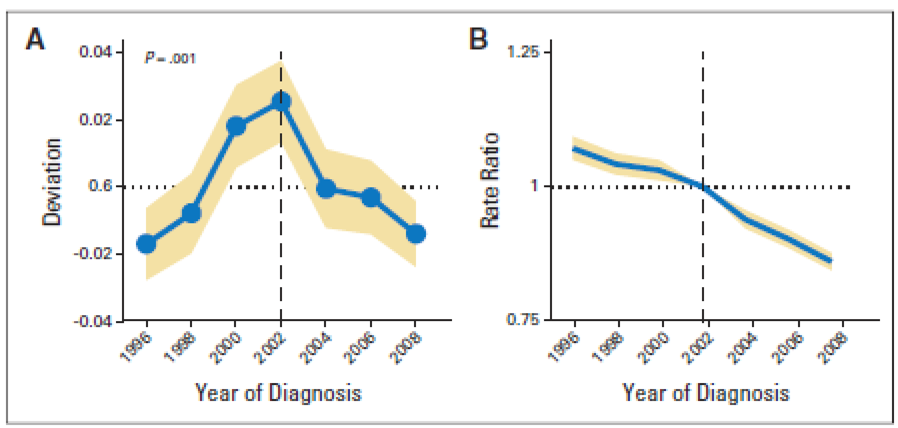


Fig 2. Age-period-cohort effects among US women age ≥ 50 years (North American Association of Central Cancer Registries Incidence, 1995 to 2008). Point estimates are shown in blue, with 95% CIs shaded in gold. (A) The significance of period deviations was assessed by contrasting the three time periods before the Women's Health Initiative (WHI; 1995 to 1996, 1997 to 1998, and 1999 to 2000) with the three time periods after WHI (2003 to 2004, 2005 to 2006, and 2007 to 2008). *P* value is for change in the slopes of the period deviations, adjusted for age and cohort effects. (B) Period relative risks were calculated as rate ratios adjusted for age and birth cohort effects, comparing the ovarian cancer incidence rates for a given time period with the rate of a referent period (the 2002 time period in this analysis). The period relative risks declined from more than 1.0 before the 2002 referent period, after which the period relative risks were significantly less than 1.0.

VOLUME 30 · NUMBER 22 · AUGUST 1 2012

JOURNAL OF CLINICAL ONCOLOGY — ORIGINAL REPORT

Cohort Effects

(Cohort RR)

Increasing Lung Cancer Death Rates Among Young Women in Southern and Midwestern States

Ahmedin Jemal, Jiemin Ma, Philip S. Rosenberg, Rebecca Siegel, and William F. Anderson

Ahmedin Jemal, Jiemin Ma, and Rebecca Siegel, Surveillance Research Program, American Cancer Society, Atlanta, GA; and Philip S. Rosenberg and William F. Anderson, National Cancer Institute, Rockville, MD.

Submitted February 23, 2012; accepted May 7, 2012; published online ahead of print at www.jco.org on June 25, 2012.

A B S T R A C T

Purpose
Previous studies reported that declines in age-specific lung cancer death rates among women in the United States abruptly slowed in women younger than age 50 years (ie, women born after the 1950s). However, in view of substantial geographic differences in antitobacco measures and sociodemographic factors that affect smoking prevalence, it is unknown whether this change in the trend was similar across all states.

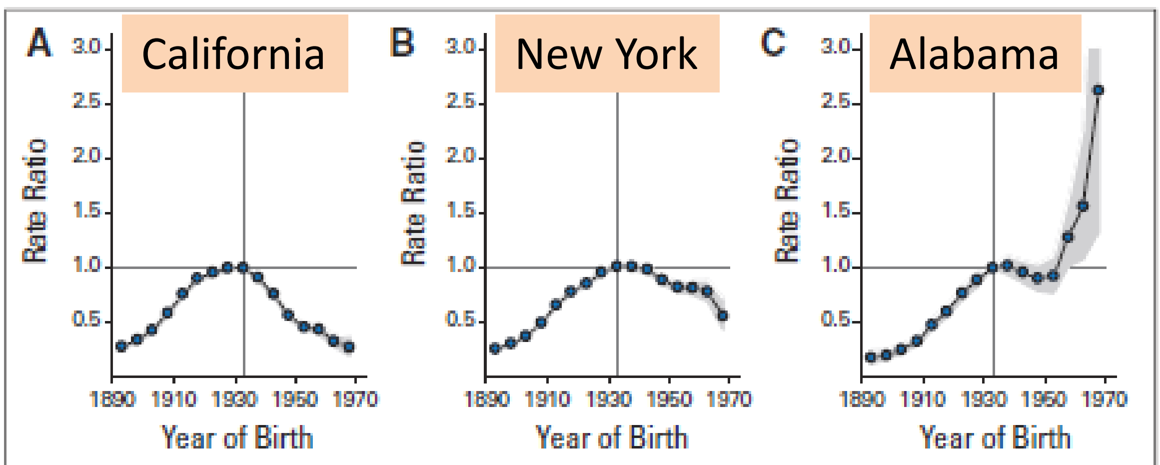


Fig 2. Rate ratios of lung cancer death rates according to birth cohort among white women for (A) California, (B) New York, and (C) Alabama. The reference group is the 1933 birth cohort, and the shaded areas denote the 95% point-wise CIs of rate ratios.

uicc global cancer control

IJC International Journal of Cancer

Divergent estrogen receptor-positive and -negative breast cancer trends and etiologic heterogeneity in Denmark

William F. Anderson[1]*, Philip S. Rosenberg[1]*, Lucia Petito[1], Hormuzd A. Katki[1], Bent Ejlertsen[2], Marianne Ewertz[3], Birgitte B. Rasmussen[4], Maj-Britt Jensen[2] and Niels Kroman[5]

Cohort

Effects

(Local Drifts)

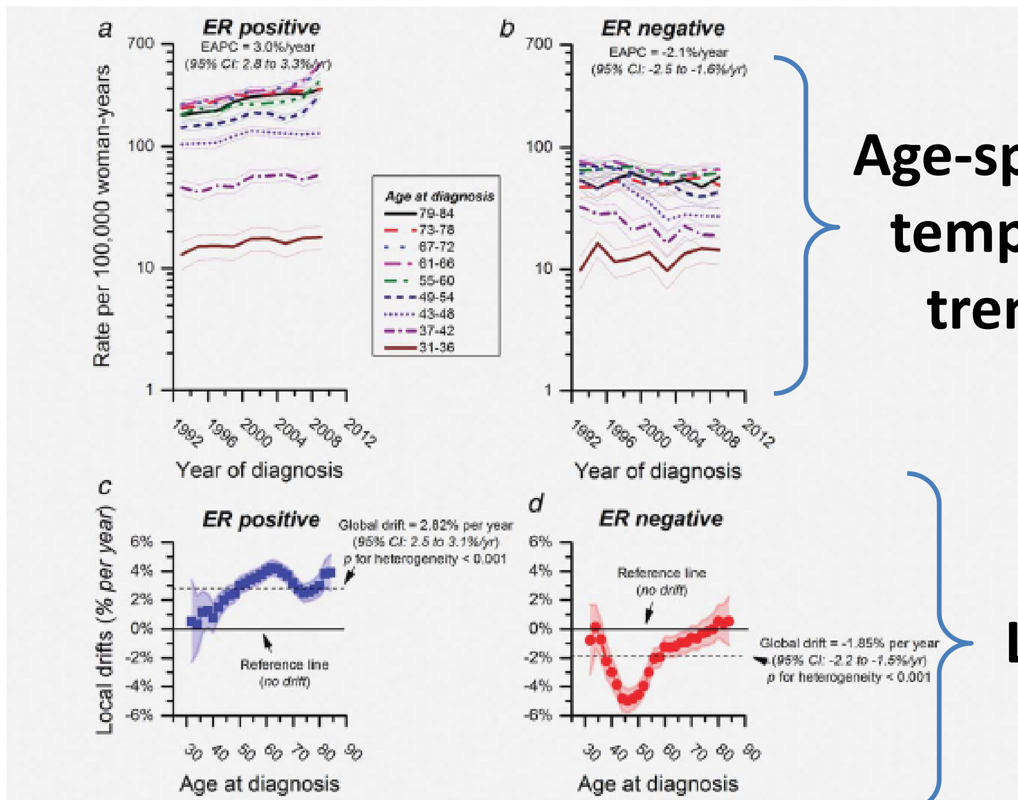


Figure 2. ER-positive age-specific incidence trends (*a*), ER-negative age-specific incidence trends (*b*), ER-positive global and age-specific (local) net drifts (*c*) and ER-negative global and age-specific (local) net drifts (*d*). The global net drift is analogous to the estimated annual percentage change (EAPC) in the age-standardized incidence rate, whereas the local net drifts provide estimates of the corresponding EAPCs for individual age groups. See text for details. The *p* value for heterogeneity tests the null hypothesis that all local drifts are no different than the global drift.

Age-specific temporal trends

Local Drifts

Outline

- The **APC** Model
  - *Overview*
  - *Examples from the literature*
- The Web Tool
  - ***What's in it (and why)***
  - *How it works*

Age Period Cohort Analysis Web Tool

All of the **APC** functions (and corresponding hypothesis tests) shown in *Examples from the literature* are produced by the Web Tool.

Key Functions

Age Period Cohort Analysis Tool - Mozilla Firefox (Private Browsing)

analysistools-dev.nci.nih.gov/apc/help.html

The most important functions calculated by the web tool are summarized in the Table of Key Functions using the following conventions:

- The APC model is defined over *A* age groups and *P* calendar periods with equal intervals.
- The central age group, calendar period, and birth cohort define standard **reference values** *a*, *p* and *c*, respectively.
- When there is an even number of age, period, or cohort categories, the reference value is the lower of the two central values.
- All values labeled **CI Lo** are lower 95% confidence limits. All values labeled **CI Hi** are upper 95% confidence limits.

Table of Key Functions

Nomenclature	Interpretation
Fitted Temporal Trend	Expected age-adjusted rates over time
Net Drift	Annual percentage change of the expected age-adjusted rates over time
Local Drifts	Annual percentage change of the expected age-specific rates over time
Cross-Sectional Age Curve (Cross Age)	Expected age-specific rates within a given calendar period
Longitudinal Age Curve (Long Age)	Expected age-specific rates for a given number of birth cohorts
Period Rate Ratios (PeriodRR)	Ratio of age-specific rates in period relative to reference period
Cohort Rate Ratios (CohortRR)	Ratio of age-specific rates in cohort relative to reference cohort

- Other parameters and functions calculated by the web tool and not listed above are described in greater mathematical detail elsewhere (click here for a PowerPoint presentation).

Hypothesis Tests

Age Period Cohort Analysis Tool - Mozilla Firefox (Private Browsing)

analysistools-dev.nci.nih.gov/apc/help.html

Statistical hypothesis tests calculated by the web tool are summarized in the Table of Hypothesis Tests.

The **Wald Tests** follow a Chi-Square distribution when the **Null Hypothesis** is true. The **df** (degrees of freedom) count the number of free parameters included in each test. The web tool reports P-values; values less than 0.05 are often considered 'statistically significant', meaning there is statistical evidence that the **Null Hypothesis** is unlikely to be correct.

Table of Hypothesis Tests

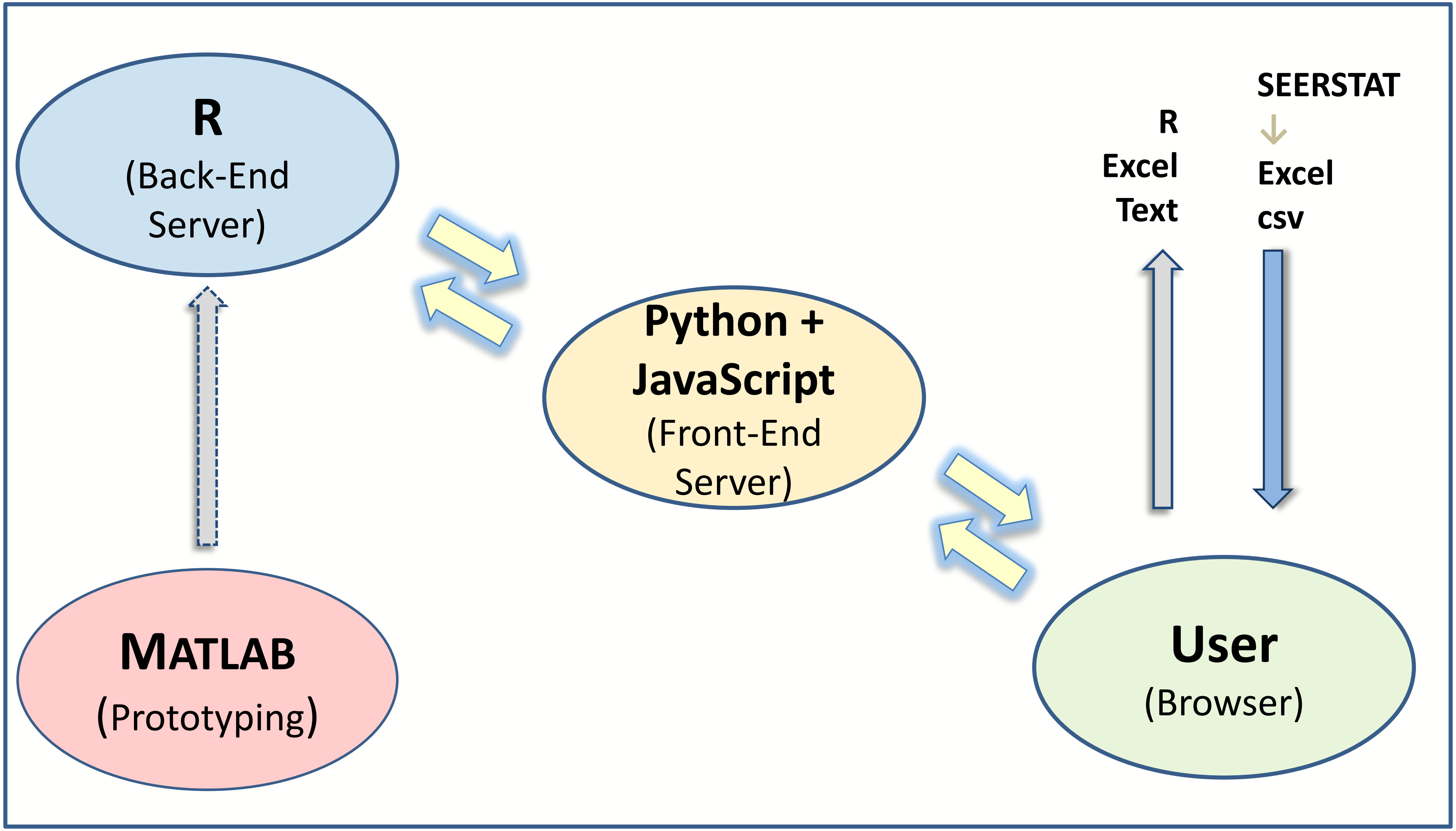
Null Hypothesis	Implications	Degrees of Freedom
Net drift = 0	Fitted temporal trends are stable (i.e., flat with no change) over time. Fitted longitudinal and cross-sectional age curves are equal.	1
All age deviations = 0	Fitted longitudinal and cross-sectional age curves are log-linear (i.e., log-additive).	*A* - 2
All period deviations = 0	Fitted temporal trends and period rate ratios are log-linear (i.e., log-additive).	*P* - 2
All cohort deviations = 0	Cohort rate ratios are log-linear; all local drifts equal the net drift.	*C* - 2
All period rate ratios = 1	Net drift is 0 and fitted temporal trends are constant; Cross-sectional age curve describes age incidence pattern in every period.	*P* - 1
All cohort rate ratios = 1	Net drift is 0 and all local drifts are 0; Longitudinal age curve describes age incidence pattern in every cohort.	*C* - 1
All local drifts = the net drift	Temporal patterns are the same in every age group.	*A*

*For APC model defined over *A* age groups, *P* calendar periods, and *C* = *P* + *A* - *1* birth cohorts.

Web Tool: Usability

- We paid a lot of attention to **workflow.**
- The Web Tool promotes **reproducible research**.
- We think it is really simple to use.

Web Tool: Architecture



Data Input

Age Period Cohort Analysis Tool - Mozilla Firefox (Private Browsing)

analysistools-dev.nci.nih.gov/apc/help.html

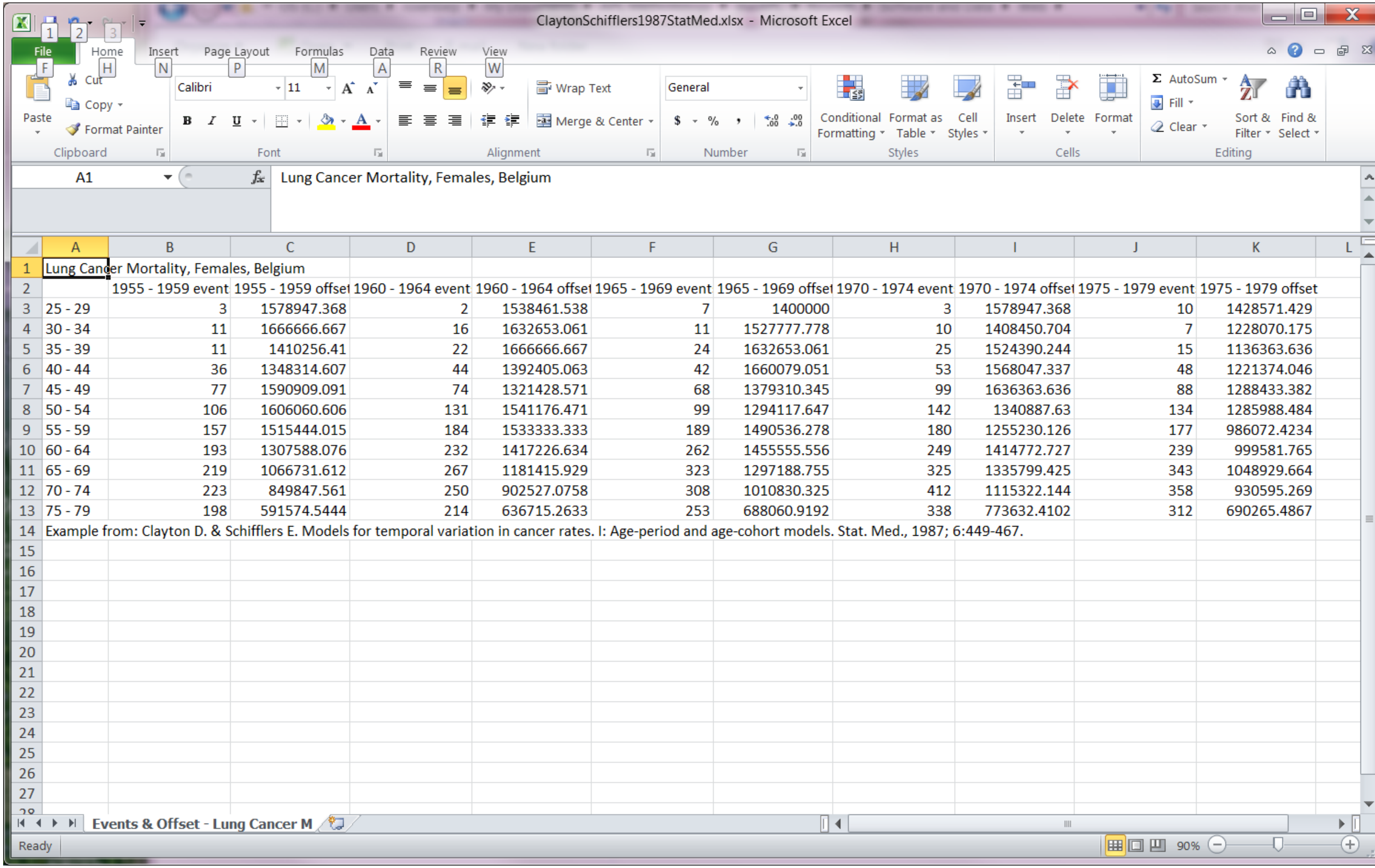
Getting started

Input data for the web tool consist of **Count and Population** data for particular age groups over calendar time, in the form of a matrix of rows with paired columns. Rows correspond to particular age groups and columns correspond to calendar time periods. The age and period intervals must all be equal, i.e. if 5-year age groups are used then 5-year calendar periods must also be used. The data can be input by copy-and-paste from an Excel worksheet or file upload of a comma-separated-values (csv) file.

To input from Excel:

1. Copy the paired columns of data you want to analyze from your spreadsheet, right-click inside the empty matrix on the Input tab, and paste your selection.
2. Fill in the information (meta-data) on the left hand side of the Input page:
   - **Title** - describe your data
   - **Description** - add optional details
   - **Start Year** - list the first calendar year of the first calendar period of your data, for example, use 1990 for the interval 1990 - 1994
   - **Start Age** - list the first age of the first age group of your data, for example use 30 for the interval 30 - 34
   - **Interval (Years)** - the width of the age and period intervals, for example use 1 for single-year data, 2 for two-year data, 5 for five-year data (e.g., 1990 - 1994), etc.
3. Click the calculate button

Data Input Using Excel



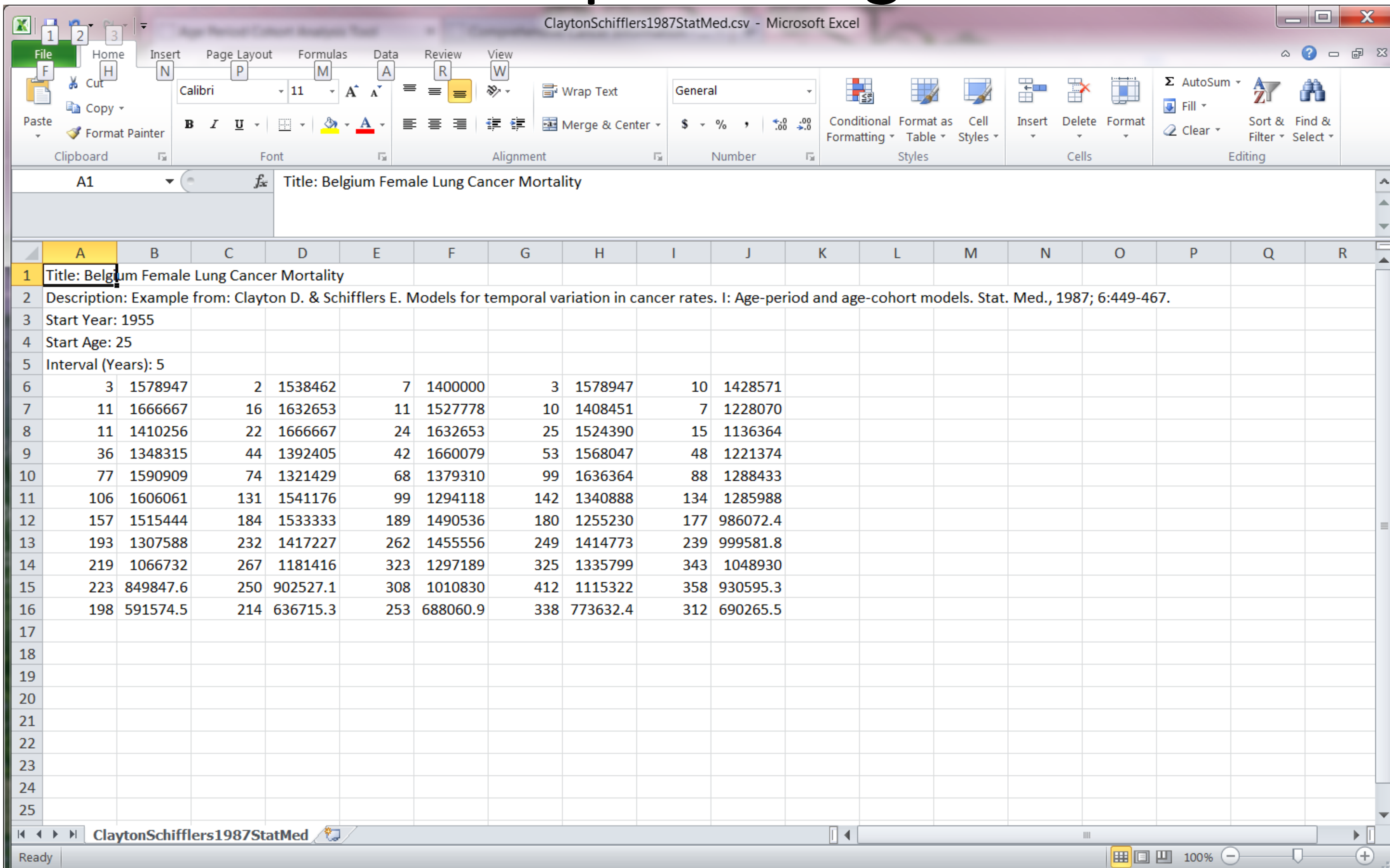
ClaytonSchifflers1987StatMed.xlsx - Microsoft Excel

A1: Lung Cancer Mortality, Females, Belgium

	A	B	C	D	E	F	G	H	I	J	K
1	Lung Cancer Mortality, Females, Belgium										
2		1955 - 1959 event	1955 - 1959 offset	1960 - 1964 event	1960 - 1964 offset	1965 - 1969 event	1965 - 1969 offset	1970 - 1974 event	1970 - 1974 offset	1975 - 1979 event	1975 - 1979 offset
3	25 - 29	3	1578947.368	2	1538461.538	7	1400000	3	1578947.368	10	1428571.429
4	30 - 34	11	1666666.667	16	1632653.061	11	1527777.778	10	1408450.704	7	1228070.175
5	35 - 39	11	1410256.41	22	1666666.667	24	1632653.061	25	1524390.244	15	1136363.636
6	40 - 44	36	1348314.607	44	1392405.063	42	1660079.051	53	1568047.337	48	1221374.046
7	45 - 49	77	1590909.091	74	1321428.571	68	1379310.345	99	1636363.636	88	1288433.382
8	50 - 54	106	1606060.606	131	1541176.471	99	1294117.647	142	1340887.63	134	1285988.484
9	55 - 59	157	1515444.015	184	1533333.333	189	1490536.278	180	1255230.126	177	986072.4234
10	60 - 64	193	1307588.076	232	1417226.634	262	1455555.556	249	1414772.727	239	999581.765
11	65 - 69	219	1066731.612	267	1181415.929	323	1297188.755	325	1335799.425	343	1048929.664
12	70 - 74	223	849847.561	250	902527.0758	308	1010830.325	412	1115322.144	358	930595.269
13	75 - 79	198	591574.5444	214	636715.2633	253	688060.9192	338	773632.4102	312	690265.4867
14	Example from: Clayton D. & Schifflers E. Models for temporal variation in cancer rates. I: Age-period and age-cohort models. Stat. Med., 1987; 6:449-467.										

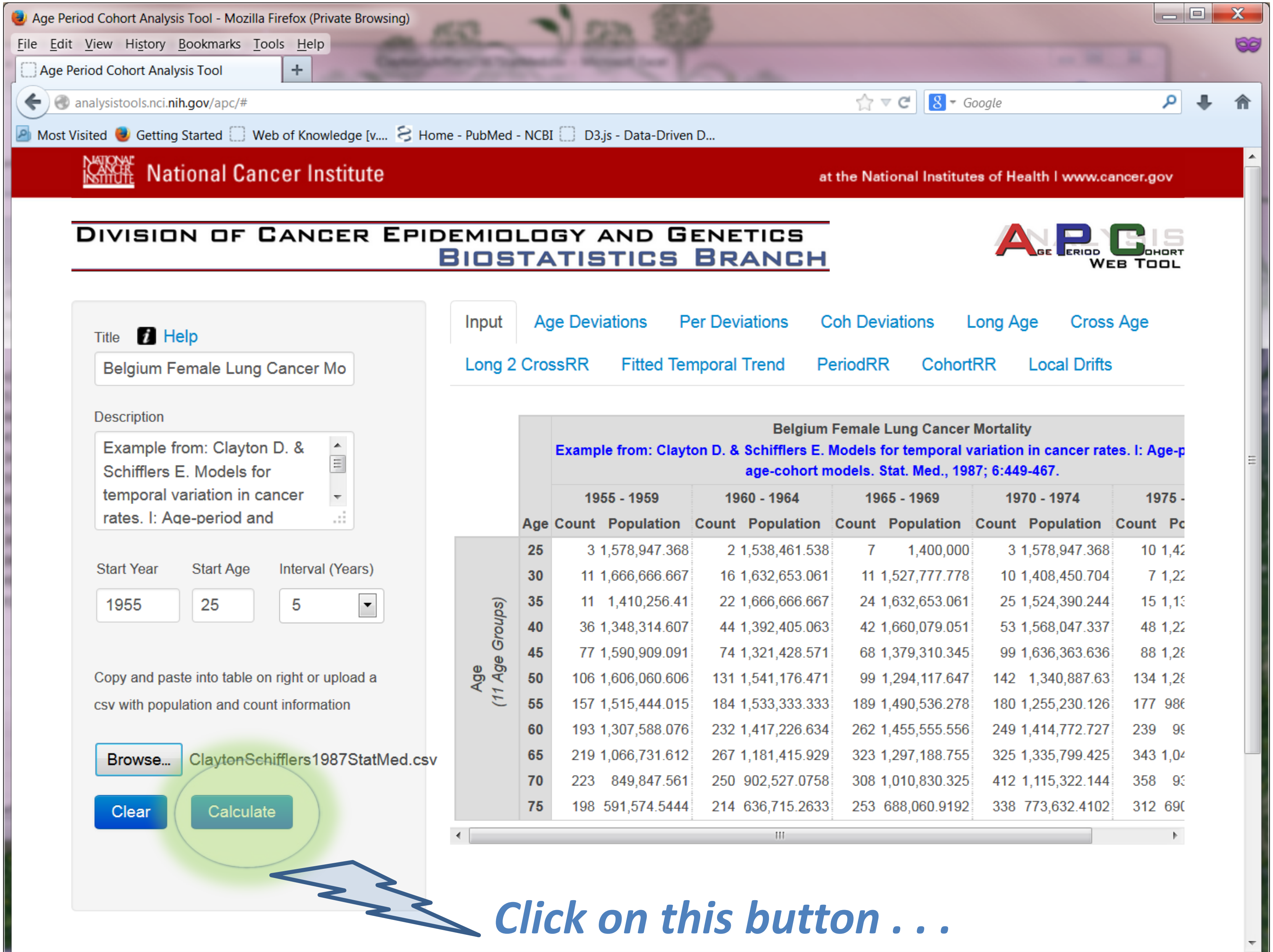
Events & Offset - Lung Cancer M

Data Input Using CSV



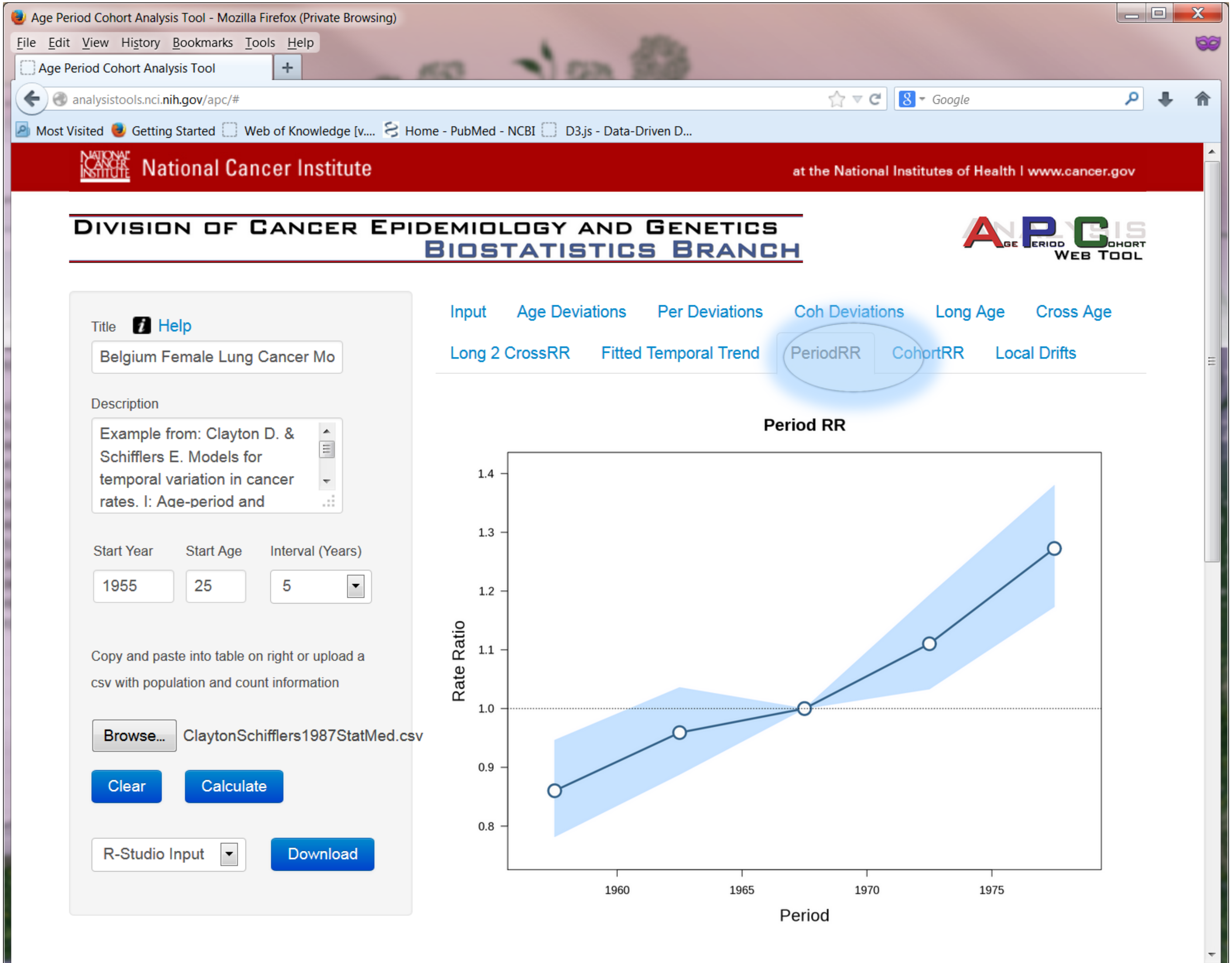
Title: Belgium Female Lung Cancer Mortality
Description: Example from: Clayton D. & Schifflers E. Models for temporal variation in cancer rates. I: Age-period and age-cohort models. Stat. Med., 1987; 6:449-467.
Start Year: 1955
Start Age: 25
Interval (Years): 5

A	B	C	D	E	F	G	H	I	J
3	1578947	2	1538462	7	1400000	3	1578947	10	1428571
11	1666667	16	1632653	11	1527778	10	1408451	7	1228070
11	1410256	22	1666667	24	1632653	25	1524390	15	1136364
36	1348315	44	1392405	42	1660079	53	1568047	48	1221374
77	1590909	74	1321429	68	1379310	99	1636364	88	1288433
106	1606061	131	1541176	99	1294118	142	1340888	134	1285988
157	1515444	184	1533333	189	1490536	180	1255230	177	986072.4
193	1307588	232	1417227	262	1455556	249	1414773	239	999581.8
219	1066732	267	1181416	323	1297189	325	1335799	343	1048930
223	849847.6	250	902527.1	308	1010830	412	1115322	358	930595.3
198	591574.5	214	636715.3	253	688060.9	338	773632.4	312	690265.5



	1955 - 1959		1960 - 1964		1965 - 1969		1970 - 1974		1975 -	
Age	Count	Population	Count	Population	Count	Population	Count	Population	Count	Po
25	3	1,578,947.368	2	1,538,461.538	7	1,400,000	3	1,578,947.368	10	1,42
30	11	1,666,666.667	16	1,632,653.061	11	1,527,777.778	10	1,408,450.704	7	1,22
35	11	1,410,256.41	22	1,666,666.667	24	1,632,653.061	25	1,524,390.244	15	1,13
40	36	1,348,314.607	44	1,392,405.063	42	1,660,079.051	53	1,568,047.337	48	1,22
45	77	1,590,909.091	74	1,321,428.571	68	1,379,310.345	99	1,636,363.636	88	1,28
50	106	1,606,060.606	131	1,541,176.471	99	1,294,117.647	142	1,340,887.63	134	1,28
55	157	1,515,444.015	184	1,533,333.333	189	1,490,536.278	180	1,255,230.126	177	986
60	193	1,307,588.076	232	1,417,226.634	262	1,455,555.556	249	1,414,772.727	239	99
65	219	1,066,731.612	267	1,181,415.929	323	1,297,188.755	325	1,335,799.425	343	1,04
70	223	849,847.561	250	902,527.0758	308	1,010,830.325	412	1,115,322.144	358	93
75	198	591,574.5444	214	636,715.2633	253	688,060.9192	338	773,632.4102	312	690

Click on this button . . .


Age Period Cohort Analysis Tool - Mozilla Firefox (Private Browsing)
File Edit View History Bookmarks Tools Help
Age Period Cohort Analysis Tool
analysistools.nci.nih.gov/apc/#
Google
Most Visited Getting Started Web of Knowledge [v.... Home - PubMed - NCBI D3.js - Data-Driven D...
National Cancer Institute
at the National Institutes of Health | www.cancer.gov
Division of Cancer Epidemiology and Genetics
Biostatistics Branch
Age Period Cohort Analysis Web Tool
Title
Help
Belgium Female Lung Cancer Mo
Description
Example from: Clayton D. & Schifflers E. Models for temporal variation in cancer rates. I: Age-period and
Start Year
Start Age
Interval (Years)
1955
25
5
Copy and paste into table on right or upload a csv with population and count information
Browse...
ClaytonSchifflers1987StatMed.csv
Clear
Calculate
R-Studio Input
Download
Input
Age Deviations
Per Deviations
Coh Deviations
Long Age
Cross Age
Long 2 CrossRR
Fitted Temporal Trend
PeriodRR
CohortRR
Local Drifts
Period RR
Rate Ratio
1.4
1.3
1.2
1.1
1.0
0.9
0.8
1960
1965
1970
1975
Period

Scroll down . . .

Age Period Cohort Analysis Tool - Mozilla Firefox (Private Browsing)

File Edit View History Bookmarks Tools Help

Age Period Cohort Analysis Tool

analysistools.nci.nih.gov/apc/#

Most Visited | Getting Started | Web of Knowledge [v.... | Home - PubMed - NCBI | D3.js - Data-Driven D...

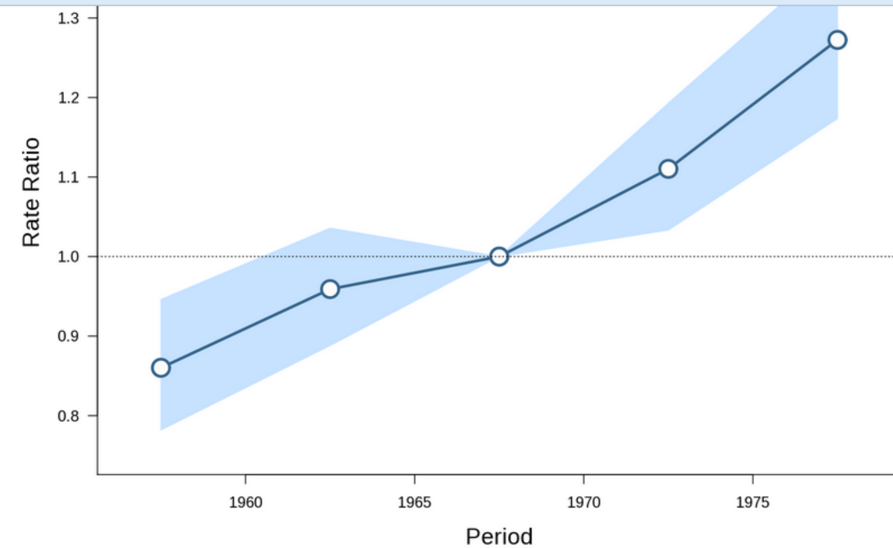
Start Year	Start Age	Interval (Years)
1955	25	5

Copy and paste into table on right or upload a csv with population and count information

Browse... ClaytonSchifflers1987StatMed.csv

Clear | Calculate

R-Studio Input | Download

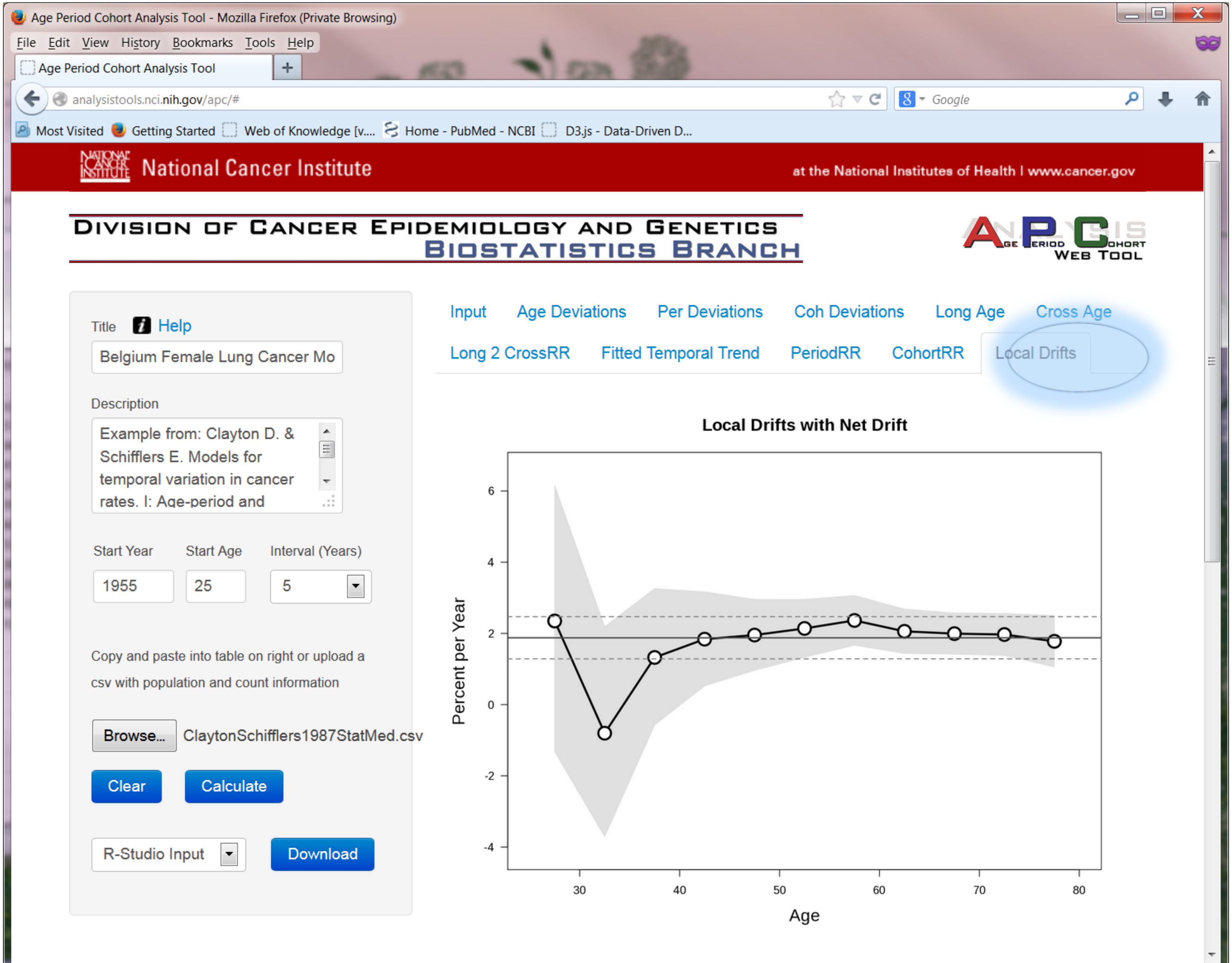


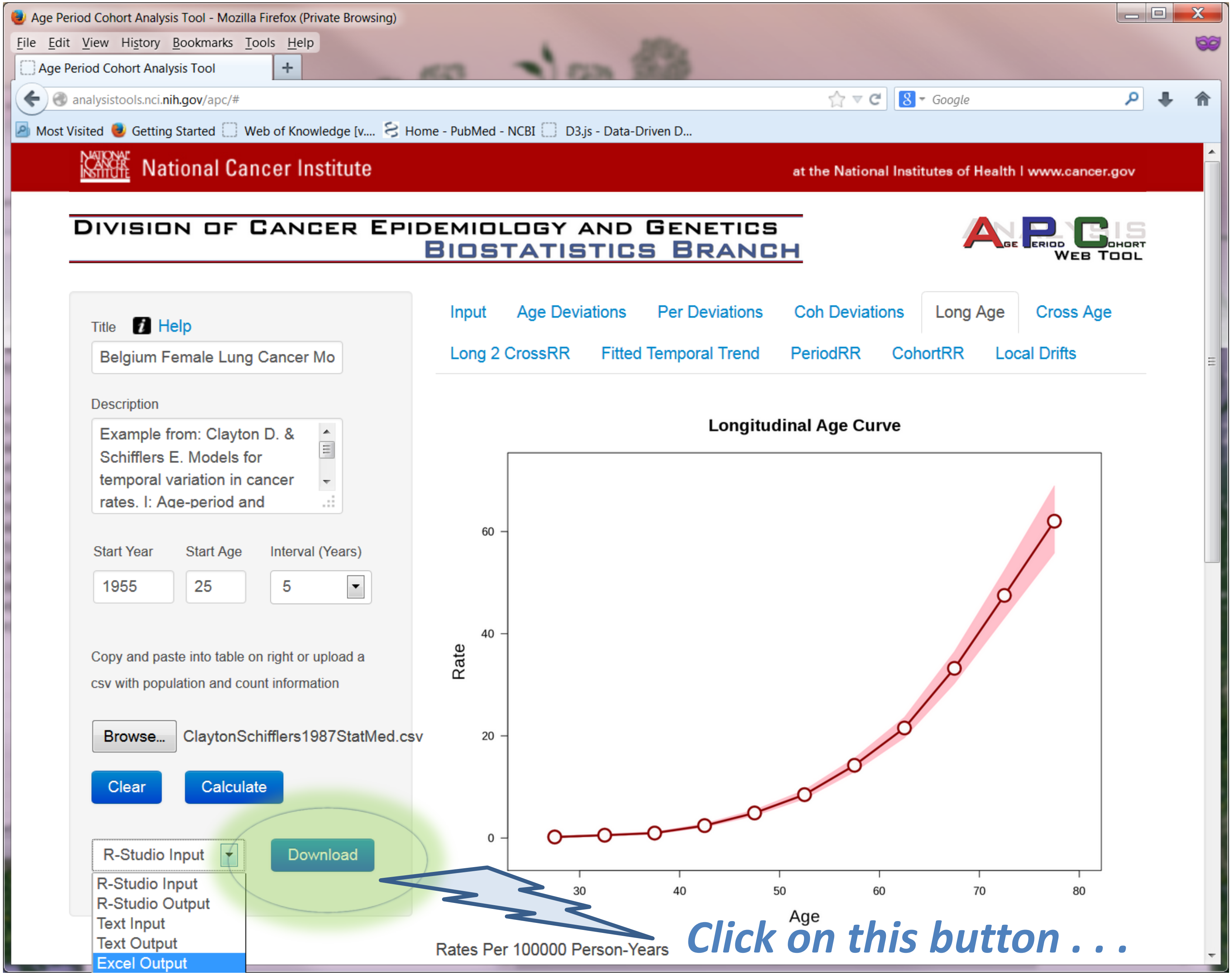
Net Drift		
Net Drift (%/year)	CI Lo	CI Hi
1.876	1.286	2.47

Period	Rate Ratio	CILo	CIHi
1957.5	0.86	0.782	0.946
1962.5	0.959	0.888	1.036
1967.5	1	1	1
1972.5	1.11	1.033	1.193
1977.5	1.272	1.173	1.38

Wald Tests			
	X2	df	P-Value
NetDrift = 0	39.275	1	0
All Age Deviations = 0	117.61	9	0
All Period Deviations = 0	3.58	3	0.3105
All Cohort Deviations = 0	11.876	13	0.5379
All Period RR = 1	46.575	4	0
All Cohort RR = 1	164.44	14	0
All Local Drifts = Net Drift	7.825	11	0.7288

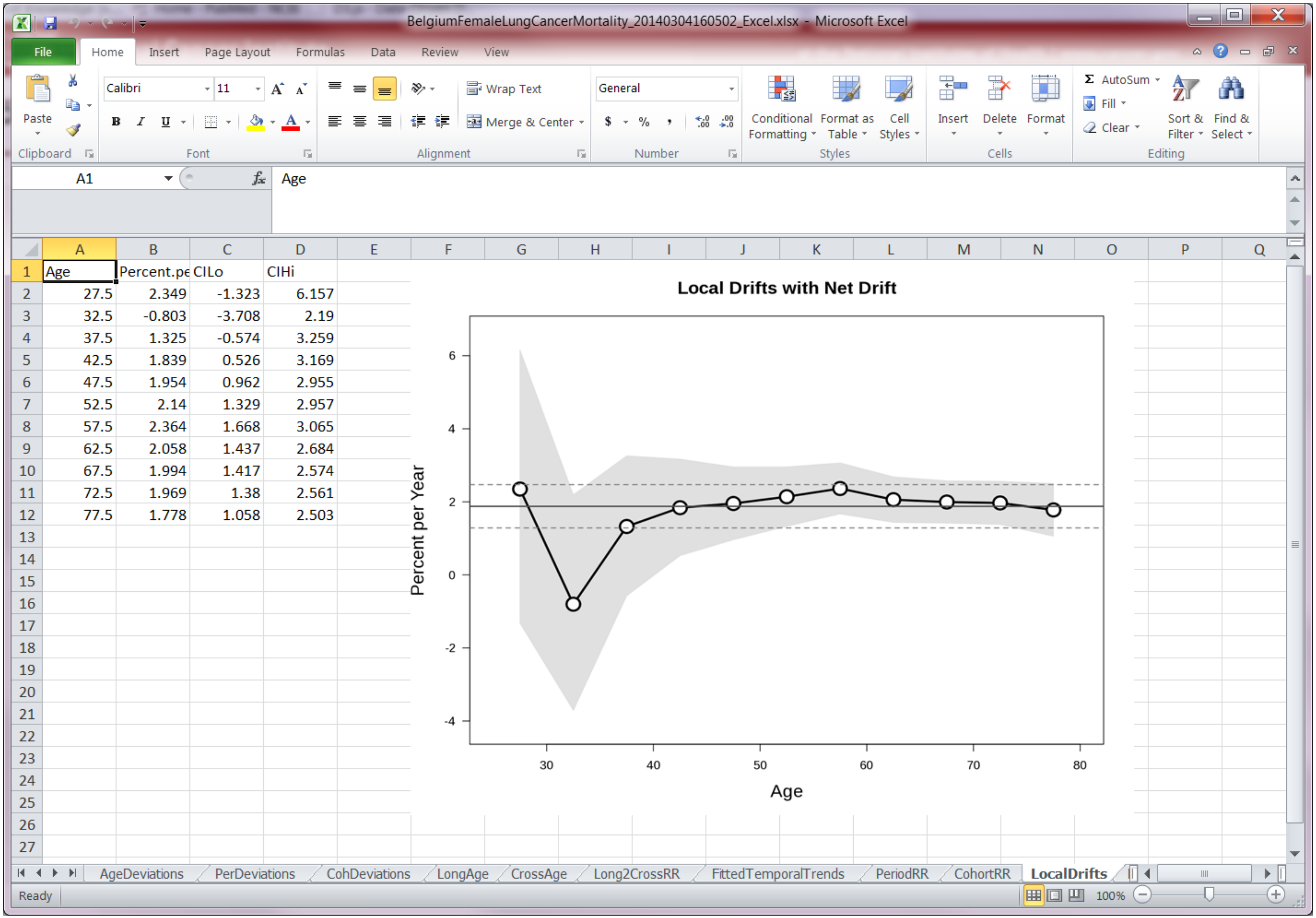
Hypothesis Tests . . .


Age Period Cohort Analysis Tool - Mozilla Firefox (Private Browsing)
File Edit View History Bookmarks Tools Help
Age Period Cohort Analysis Tool
analysistools.nci.nih.gov/apc/#
Google
Most Visited
Getting Started
Web of Knowledge [v....
Home - PubMed - NCBI
D3.js - Data-Driven D...
National Cancer Institute
at the National Institutes of Health | www.cancer.gov
Division of Cancer Epidemiology and Genetics
Biostatistics Branch
Age Period Cohort Analysis Web Tool
Title
Help
Belgium Female Lung Cancer Mo
Description
Example from: Clayton D. & Schifflers E. Models for temporal variation in cancer rates. I: Age-period and
Start Year
Start Age
Interval (Years)
1955
25
5
Copy and paste into table on right or upload a csv with population and count information
Browse...
ClaytonSchifflers1987StatMed.csv
Clear
Calculate
R-Studio Input
Download
Input
Age Deviations
Per Deviations
Coh Deviations
Long Age
Cross Age
Long 2 CrossRR
Fitted Temporal Trend
PeriodRR
CohortRR
Local Drifts
Local Drifts with Net Drift
Percent per Year
Age
6
4
2
0
-2
-4
30
40
50
60
70
80



Click on this button . . .

Model Outputs in Excel



Age	Percent.pe	CILo	CIHi
27.5	2.349	-1.323	6.157
32.5	-0.803	-3.708	2.19
37.5	1.325	-0.574	3.259
42.5	1.839	0.526	3.169
47.5	1.954	0.962	2.955
52.5	2.14	1.329	2.957
57.5	2.364	1.668	3.065
62.5	2.058	1.437	2.684
67.5	1.994	1.417	2.574
72.5	1.969	1.38	2.561
77.5	1.778	1.058	2.503

Conclusions

- *What does it do?*

  The tool fits the APC Model and serves up Model Outputs.

- *Why is that important?*

  Many cancers present complicated patterns. The outputs complement and extend standard descriptive methods.

- *Who built it, and how?*

  BB – concept, design, computations
  CBIIT – "Webification"

- *How do I use it?*

  http://analysistools.nci.nih.gov/apc/