EDEN – European Dermato-Epidemiology Network

Statistics in skin cancer epidemiology

Loes Hollestein, PhD
Erasmus MC Cancer Institute
Netherlands Comprehensive Cancer Center

Outline

- Standardised incidence rates
- Standardised Incidence Ratio (SIR)
- Trend analysis using joinpoint regression

Skin Cancer incidence

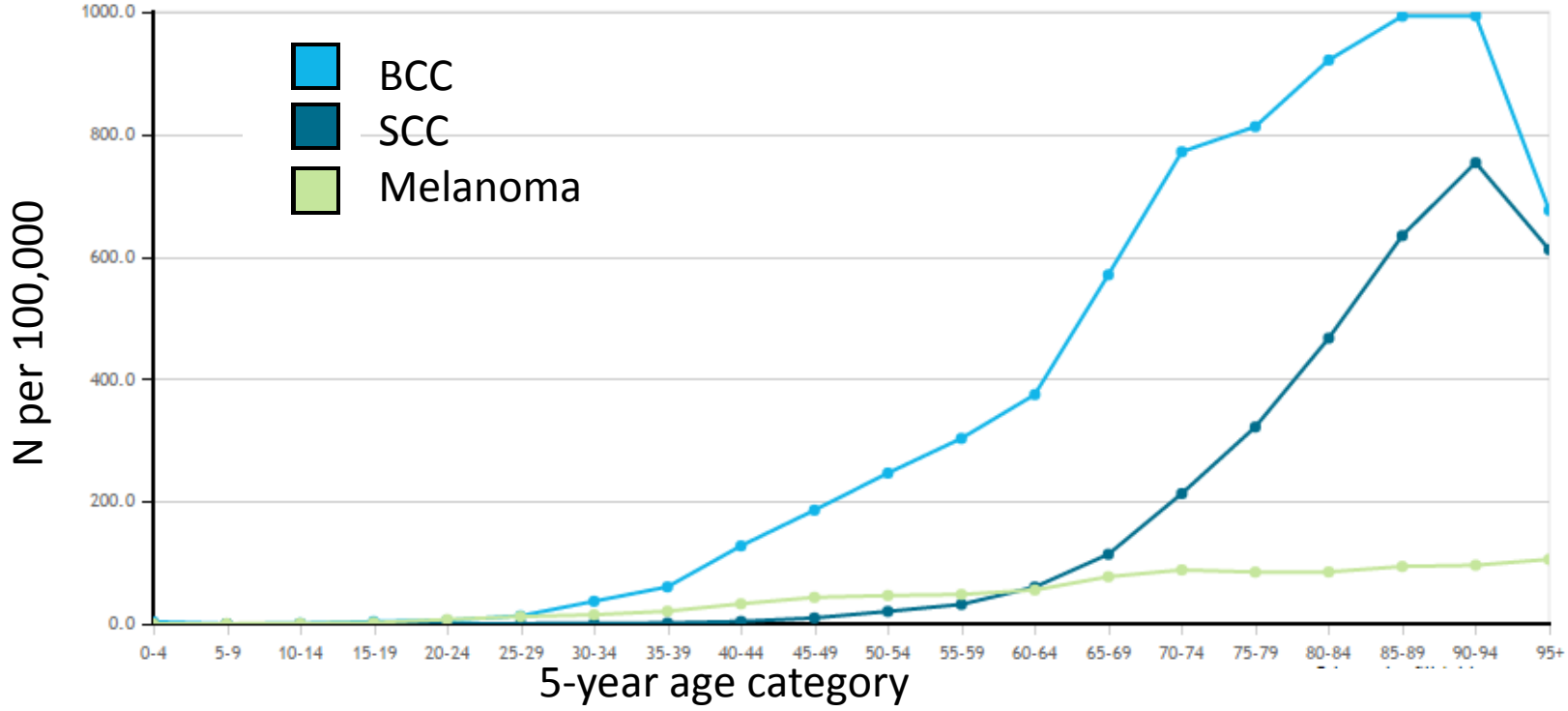
- "The number of new cancer patients diagnosed in a certain time period in a certain population"

<u>Numbers of newly diagnosed cases</u>

Total number of people in population

- Usually expressed per 100,000 person-years
- Possibility to stratify by age, sex, etc.

Age specific patterns of skin cancer incidence



Age and sex specific incidence rates

	Males		Females		Total	
	N	rate	N	rate	N	rate
all ages	11	0,3	2102	43,2	2114	23,0
0-50	3	0,1	735	18,2	738	9,6
50-85+	8	1,2	1367	163,7	1376	92,3

- Overall rate: 23, but driven by females
- Agegroup <50: much lower rates
- Agegroup >= 50: very high rates

- Conclusion: If you would only present overall rate you would get very different interpretations!

- It is important to always inspect sex- and age-specific patterns

Why standardising?

- Cancer is generally more common among the elderly

- When comparing different populations (by region, by time period, by socio-economic status, risk factor, etc), just the differences in age-distribution of the population can ‘cause’ differences in incidence rates!

A numerical example of standardization

Cancer incidence in an imaginary population A: 68 per 100,000 person-years

Age	Population	New cases	Incidence rate (per 100,000)
0-14	160 000	32	
15-44	120 000	60	
45-64	80 000	80	
65+	40 000	100	
TOTAL	**400 000**	**272**	$\frac{272}{400{,}000} \times 100{,}000 = 68$

A numerical example of standardization

Cancer incidence in an imaginary population A: increases with age

Age	Population	New cases	Incidence rate (per 100,000)
0-14	160 000	32	$\frac{32}{160{,}000} \times 100{,}000 = 20$
15-44	120 000	60	$\frac{60}{120{,}000} \times 100{,}000 = 50$
45-64	80 000	80	$\frac{80}{80{,}000} \times 100{,}000 = 100$
65+	40 000	100	$\frac{100}{40{,}000} \times 100{,}000 = 250$
TOTAL	**400 000**	**272**	$\frac{272}{400{,}000} \times 100{,}000 = 68$

A numerical example of standardization

Pop. A

Age	Population	New cases	Incidence rate (per 100,000)
0-14	160 000	32	20
15-44	120 000	60	50
45-64	80 000	80	100
65+	40 000	100	250
TOTAL	**400 000**	**272**	**68**

Pop. B

Age	Population	New cases	Incidence rate (per 100,000)
0-14	40 000	32	80
15-44	80 000	160	200
45-64	120 000	480	400
65+	160 000	1600	1000
TOTAL	**400 000**	**2272**	**568**

Comparing population A vs B

- Population A
- Incidence rate: 272

- Population B
- Incidence rate: 2272

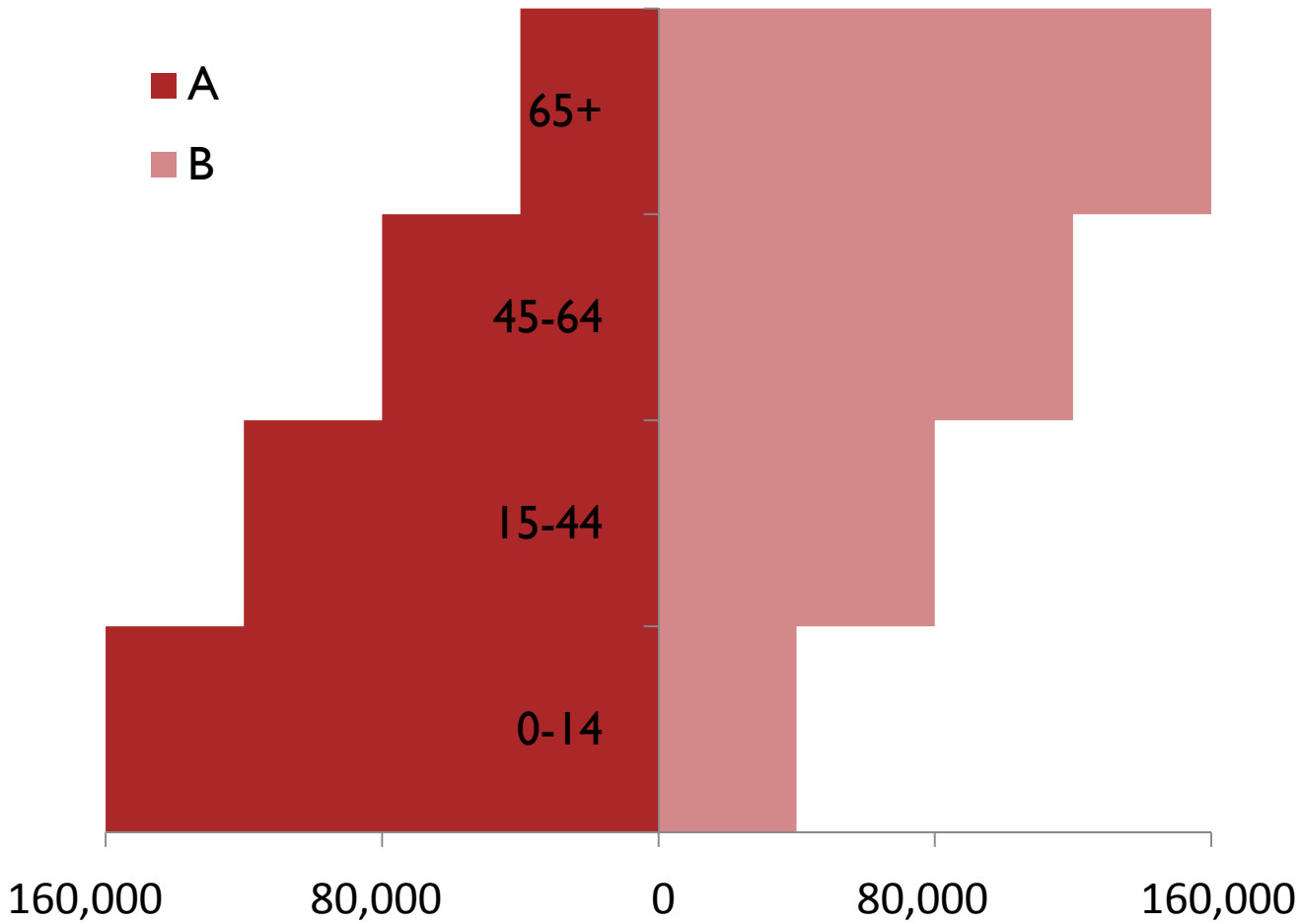
Incidence rate around 8x higher in population B

Age	Incidence A	Incidence B
0-14	20	80
15-44	50	200
45-64	100	400
65+	250	1000

Age-specific: 4x higher!

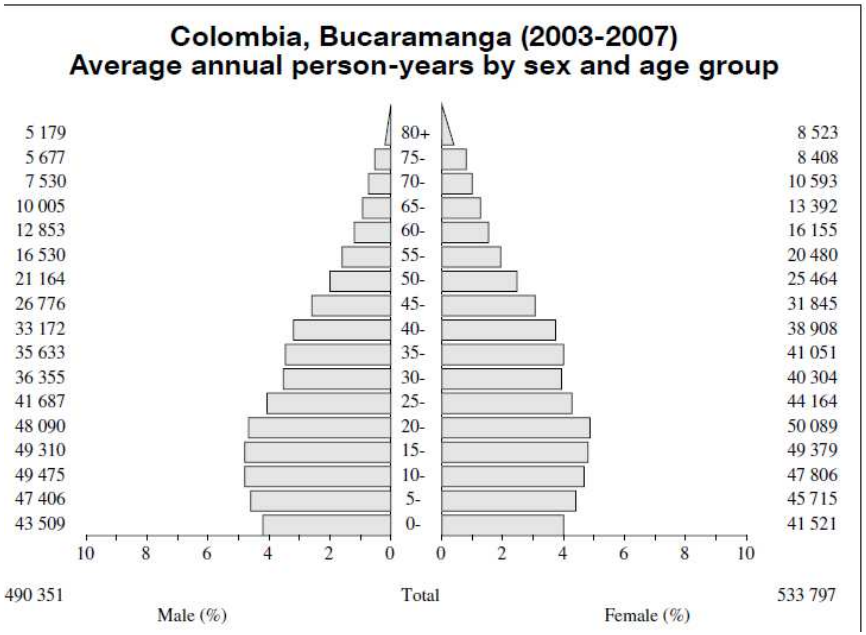
Bias because of differences in age-distribution

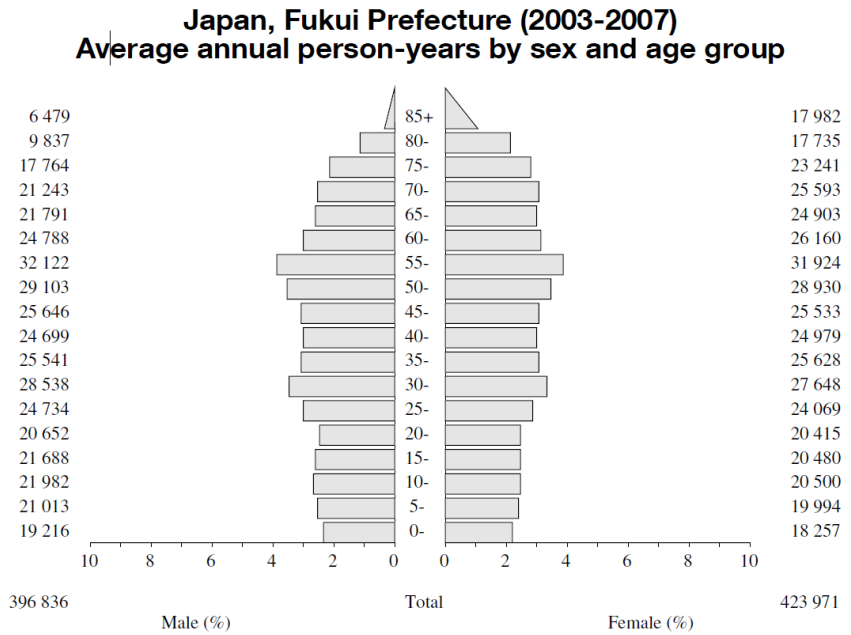
Comparing age-structure of population A vs B



Example with 2 populations

- Cancer Incidence in 5 continent vol X





日本で皮膚がんが少ない
Nihon de hifu gan ga sukunai

SCC incidence (WSR2000-2025) women

Japan:	2 per 100,000
Netherlands:	11 per 100,000

What is the most important reason for this difference?

A. Skin type
B. Population distribution
C. Fashion
D. Geographic location

Fashion & Geographic location!

Standardization

- Standardization: choosing equal weighting schemes derived from a common reference population
- Reference population could be any, e.g.
  - European Standard Population 1976 (WHO)
  - European Standard Population 2013 (WHO)
  - World Standard Population 1968 (WHO)
  - World Standard Population 2000-2025 (WHO)
  - US 2000 population (SEER, USA)

Standardization

- We make our own standard population, with 250,000 persons in each age group

Age	STANDARD Population		
0-14	250 000		
15-44	250 000		
45-64	250 000		
65+	250 000		
TOTAL	**1 000 000**		

Direct Standardization

(mostly used in cancer registry data)

- What over-all incidence rate would STANDARD have if the age-specific mortality rates in the standard population were the same as in A or B?
- Weighting scheme: derived from the age-distribution of the reference population ‘STANDARD’

Standardization

- What over-all incidence rate would STANDARD have if the age-specific mortality rates in the standard population were the same as in A?

Age	STANDARD Population	Incidence rates A (per 100,000)	Expected cases in STANDARD population
0-14	250 000		
15-44	250 000		
45-64	250 000		
65+	250 000		
TOTAL	**1 000 000**		

Standardization

- What over-all incidence rate would STANDARD have if the age-specific mortality rates in the standard population were the same as in A? 1050 (vs 68 in A)

Age	STANDARD Population	Incidence rates A (per 100,000)	Expected cases in STANDARD population
0-14	250 000	20	$\frac{20}{100{,}000} \times 250{,}000 = 50$
15-44	250 000	50	$\frac{50}{100{,}000} \times 250{,}000 = 125$
45-64	250 000	100	$\frac{100}{100{,}000} \times 250{,}000 = 250$
65+	250 000	250	$\frac{250}{100{,}000} \times 250{,}000 = 625$
TOTAL	**1 000 000**		$\mathbf{50 + 125 + 250 + 625 = 1050}$

Standardization

(mostly used in cancer registry data)

- Observed incidence in A: 68
- Expected number of cases if A would have had population structure 'STANDARD': 1050
- Standardized Incidence Rate : $\frac{1{,}050}{1{,}000{,}000} \times 100{,}000 = 105$

Standardization

- Similar for population B: expected number of cases in STANDARD:4200 (vs 568 in B)

Age	STANDARD Population	Incidence rates B (per 100,000)	Expected cases STANDARD
0-14	250 000	80	$\frac{80}{100{,}000} \times 250{,}000 = 200$
15-44	250 000	200	$\frac{200}{100{,}000} \times 250{,}000 = 500$
45-64	250 000	400	$\frac{400}{100{,}000} \times 250{,}000 = 1000$
65+	250 000	1000	$\frac{1000}{100{,}000} \times 250{,}000 = 2500$
TOTAL	**1 000 000**		$\mathbf{200 + 500 + 1000 + 2500 = 4200}$

Comparing population A vs B

- Population A
- Incidence rate: 272

- Population B
- Incidence rate: 2272

Crude Incidence rate ~ 8x higher in B
Age-specific rates: 4x higher in B

Bias because of differences in age-distribution

Age-standardised rates:
Population A: 105
Population B: 420

Difference: factor 4!

Standardized incidence RATIO (SIR)

- Ratio of standardized incidence rates of two populations
- Example: population B vs A:

- Conclusion: incidence is 4x higher in population B compared with population A

$$SIR = \frac{420(B)}{105(A)} = 4$$

Comparability

- SIR only applicable to the same standard populations

- Different standardardised rates not comparable

  e.g. WSR cannor be compared to ESR.

- Problem in literature.

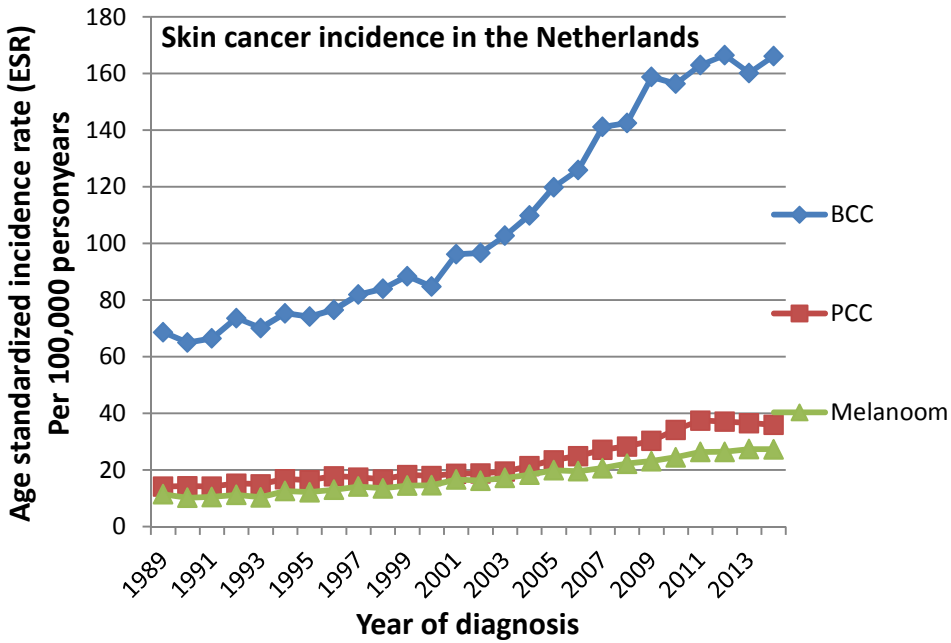
- Solution: Provide ESR/ESR2013/WSR/WSR2000-2025/USSR in supplementary material

95% Confidence Interval

Should a 95% CI be provided?

No, not a sample: nationwide data

Yes, when extrapolating



- 95% CI:
  - Normal approximation
  - Exact method (holds with small counts)

    (When count is large the 2 methods produce similar results)

- Formulas: http://seer.cancer.gov/seerstat/WebHelp/Rate_Algorithms.htm

Cijfersoverkanker.nl

Prediction

Table III. *Predicted basal cell carcinoma (BCC) incidence rates (European Standardized Rates per 100,000 person years) and num of newly diagnosed tumours in the Netherlands*

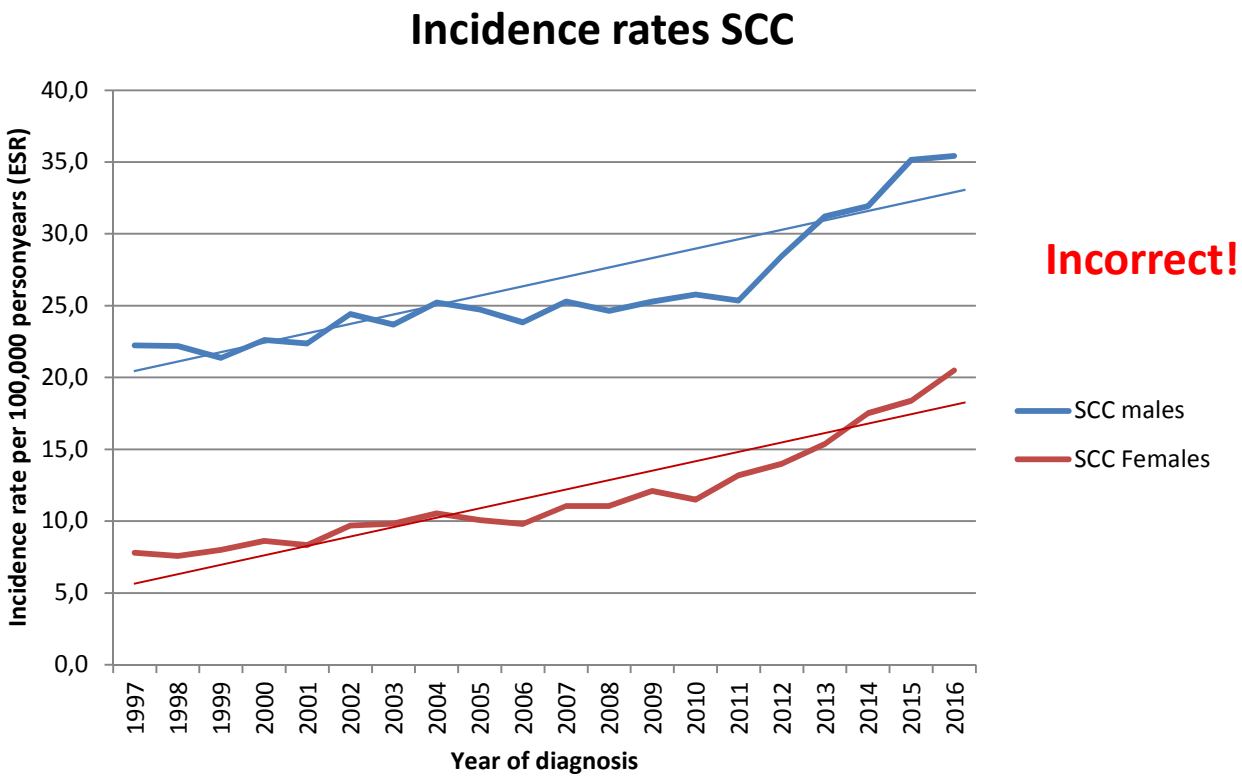
	2005 Observed	2010 Expected (95% PI)	2015 Expected (95% PI)	2020 Expected (95% PI)
BCC incidence rate				
Men				
15–34 years	7.8	9.8 (6.5–12.9)	12.2 (7.1–17.3)	14.6 (7.5–21.7)
35–64 years	136.5	167.6 (155.8–179.3)	198.2 (178.9–217.5)	228.8 (201.5–256.1)
>65 years	658.1	859.9 (813.6–906.3)	1,067.5 (992.4–1,142.6)	1,275.0 (1,169.7–2,168.8)
All	139.5	162.7 (155.8–169.5)	198.1 (186.9–209.3)	233.5 (217.7–249.2)

Flohil et al. JEADV 2011

Summary standardising incidence rates

- 5-year age categories
- World, European or US standard populations
- Usually expressed per 100,000 person-years
- Needed to compare incidence between populations with different age structures
  - Different countries
  - Different time periods, etc

Trends in incidence rates



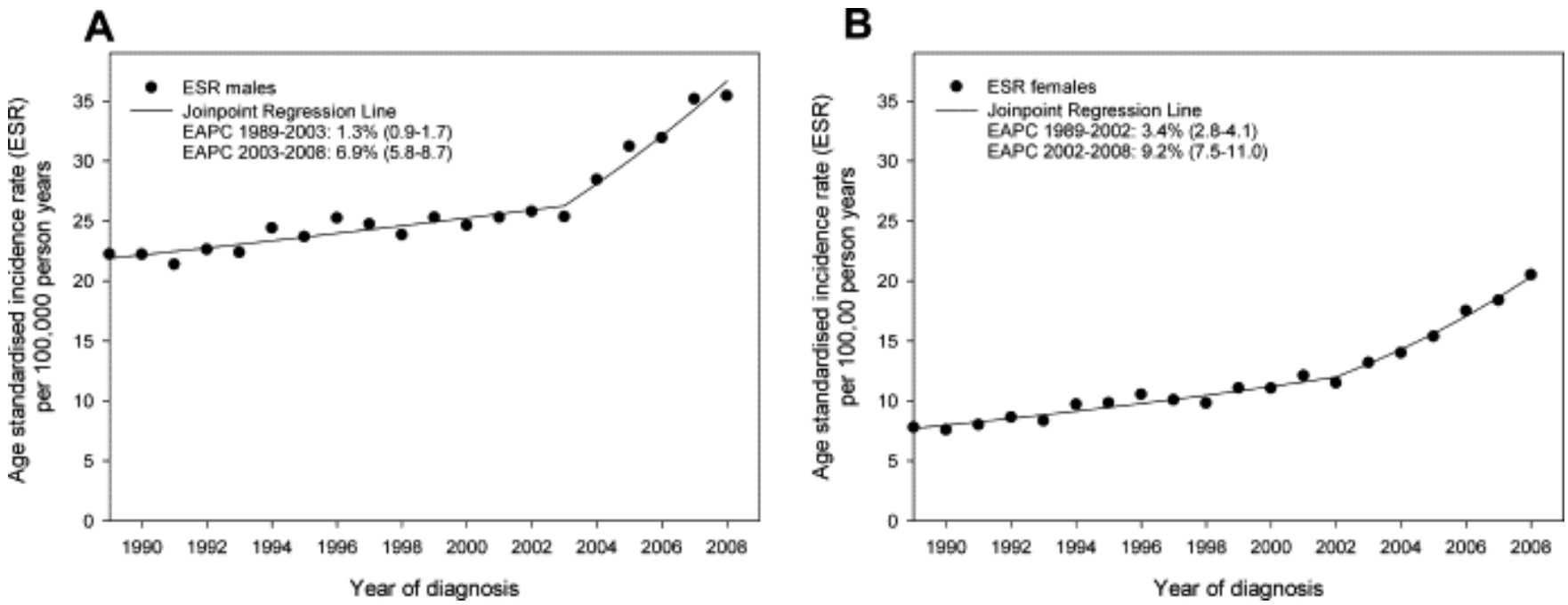
Incorrect!

Estimated Annual Percentage Change (EAPC)

Simple linear regression on LN(incidence rate)

Joinpoint regression

- Identifies the timepoint when trends changes



- Free software:
- https://surveillance.cancer.gov/joinpoint/

Hollestein et al, 2012, EJC

Practical Incidence rates

Incidence rate

- Calculate the crude and standardised incidence rate

Trends

- Calculate the EAPC

$y = ax + b$

$y = \ln(\text{rate})$

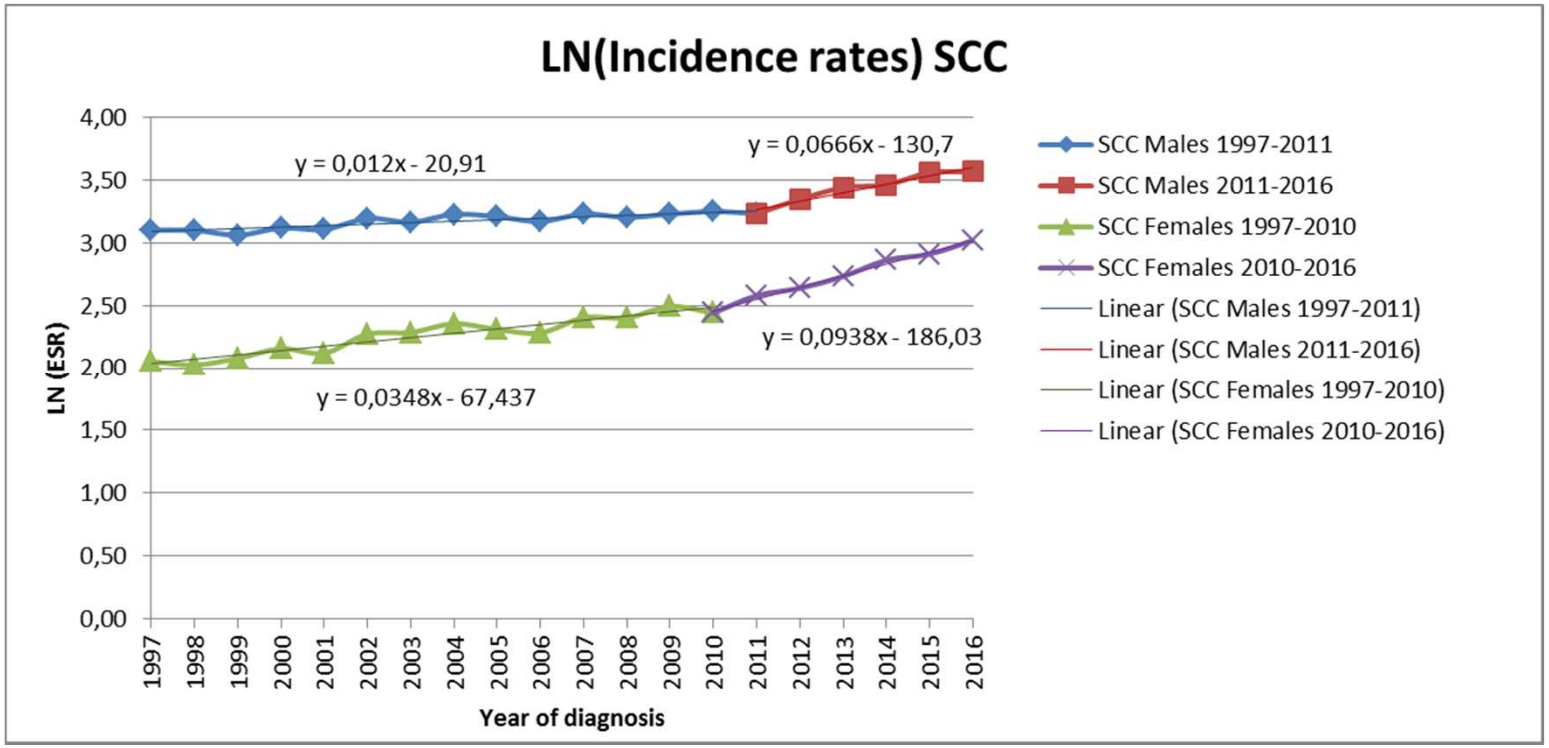
$x = \text{calendar year}$

then $\text{EAPC} = 100 \times (e^{a} - 1)$

Practical Incidence Rate

	A	B	C	D
1			Men	
2	5 years agecategory	Newly diagnosed skin cancer patients	Population size	WHO-2000-2025 Standard population
3	0 - 4 years	0	459794	8856,9
4	5 - 10 years	0	479135	8687
5	10 - 14 years	0	519726	8597
6	15 - 20 years	2	509469	8467
7	20 - 24 years	0	537679	8217,1
8	25 - 30 years	0	522520	7927,2
9	30 - 34 years	1	504790	7607,3
10	35 - 40 years	6	504694	7147,5
11	40 - 44 years	31	616191	6587,7
12	45 - 50 years	42	648065	6037,9
13	50 - 54 years	78	629198	5368,1
14	55 - 60 years	165	565898	4548,4
15	60 - 64 years	321	522346	3718,7
16	64 - 70 years	433	488688	2959
17	70 - 74 years	555	331824	2209,2
18	74 - 80 years	675	238308	1519,5
19	80 - 84 years	651	153762	909,7
20	84 - 90 years	351	74521	439,8
21	90 - 94 years	117	24176	150
22	95 years or older	25	3601	45
23	Total	3453	8334385	100000

Practical Trend



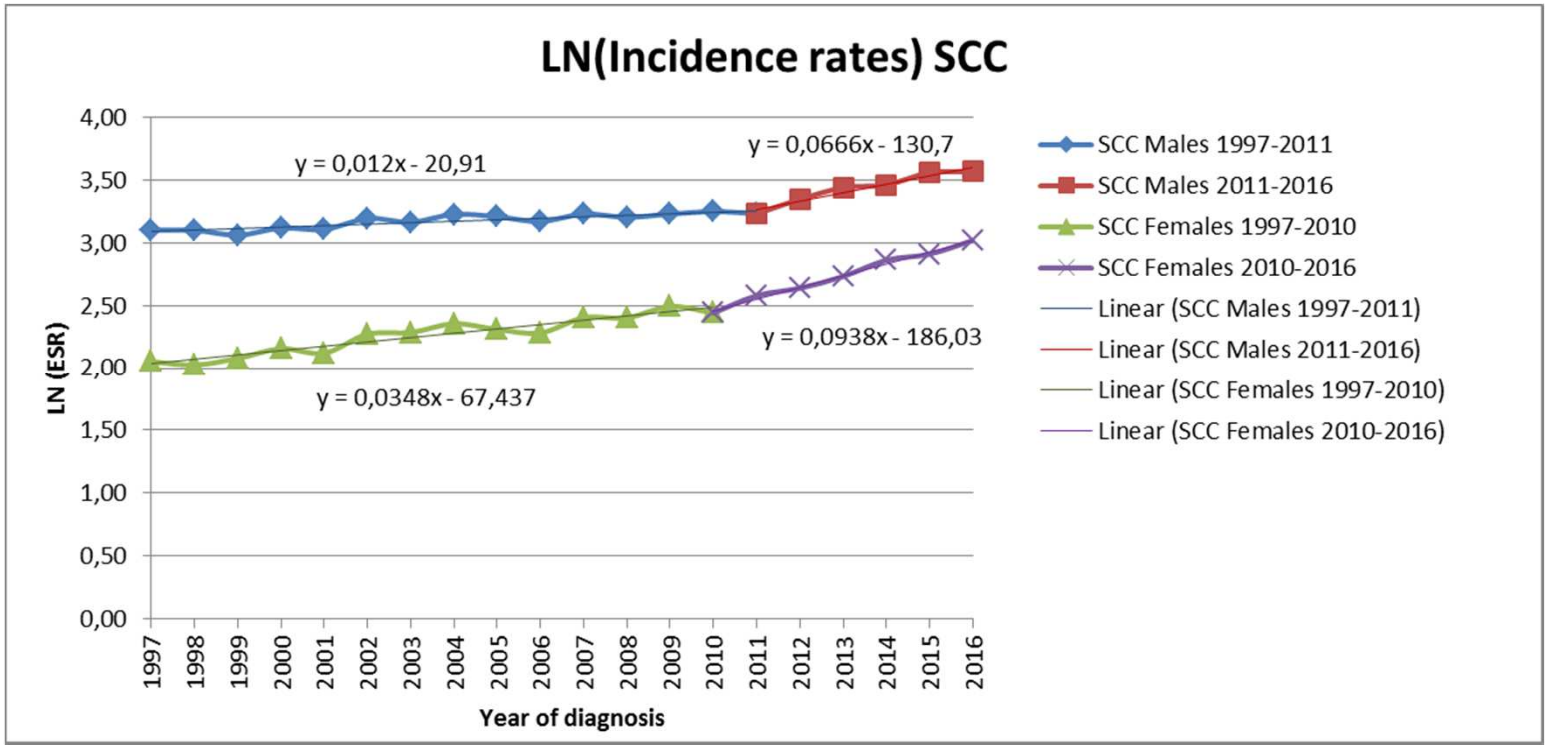
Practical Incidence Rate

	A	B	C	D	E	F
1			Men			
2	5 years agecategory	Newly diagnosed skin cancer patients	Population size	WHO-2000-2025 Standard population	Age specific incidence rate (column B/C)*100,000	Expected cases (column E/100,000)*D
3	0 - 4 years	0	459794	8856,9	0,0	0,0
4	5 - 10 years	0	479135	8687	0,0	0,0
5	10 - 14 years	0	519726	8597	0,0	0,0
6	15 - 20 years	2	509469	8467	0,4	0,0
7	20 - 24 years	0	537679	8217,1	0,0	0,0
8	25 - 30 years	0	522520	7927,2	0,0	0,0
9	30 - 34 years	1	504790	7607,3	0,2	0,0
10	35 - 40 years	6	504694	7147,5	1,2	0,1
11	40 - 44 years	31	616191	6587,7	5,0	0,3
12	45 - 50 years	42	648065	6037,9	6,5	0,4
13	50 - 54 years	78	629198	5368,1	12,4	0,7
14	55 - 60 years	165	565898	4548,4	29,2	1,3
15	60 - 64 years	321	522346	3718,7	61,5	2,3
16	64 - 70 years	433	488688	2959	88,6	2,6
17	70 - 74 years	555	331824	2209,2	167,3	3,7
18	74 - 80 years	675	238308	1519,5	283,2	4,3
19	80 - 84 years	651	153762	909,7	423,4	3,9
20	84 - 90 years	351	74521	439,8	471,0	2,1
21	90 - 94 years	117	24176	150	484,0	0,7
22	95 years or older	25	3601	45	694,3	0,3
23	Total	3453	8334385	100000		SUM(columnF)
24		Crude rate=	41,4		WSR2000-2025=	22,7

Practical Incidence Rate

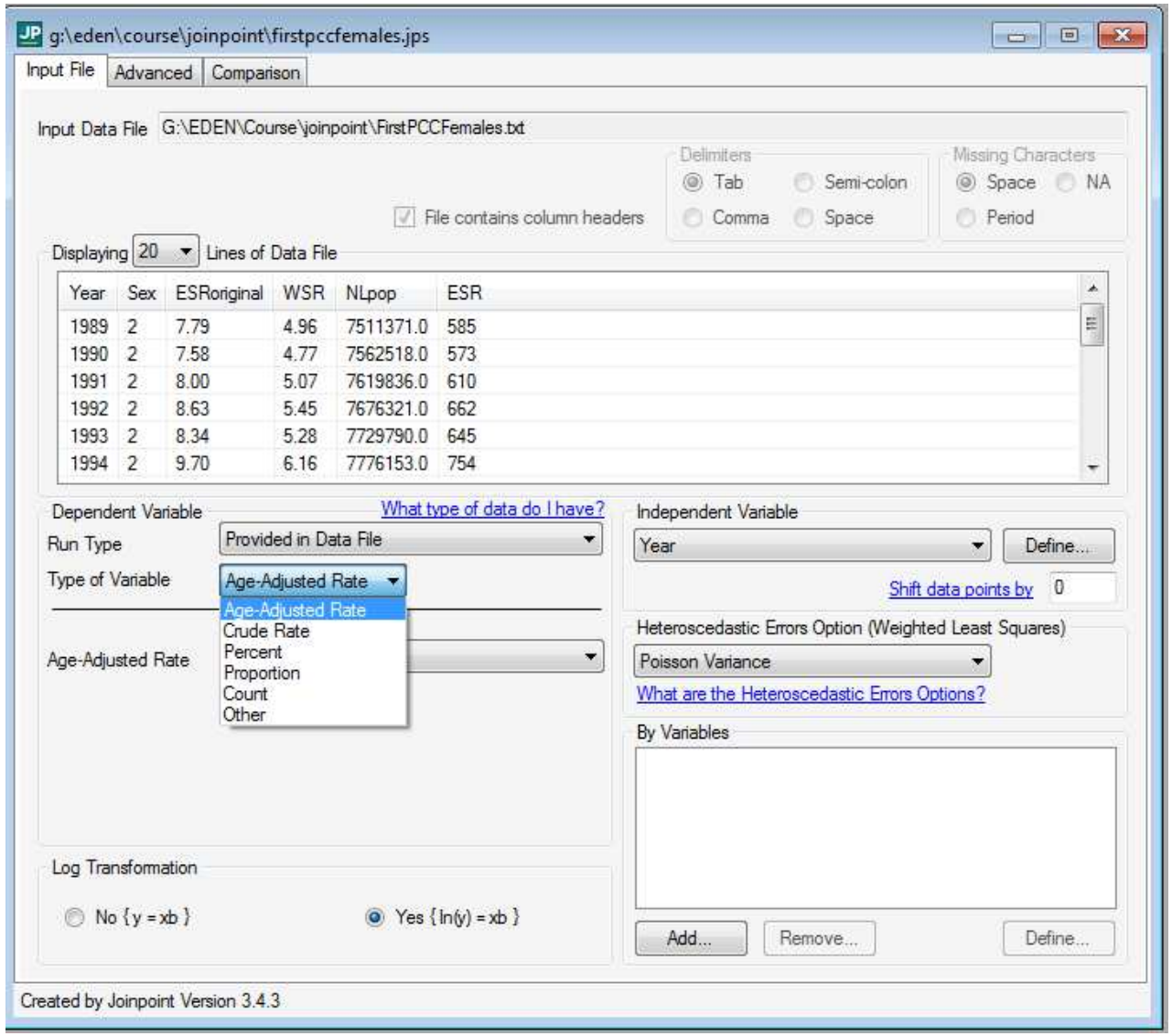
	Men				
5 years agecategory	**Newly diagnosed skin cancer patients**	**Population size**	**WHO-2000-2025 Standard population**	**column B/C**	**column E*D**
0 - 4 years	0	459794	8856,9	0,0000	0,0000
5 - 10 years	0	479135	8687	0,0000	0,0000
10 - 14 years	0	519726	8597	0,0000	0,0000
15 - 20 years	2	509469	8467	0,0000	0,0332
20 - 24 years	0	537679	8217,1	0,0000	0,0000
25 - 30 years	0	522520	7927,2	0,0000	0,0000
30 - 34 years	1	504790	7607,3	0,0000	0,0151
35 - 40 years	6	504694	7147,5	0,0000	0,0850
40 - 44 years	31	616191	6587,7	0,0001	0,3314
45 - 50 years	42	648065	6037,9	0,0001	0,3913
50 - 54 years	78	629198	5368,1	0,0001	0,6655
55 - 60 years	165	565898	4548,4	0,0003	1,3262
60 - 64 years	321	522346	3718,7	0,0006	2,2853
64 - 70 years	433	488688	2959	0,0009	2,6218
70 - 74 years	555	331824	2209,2	0,0017	3,6950
74 - 80 years	675	238308	1519,5	0,0028	4,3039
80 - 84 years	651	153762	909,7	0,0042	3,8515
84 - 90 years	351	74521	439,8	0,0047	2,0715
90 - 94 years	117	24176	150	0,0048	0,7259
95 years or older	25	3601	45	0,0069	0,3124
Total	3453	8334385	100000		**SUM(columnF)**
	Crude rate=	41,4		WSR2000-2025=	22,7

Practical Trend

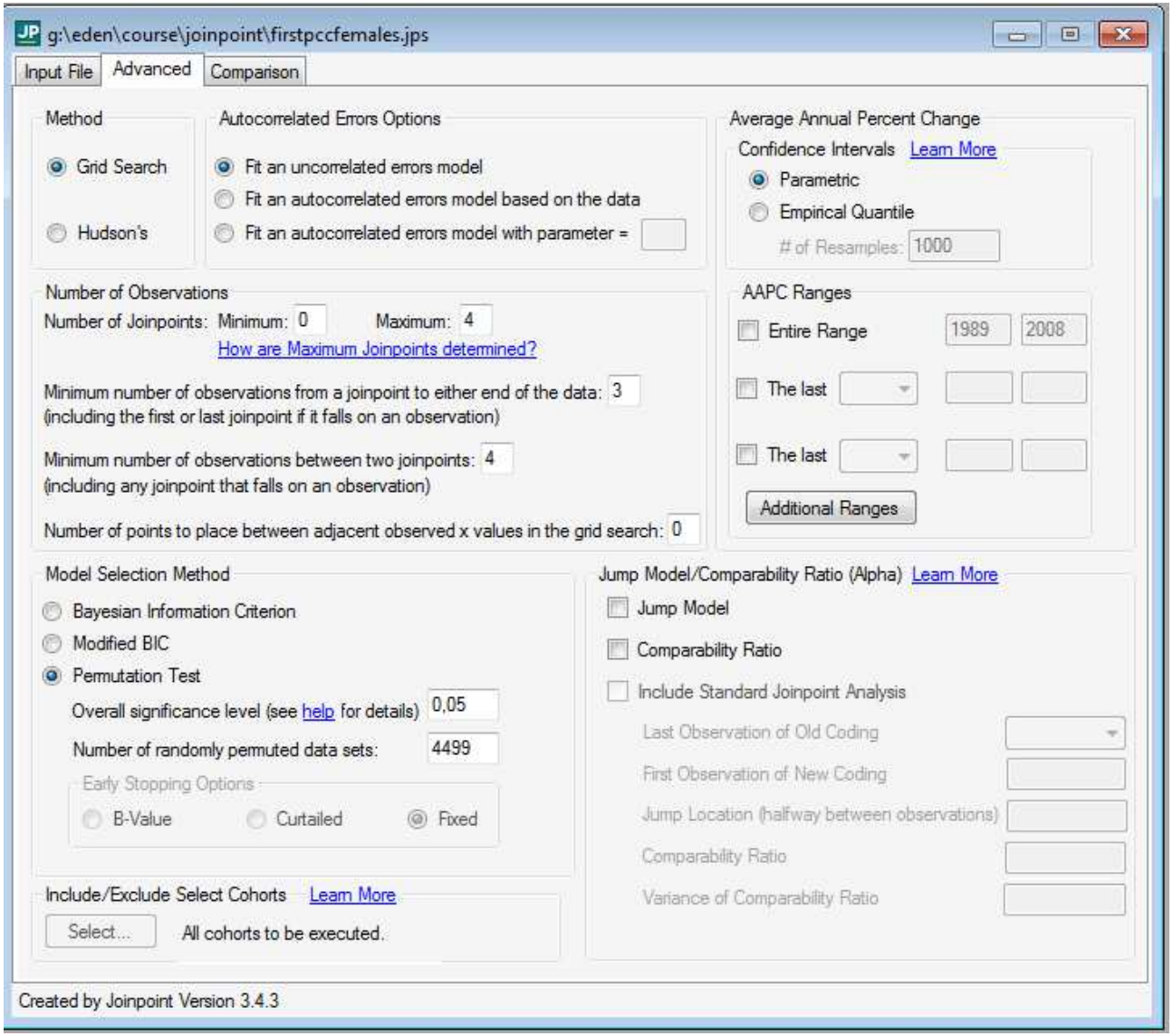


		Regression coefficient (a)	$100 \times (e^a - 1)$
Males	EAPC 1997-2011	0,012	1,2%
	EAPC 2011-2016	0,066	6,8%
Females	EAPC 1997-2010	0,0348	3,5%
	EAPC 2010-2016	0,0938	9,8%

Joinpoint regression



Joinpoint regression



Joinpoint Regression

Bestand Bewerken Beeld Geschiedenis Bladwijzers Extra Help

Number of Joinpoints — J... | Number of Joinpoints — J...

https://surveillance.cancer.gov/help/joinpoint/setting-parameters/advanced-tab/number-of-joinpoints

Create PDF

NIH NATIONAL CANCER INSTITUTE
Division of Cancer Control & Population Sciences

Search Site Search
only in current section

Surveillance Research Program

- Home
- Setting Parameters
  - Input File Tab
  - Advanced Tab
    - Method
    - Number of Joinpoints
    - Autocorrelated Errors Option
    - Average Annual Percent Change (AAPC)
    - AAPC Confidence Intervals
    - Specify AAPC Ranges
    - Number of Observations
    - Model Selection Method
    - Jump Model / Comparability Ratio Model
    - Include/Exclude Select Cohorts
  - Comparison Tab
  - Preferences
- Statistical Notes
- Executing a Session

Home › Setting Parameters › Advanced Tab › Number of Joinpoints

Number of Joinpoints

Enter the minimum and maximum number of joinpoints to fit where:

0 = min = max = 9 when the Grid Search method is selected.

-or-

0 = min = max = 4 when Hudson's method is selected.

As of Version 3.5, the default value for the maximum number of joinpoints depends on the number of data points. This value can be changed by the user, subject to having a minimum number of data points necessary to satisfy two pre-set rules:

- A joinpoint cannot occur within a user-specified number (default: 3) of data points from the beginning or end of a series.
- There must be at least a user-specified number (default: 4) of data points between two joinpoints.

The default maximum number of joinpoints is a recommendation based on the same metrics used to determine the minimum number of data points (or conversely the maximum number of joinpoints for a given number of data points). The default is based on the following recommendations:

- At least seven data points should be observed in order to consider allowing a joinpoint.
- There should be, on average, at least four data points between consecutive joinpoints.

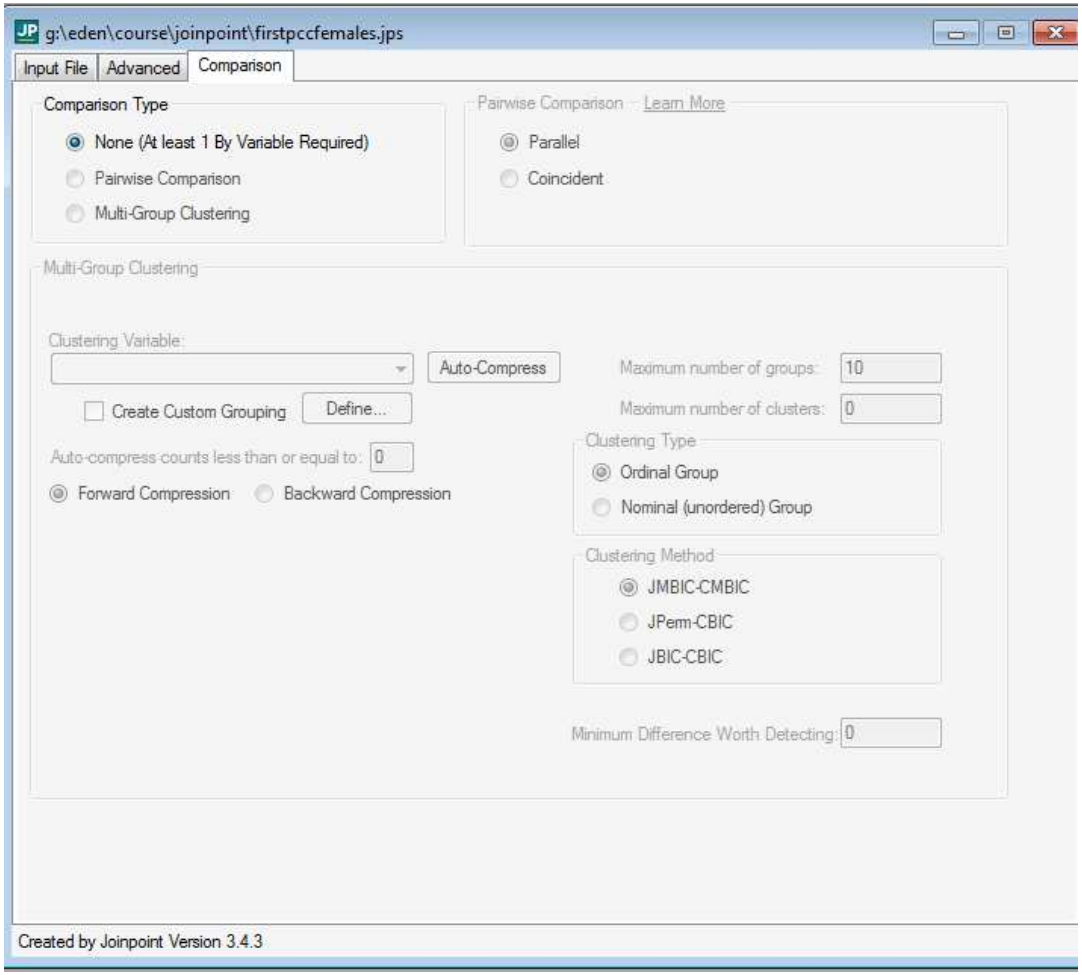
These algorithmic recommendations lead to the following default maximum number of joinpoints.

Number of Data Points	Default Maximum Number of Joinpoints
0 - 6	0
7 -11	1
12 -16	2
17 - 21	3

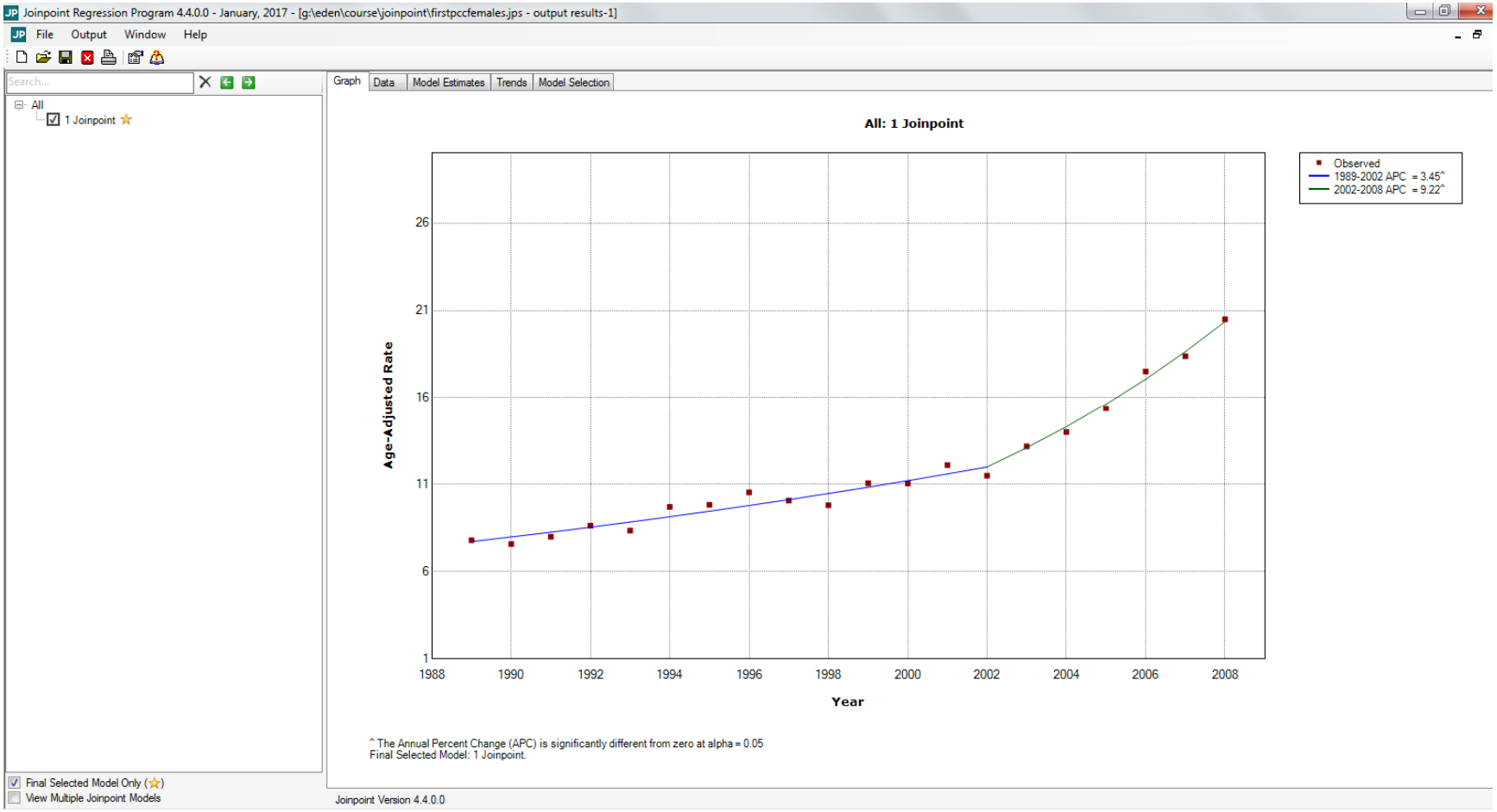
22:40
28-3-2017

https://surveillance.cancer.gov/help/joinpoint/setting-parameters/advanced-tab/number-of-joinpoints

Joinpoint regression



Joinpoint regression



Joinpoint regression

Graph | Data | Model Estimates | Trends | Model Selection

Model Statistics

Cohort	Number of Joinpoints	Number of Observations	Number of Parameters	Degrees of Freedom	Sum of Squared Errors	Mean Squared Error	Autocor Parar
	1	20	4	16	0.28577	0.01786	Uncom

Estimated Joinpoints

Cohort	Joinpoint	Estimate	Lower CI	Upper CI
	1	2002	2001	2004

Estimated Regression Coefficients (Beta)

Standard Parameterization

Parameter	Parameter Estimate	Standard Error	Test Statistic (t)	Prob > |t|
Intercept 1	-65.483405	6.207528	-10.549030	0.000000
Slope 1	0.033950	0.003111	10.913790	0.000000
Slope 2 - Slope 1	0.054242	0.008242	6.581193	0.000009

General Parameterization

Parameter	Parameter Estimate	Standard Error	Test Statistic (t)	Prob > |t|
Intercept 1	-65.483405	6.207528	-10.549030	0.000000
Intercept 2	-174.076768	15.308904	-11.370949	0.000000
Slope 1	0.033950	0.003111	10.913790	0.000000
Slope 2	0.088192	0.007632	11.554920	0.000000

Joinpoint regression

Graph | Data | Model Estimates | Trends | Model Selection

Estimated Joinpoints

Cohort	Joinpoint	Estimate	Lower CI	Upper CI				
	1	2002	2001	2004				

Annual Percent Change (APC)

Cohort	Segment	Lower Endpoint	Upper Endpoint	APC	Lower CI	Upper CI	Test Statistic (t)	Prob > |t|
	1	1989	2002	3.5^	2.8	4.1	10.9	0.0
	2	2002	2008	9.2^	7.5	11.0	11.6	0.0

^ The Annual Percent Change (APC) is significantly different from zero at alpha = 0.05

Joinpoint regression

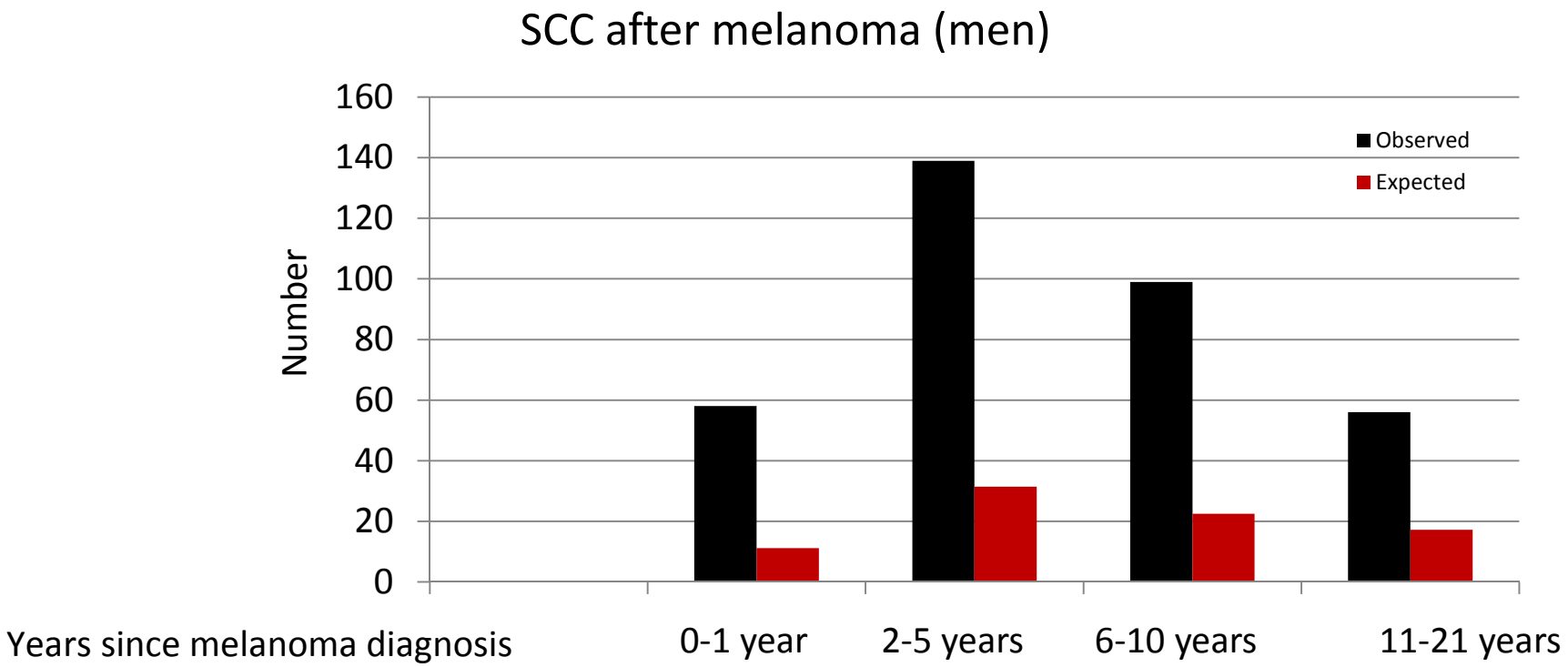
Graph | Data | Model Estimates | Trends | Model Selection

Test For Number of Joinpoints

Test Number	Null Hypothesis	Alternate Hypothesis	Numerator Degrees of	Denominator Degrees of	Number of Permutations	P-Value	Significance Level~
#1	0 Joinpoint(s)	4 Joinpoint(s) *	8	10	4500	0.0004444	0.0125000
#2	1 Joinpoint(s) *	4 Joinpoint(s)	6	10	4500	0.4073333	0.0166667
#3	1 Joinpoint(s) *	3 Joinpoint(s)	4	12	4500	0.5371111	0.0166667
#4	1 Joinpoint(s) *	2 Joinpoint(s)	2	14	4500	0.4055556	0.0166667

Final Selected Model: - 1 Joinpoint(s)

Risk of subsequent different skin cancer



Cumulative incidence	0.3%	1%	2%	5%
Standardised incidence ratio	5	4	4	3
Absolute exess risk (per 10,000 personyears)	24	20	22	18

Van der Leest, 2016, submitted

Standardized Mortality Ratio

$$SMR = \frac{Observed\ mortality}{Expected\ mortality}$$

Table 1. All-cause mortality after BCC diagnosis

		0-6 Months	0-12 Months			0-24 Months		
Age at diagnosis	N total BCC	Observed deaths (% of total)	Observed deaths (% of total)	Expected deaths (% of total)	SMR (95% CI)*	Observed deaths (% of total)	Expected deaths (% of total)	SMR (95% CI)*
< 80	9465	62 (0.7%)	140 (1.5%)	132 (1.4%)	1.1 (0.9-1.4)	291 (3.1%)	268 (2.8%)	1.1 (1.0-1.2)
80-84	974	25 (2.6%)	55 (5.6%)	68 (7.0%)	0.8 (0.6-1.1)	126 (12.9%)	135 (13.9%)	0.9 (0.8-1.1)
85-89	557	24 (4.3%)	58 (10.4 %)	66 (11.8%)	0.9 (0.7-1.1)	132 (23.7%)	126 (22.6%)	1.0 (0.9-1.2)
> 90	174	15 (8.6%)	32 (18.4%)	36 (20.7%)	0.9 (0.6-1.2)	65 (37.4%)	63 (36.2%)	1.0 (0.8-1.3)
Total	11170	126 (1.1%)	285 (2.5%)	302 (2.7%)	0.9 (0.8-1,1)	614 (5.5%)	592 (5.3%)	1.0 (1.0-1.1)

* Expected deaths based on probability of death of the Dutch general population using data from Statistics Netherlands.
CI= confidence interval, SMR= Standardized Mortality Rate

Waalboer R et al, 2017, BJD

Literature

- WHO Cancer Epidemiology: Principles and methods

Free PDF download:

http://www.iarc.fr/en/publications/pdfs-online/epi/cancerepi/