

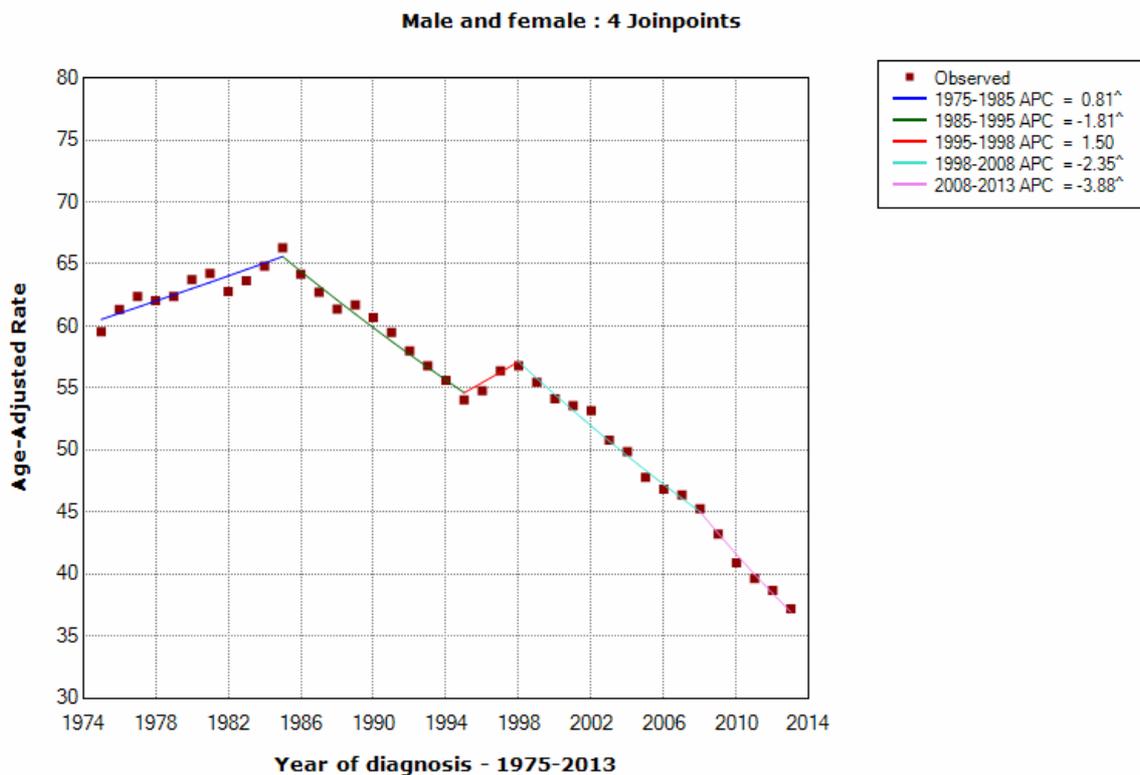
Joinpoint Help Manual

4.8.0.1

Joinpoint

The Joinpoint Regression Program is 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 the selected trend data (e.g., cancer rates) into the simplest joinpoint model that the data allow. The resulting graph is like the figure below, where several different lines are connected together at the "joinpoints."



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

Joinpoint also allows the user to view one graph for each joinpoint model, from the model with the minimum number of joinpoints to the model with maximum number of joinpoints.

You may use any software to create data files to be analyzed by Joinpoint (see Input Data File). The SEER*Stat software provides an easy mechanism for creating data files containing cancer rates or counts that can be analyzed by Joinpoint. The controls on the Joinpoint **Specifications** tab will be set automatically if you load data files exported from SEER*Stat. To learn more about the SEER*Stat software visit the SEER*Stat Web site at:

<https://seer.cancer.gov/seerstat/>

Related content

- [Setting The Joinpoint Parameters](#)
- [Citation](#)
- [Technical Support](#)
- [System Requirements](#)

Setting The Joinpoint Parameters

When using Joinpoint, work on each tab in sequence - from left to right, and from top to bottom within each tab.

Input File Tab

On the **Input File** tab you will describe the data file to be analyzed by the Joinpoint program and select the options for the modeling of the data.

Method and Parameters Tab

Use the **Method and Parameters** tab to specify the modeling method, the criteria used to determine the locations of the joinpoints, autocorrelated errors options, the model selection method, APC/AAPC/Tau Confidence Intervals, and the AAPC segment ranges.

Advanced Analysis Tools Tab

Use the **Advanced Analysis Tools** tab to set options for pairwise comparisons, Jump Model/Comparability Ratio Model or multi group clustering.

You can save all your settings in a Joinpoint session file by selecting **File > Save As...** from the top of the main window and choosing a name for the session file.

When you have made all of your selections on the Joinpoint tabs, click  to execute the Joinpoint session.

Note, the session file is saved as part of the output file, so you can select **Output > Retrieve Session** to retrieve the session that was used to create the output, even if that session file wasn't specifically saved.

Once you have created your session, and you want to create a copy of this session that you can use in the future with different data sets, select **File > Save Session Template As**. See [Joinpoint Session Template](#) for more information.

Related content

- [Preferences](#)
- [Joinpoint Session Template](#) A file type that allows users to save a Joinpoint Session and reopen it with a different data file.
- [Input File Tab](#)
- [Method and Parameters Tab](#)
- [Advanced Analysis Tools Tab](#)

Input File Tab

The purpose of the Input File tab is to set up your regression analysis by identifying the input data file, selecting the model type, selecting and defining the dependent, independent, and any by variables, selecting the heteroscedastic errors option.

Related content

- [Input Data File](#)
- [By-variables](#)
- [Independent Variable](#)
- [Dependent Variable](#)
- [Heteroscedastic Errors Option \(Weighted Least Squares\)](#)
- [Log Transformation](#)
- [How to Compute Age-Adjusted Rates in Joinpoint](#)
- [Delay Adjusted Statistics](#)
- [Data Grid](#)
- [Data Files Containing Cohorts with an Unequal Number of Observations](#) As of version 4.4.0.0, Joinpoint can process data files that contain cohorts with an unequal number of observations.

Input Data File

The input data file contains a dependent variable and an independent variable or covariate. The dependent variable values can either be directly provided in the input data file or Joinpoint can calculate the values if their components are provided. For example, Joinpoint can calculate age-adjusted rates if the associated count, population, standard population, and adjustment variable (typically age group) are provided.

The dependent variable may be an age-adjusted rate, crude rate, count, percentage, proportion, or other numeric value. The independent variable is typically the year. The data file may also contain standard errors (for rates) and by-variables such as age group, sex, or race. These variables (or columns of data) do not have to be in any particular order, but the data file (or rows of data) must be sorted by the by-variables and the independent variable. You identify the dependent variable by specifying whether it is calculated or provided, the type of variable to be modeled (i.e., crude rate, percent), and then selecting the appropriate items on the right (i.e. count, population, rate).

SEER*Stat provides an easy mechanism for creating Joinpoint data files and setting the controls on the Joinpoint **Input File** tab. In SEER*Stat, select "Matrix | Export" when viewing the SEER*Stat matrix to create a file in the format of a Joinpoint Input Data File. In Joinpoint, select "New">"Session..." from the **File** menu and then select the SEER*Stat Dictionary file. The input data file name, variable names and locations, and file format information will be loaded from the SEER*Stat Dictionary.

The Heteroscedastic Errors Option is automatically set from the SEER*Stat Dictionary, depending on whether or not the standard errors are included. If they are not included, Constant Variance (Homoscedasticity) is selected. If they are included, Standard Error is selected.

Related content

- [Format of the Joinpoint Input Data File](#)
- [Using SEER*Stat to Create Joinpoint Input Data Files](#)
- [Data Errors That Stop An Analysis](#)

Format of the Joinpoint Input Data File

Each record in the data file has the following layout:

Variable(1) Variable(2) Variable(k)

The records must be sorted by: by-variables, independent variable.

File Contains Column Headers

Joinpoint can process data files where the first record contains the names of each variable in the data file. The names should be delimited with the same field delimiter used in the rest of the file. If the first record in your data file contains the variable names, please check the "File contains column headers" checkbox. Once this is checked, the data grid on the Input File tab will be updated accordingly.

Field Delimiters

The variables may be separated by a single character. This field delimiter may be a tab, space, comma, or semi-colon.

Missing Characters

Missing values may be represented by a space, a period, or the text string NA.

Using SEER*Stat to Create Joinpoint Input Data Files

Joinpoint requires a specific file format for the input data file. Therefore, the following export options *must* be used in SEER*Stat when creating data files to be used in Joinpoint:

- The data file must be a text file (use a .txt extension or a gzipped text file (use a .gz extension))
- The Output Variables option must be set to "Numeric Representation"
- The Line Delimiter option must be set to "DOS/Windows (CR/LF)"
- The Field Delimiter may be any one of the choices. However, if commas are used as the field delimiter then the option to "Remove All Thousand Separators (Commas)" must be checked.
- "Remove Flags (Footnote Characters)" must be selected.

The independent variable (year) must be an individual value. Therefore, we recommend you create a user-defined variable in SEER*Stat that does not contain any ranges. For example, when creating a new "Year of Diagnosis" variable remove the range variable (1973-2004) for all years combined. When using a SEER*Stat dictionary file (*.DIC), any independent variable value with a dash (*-*) located in its format label will not be used in the analysis. The label will be updated to contain the following text: (not used in calculations). This functionality was added so that users would not have to reproduce their SEER*Stat analyses in order to remove totals or sub-totals in their independent variable.

The records for each by-group must be contiguous. If N = number of records for the first by-group, lines 1 through N must contain all values for the first by-group. Within each by-group, the records must be sorted by the independent variable (this is typically year).

To be sure that the sort order is correct, on the **SEER*Stat Table** tab, put all table variables in the row dimension. Make the independent variable (year) the last variable in the list of row variables. Alternatively, you could set the independent variable as the column variable.

The SEER Cancer Statistics Review (CSR) uses the same Joinpoint Regression Program. Their analysis uses the "Standard Error" setting for the Heteroscedastic Errors Option with the standard error of the rate used as an estimate of the standard deviation. Using these options, and the current default number of permutations (4499), if the Joinpoint Regression Program chooses the model with 0 joinpoints, the annual percentage rate change will agree with the calculation of this value given in the CSR.

Note: to obtain the SEER*Stat Program visit the SEER web page at:

<http://seer.cancer.gov/seerstat/>

Dependent Variable

The dependent, or response, variable is the variable being tested in the model. You can either provide your dependent variable values in the input data set or you can have Joinpoint calculate them from other variables in the input data set (e.g. a crude rate can be calculated from variables with the count and the population).

Run Type

Calculated From Data File This option should be selected if you want Joinpoint to calculate your dependent variable. Joinpoint can calculate Crude Rates, Age-Adjusted Rates, Proportions and Percentages. The [table below](#) documents which variables are required for each type of dependent variable. Please note that if Joinpoint calculates the dependent variable, it will also compute the associated standard errors.

Provided In Data File This option should be selected if you are providing Joinpoint with your dependent variable values. For some analysis types, the standard error is also required. Please reference the [table below](#) for more information.

Delay Adjusted Data This option is only available if a SEER*Stat dictionary (DIC) file is used for the session that contains delay adjusted statistics. Joinpoint will read in and analyze the delay statistics for each cohort. See [Delay Adjusted Statistics](#) for more information.

Delay & Non-Delay Data This option is only available if a SEER*Stat dictionary (DIC) file is used for the session that contains delay adjusted statistics. Joinpoint will read in and analyze both the delay and non-delay statistics for each cohort. See [Delay Adjusted Statistics](#) for more information.

Types of Dependent Variables Accepted by Joinpoint:

The following dependent variable types can be analyzed by Joinpoint:

Count A count is typically a number of “events” (i.e. cancer cases or deaths). Joinpoint will accept counts in fractional form (e.g. 410.375).

Crude Rate A crude rate is the number of cases (or deaths) occurring in a specified population, usually expressed as the number of cases per 100,000 population at risk.

Joinpoint will accept counts and populations in fractional form (e.g., 410.375).

$$cruderate = \frac{count}{pop} \times 100,000$$

Age-Adjusted Rate An age-adjusted rate is a weighted average of crude rates, where the crude rates are calculated for different age groups and the weights are the proportions of persons in the corresponding age groups of a standard population. The age-adjusted rate for an age group comprised of the ages x through y is calculated using the following formula.

$$aarate_{x-y} = \sum_{i=x}^y \left[\left(\frac{count_i}{pop_i} \right) \times 100,000 \times \left(\frac{stdpop_i}{\sum_{j=x}^y stdpop_j} \right) \right]$$

Where $count_i$ is the number of cases in the i th age group, pop_i is the relevant population for the same age group, and $stdpop_i$ is the standard population for the same age group.

If $count_i$ and pop_i are both zero, Joinpoint treats that rate as zero. Joinpoint will not allow the unusual event where pop_i is zero but $count_i$ is non-zero.

If $count_i$ is greater than zero and pop_i is zero, Joinpoint will set pop_i to be equal to $count_i$ in order to compute the interim crude rate.

Joinpoint will accept counts, populations and standard populations in fractional form (e.g., 410.375).

For a detailed explanation on how to compute age-adjusted rates, see [How to Compute Age-Adjusted Rates](#).

Proportion A proportion is a ratio of a numerator to a denominator where the result is between 0 and 1. Joinpoint will not analyze cohorts that have proportions greater than 1 when such proportions are calculated. Joinpoint will accept numerators and denominators in fractional form (e.g., 410.375).

$$proportion = \frac{numerator}{denominator}$$

Percent A percent is a fraction or ratio with 100 understood as the denominator. Joinpoint will not allow percentage values greater than 100 when such percentages are calculated. Joinpoint will accept numerators and denominators in fractional form (e.g., 410.375).

$$percent = \frac{numerator}{denominator} \times 100$$

Other Select this category for all other types of numeric variables that do not fall into one of the previous categories (i.e. temperatures over time).

Standard Error Calculations:

When Joinpoint computes the Dependent Variable, the following formulas will be used to compute the associated standard errors.

Standard Error for Crude Rate

This calculation assumes that the counts have Poisson distributions.

$$SE_{crude} = \frac{\sqrt{count}}{population} \times 100,000$$

Standard Error for an Age-Adjusted Rate

This calculation assumes that the counts have Poisson distributions. Suppose that the age-adjusted rate is for an age group comprised of ages x through y .

$$SE_{AArate} = \sqrt{\sum_{i=x}^y \left(\frac{stdpop_i}{\sum_{j=x}^y stdpop_j} \right)^2 \times \left(\frac{count_i}{population_i^2} \right)} \times 100,000$$

Standard Error for Proportions and Percentages

For consistency, assume that percents are transformed to proportions before analysis.

$$SE_{proportion} = \sqrt{\frac{p_i(1-p_i)}{n_i}}$$

Where p_i is the proportion and n_i is the sample size of the i th observation.

$$SE_{percentage} = \sqrt{\frac{p_i(1-p_i)}{n_i}} \times 100$$

Edits performed on the Dependent Variable:

- When a session is executed, Joinpoint will first edit the input data file. For a list and description of all data errors and warnings produced by Joinpoint, please see the "Job Execution - Errors and Warnings" help section.

Variables:

One or more of the following variables will need to be specified in the input data file depending on the Calculated/Provided, Type, and Heteroscedastic Errors Options selections:

- Count Variable
- Population Variable
- Age-Adjusted Rate
- Standard Error
- Adjustment Variable
- Standard Population
- Numerator Variable – this variable replaces the “Count” variable when appropriate
- Denominator Variable – this variable replaces the “Population” variable when appropriate

Joinpoint does allow any variable value to be provided in fractional form (e.g. 410.375).

| Calculate | Type | Heteroscedastic Errors | Required Variables | Required Variables | Required Variables | Required Variables |
|-------------|--------------------|------------------------|--------------------|--------------------|--------------------|--------------------|
| No | Count | Constant Variance | Count | | | |
| No | Count | Standard Error | Count | Standard Error | | |
| No | Count | Poisson Variance | Count | | | |
| No | Crude Rates | Constant Variance | Rate | | | |
| No | Crude Rates | Standard Error | Rate | Standard Error | | |
| Crude Rates | Poisson Variance | n/a | | | | |
| No | Age-Adjusted Rates | Constant Variance | Rate | | | |
| No | Age-Adjusted Rates | Standard Error | Rate | Standard Error | | |
| No | Age-Adjusted Rates | Poisson Variance | n/a | | | |
| No | Proportion | Constant Variance | Proportion | | | |

| | | | | | | |
|-----|--------------------|-------------------|------------|----------------|-----------|--------------|
| No | Proportion | Standard Error | Proportion | Standard Error | | |
| No | Proportion | Poisson Variance | n/a | | | |
| No | Percent | Constant Variance | Percent | | | |
| No | Percent | Standard Error | Percent | Standard Error | | |
| No | Percent | Poisson Variance | n/a | | | |
| No | Other | Constant Variance | Other | | | |
| No | Other | Standard Error | Other | Standard Error | | |
| No | Other | Poisson Variance | n/a | | | |
| Yes | Count | Constant Variance | n/a | | | |
| Yes | Count | Standard Error | n/a | | | |
| Yes | Count | Poisson Variance | n/a | | | |
| Yes | Crude Rates | Constant Variance | Count | Population | | |
| Yes | Crude Rates | Standard Error | Count | Population | | |
| Yes | Crude Rates | Poisson Variance | Count | Population | | |
| Yes | Age-Adjusted Rates | Constant Variance | Count | Population | Age Group | Standard Pop |
| Yes | Age-Adjusted Rates | Standard Error | Count | Population | Age Group | Standard Pop |
| Yes | Age-Adjusted Rates | Poisson Variance | n/a | | | |
| Yes | Proportion | Constant Variance | Count | Population | | |
| Yes | Proportion | Standard Error | Count | Population | | |
| Yes | Proportion | Poisson Variance | n/a | | | |
| Yes | Percent | Constant Variance | Count | Population | | |
| Yes | Percent | Standard Error | Count | Population | | |

| | | | | | | |
|-----|---------|-------------------|-----|--|--|--|
| Yes | Percent | Poisson Variance | n/a | | | |
| Yes | Other | Constant Variance | n/a | | | |
| Yes | Other | Standard Error | n/a | | | |
| Yes | Other | Poisson Variance | n/a | | | |

Related content

- [Data Warnings Displayed by Joinpoint](#)
- [Data Errors That Stop An Analysis](#)

Independent Variable

The independent variable is defined on the [Input File](#) tab by its name or column position. These controls will automatically be set when the SEER*Stat Export Dictionary is loaded. If you created your data file with software other than SEER*Stat you must set these controls.

The data file must be sorted first by all by-variables and then the independent variable, unless you want Joinpoint to calculate age-adjusted rates -- in which the data file would also be sorted by the adjustment variable. For a detailed description of how age-adjusted rates are calculated and what the required variable order, please go to the [How to Compute Age-Adjusted Rates](#) help section.

The independent variable or covariate is typically the year of diagnosis or year of death. You are limited to a 40-character label for the independent variable name.

The "Shift Data Points by" option allows all the values for the independent variable to be shifted up by a fixed value. This is done by simply entering a value for "Shift Data Points by". For example, if your independent variable is years (input as 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 you would 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.

Note: this value cannot be greater than the maximum interval between data points.

One reason for employing this option would be because each data point represents a summary of data collected over a time interval. For example, cancer incidence or mortality data is often collected over the course of a year, and is usually entered as a whole year value, e.g. 1990, 1991. Instead, one may want to shift by half a year so that the data point is represented as the midpoint of the interval, e.g. 1990.5, 1991.5). 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). If the data points are not shifted, the results may be counterintuitive. For example, without a shift, a joinpoint located at 1989.50 represents the beginning of 1990, and 1990.49 represents the end of 1990. If the points are shifted by half a year, then 1990.00 represents the start of the year and 1990.99 represents the end of the year.

Related content

- [Data Warnings Displayed by Joinpoint](#)
- [Data Errors That Stop An Analysis](#)
- [Define Format](#) Joinpoint allows the independent variable to be renamed, its values to be recoded, and display labels assigned to each value. If the variable is recoded, the recoded values are used in the Joinpoint calculations.

Define Format

Joinpoint allows the independent variable to be renamed, its values to be recoded, and display labels assigned to each value. If the variable is recoded, the recoded values are used in the Joinpoint calculations.

The following attributes can be set for the Independent variable:

Variable name:

The name must be of reasonable length and it is recommended that the label be concise (short) for ease of viewing in the output.

Independent Variable Coding Information (Value = Recode (Calculation value) = Display Label):

An example of an Independent Variable format would be:

Variable Name: Year of Diagnosis - 1993-2013

| Value | Value Used In Calculation | Use Custom Display Labels |
|-------|---------------------------|---------------------------|
| 0 | 1993 | Ninety Three |
| 1 | 1994 | Ninety Four |
| 2 | 1995 | Ninety Five |
| Etc. | | |

Recode/Calculation values MUST be numeric. Recode values and Display Labels must also be a reasonable length and it is recommended they be concise for ease of viewing in the output.

When using a SEER*Stat dictionary file (*.DIC), any independent variable value with a dash (*-*) located in its format label will not be used in the analysis. The label will be updated to contain the following text: "not used in calculations". This functionality was added so that users would not have to reproduce their SEER*Stat analyses in order to remove totals or sub-totals in their independent variable.

By-variables

A separate analysis will be performed for each by-group. The by-variable information is automatically loaded when you open the SEER*Stat Export dictionary associated with your data file. If you created your data file with software other than SEER*Stat, you must manually add by-variables. To add a by-variable, click the "Add..." button, select the desired variable, and click OK. Select the "Exclude..." button if you want to remove cohorts from the next processing.

Related content

- [Format of the Joinpoint Input Data File](#)
- [Define Format](#)
- [Include/Exclude Select Cohorts](#) Specify which cohorts from the data should actually be processed.

Define Format

Joinpoint allows by-variables to be renamed and labels to be assigned to each of their associated values. If the format for a variable is changed, the new format will be saved with the session and when the session is executed, the resultant output will reflect the changes.

The following attributes can be set for each By-variable:

Variable name:

The name must be of reasonable length and it is recommended that the label be concise (short) for ease of viewing in the output.

Coding Information (Value = Label):

Each By-variable must consist of a series of values. Each value can be assigned a label. Users are allowed to change the label for any value. If no labels were initially found for the by-variable, then the label for each value is defaulted to the associated value.

A By-variable format example would be:

1 = Male and Female

2 = Male

3 = Female

Labels may have any meaningful text including numbers, letters, and special characters. They must also be a reasonable length and it is recommended that the label be concise for ease of viewing in the output.

Include/Exclude Select Cohorts

Specify which cohorts from the data should actually be processed.

The “Exclude...” option should be used when you want Joinpoint to analyze a subset of the cohorts found in your input data file. With this feature, you can select those cohorts you wish to analyze and those that you do not.

Regardless of the number of cohorts you select to include in your analysis, all cohorts will still be present in the output data file. Those that were “excluded” from the analysis, will be marked with an ✕ and labeled as “Excluded By User”. Excluded cohorts will only display the observed data values in the graph and data grids.

If one or more cohorts are excluded when you execute your session, Joinpoint will display those cohorts in the “Issues Found in Cohort Data” window. The cohorts will be marked as “User Selected to Exclude This Cohort” and the user can choose to cancel the job.

Selecting Cohorts to include using the “By-Variable” categories:

Choose By-Variables to Include

This list will contain all categories for each of your by-variables. You can use this list to include or exclude cohort categories (groups) in your Joinpoint analysis.

Options:

Add By-Group - This button will add all cohorts containing the selected By-Variable category to the “Cohorts to Be Included” list.

Remove By-Group - This button will remove all cohorts containing the selected By-Variable category from the “Cohorts to Be Included” list.

Selecting Individual Cohorts to Include:

Choose Individual Cohorts to Include

This list will contain all the individual cohorts found in your input data file. By using this list, you can include or exclude individual cohorts from your Joinpoint analysis.

Options:

Add All - this button will add all items in the “Choose Individual Cohorts to Include” list to the “Cohorts to Be Included” list.

Add -> - this button will add any selected items in the “Choose Individual Cohorts to Include” list to the “Cohorts to Be Included” list.

<- Remove - this button will remove all selected items from the “Cohorts to Be Included” list.

Remove All - this button will remove all items in the “Cohorts to Be Included” list.

Cohorts That Will Be Modeled By Joinpoint:

Cohorts to Be Included: This list will contain all cohorts to be analyzed by Joinpoint. If the list is empty, then the Joinpoint session cannot be executed. There must be one or more cohorts listed.

be the number of cohorts found in your input data file.

Total Number of Cohorts to be Processed: In the lower right-hand corner of the Select Cohorts To Run” dialog will be the total number of cohorts you have selected to have Joinpoint analyze.

How to Compute Age-Adjusted Rates in Joinpoint

For a detailed explanation on how to calculate age-adjusted rates, please refer to the following web site:

<http://seer.cancer.gov/seerstat/tutorials/aarates/step1.html>

For Joinpoint to compute age-adjusted rates, each input data record must have the following variables IN THIS ORDER:

1. All cohort-defining variables
2. Independent variable
3. Adjustment variable (this is typically an age variable)
4. Count
5. Population
6. Standard Population

The following requirements will be applied to each cohort when age-adjusted rates are computed by Joinpoint:

- The data has to be sorted using the following variable order:
 - All cohort defining variables
 - Independent variable
 - Adjustment variable
- Count values must be greater than or equal to zero.
- Population values must be greater than or equal to zero
- If the population value is zero, the associated count must also be zero.
- The count must not be greater than the associated population.
- Standard Populations values must be greater than zero.
- The Adjustment Variable must have the same categories (in the same order) for each By Variable/Independent Variable combination.

Below is an example of an input data file correctly structured for Joinpoint to compute age-adjusted rates. The data file contains one by variable (sex:1 = Male, 2 = Female) and the independent variable is Year (values are 1980 thru 2000 inclusive). For each Sex/Year combination, there is a complete set of adjustment variable values (00 years to 85+ years). Please note that the standard population appears on every data record and the value corresponds to the associated age group.

| By Variable (Sex) | Independent Variable (Year) | Adjustment Variable (Age) | Count | Population | US 2000 Standard Populations |
|--------------------------|------------------------------------|----------------------------------|--------------|-------------------|-------------------------------------|
| 1 | 1980 | 00 years | 29 | 139,879 | 3,794,901 |
| 1 | 1980 | 01-04 years | 87 | 553,189 | 15,191,619 |
| 1 | 1980 | 05-09 years | 67 | 736,212 | 19,919,840 |
| 1 | 1980 | 10-14 years | 71 | 770,999 | 20,056,779 |
| 1 | 1980 | 15-19 years | 87 | 651,309 | 19,819,518 |

| | | | | | |
|-----|------|-------------|-------|---------|------------|
| 1 | 1980 | 20-24 years | 177 | 639,159 | 18,257,225 |
| 1 | 1980 | 25-29 years | 290 | 676,354 | 17,722,067 |
| 1 | 1980 | 30-34 years | 657 | 736,557 | 19,511,370 |
| 1 | 1980 | 35-39 years | 1,072 | 724,826 | 22,179,956 |
| 1 | 1980 | 40-44 years | 1,691 | 700,200 | 22,479,229 |
| 1 | 1980 | 45-49 years | 2,428 | 617,437 | 19,805,793 |
| 1 | 1980 | 50-54 years | 2,931 | 516,541 | 17,224,359 |
| 1 | 1980 | 55-59 years | 2,881 | 361,170 | 13,307,234 |
| 1 | 1980 | 60-64 years | 2,817 | 259,440 | 10,654,272 |
| 1 | 1980 | 65-69 years | 2,817 | 206,204 | 9,409,940 |
| 1 | 1980 | 70-74 years | 2,744 | 172,087 | 8,725,574 |
| 1 | 1980 | 75-79 years | 2,634 | 142,958 | 7,414,559 |
| 1 | 1980 | 80-84 years | 1,884 | 99,654 | 4,900,234 |
| 1 | 1980 | 85+ years | 1,701 | 92,692 | 4,259,173 |
| 1 | 1981 | 00 years | 34 | 141,456 | 3,794,901 |
| 1 | 1981 | 01-04 years | 97 | 562,059 | 15,191,619 |
| 1 | 1981 | 05-09 years | 72 | 738,302 | 19,919,840 |
| ... | ... | ... | ... | ... | ... |
| 1 | 2000 | 80-84 years | 2,176 | 100,249 | 4,900,234 |
| 1 | 2000 | 85+ years | 2,001 | 94,875 | 4,259,173 |
| 1 | 2000 | 80-84 years | 2,176 | 100,249 | 4,900,234 |
| 2 | 1980 | 00 years | 37 | 159,829 | 3,794,901 |
| 2 | 1980 | 01-04 years | 101 | 562,001 | 15,191,619 |
| ... | ... | ... | ... | ... | ... |
| 2 | 2000 | 85+ years | 2,521 | 96,228 | 4,259,173 |

Below is an example of how an age-adjusted rate is computed. The age-adjusted rate is 400.3 and is the sum of the interim values over all age groups. The interim value in each row is equal to (Crude Rate * Age Distribution of Std Pop).

| Age | Count | Population | Crude Rate | US Standard Populations | Age Distribution of Std Pop | Interim Value |
|-------------|-------|------------|------------|-------------------------|-----------------------------|---------------|
| 00 years | 29 | 139,879 | 20.7 | 3,794,901 | 0.013818 | 0.29 |
| 01-04 years | 87 | 553,189 | 15.7 | 15,191,619 | 0.055316 | 0.87 |

| | | | | | | |
|-------------|-------|---------|---------|-------------|----------|--------------|
| 05-09 years | 67 | 736,212 | 9.1 | 19,919,840 | 0.072532 | 0.66 |
| 10-14 years | 71 | 770,999 | 9.2 | 20,056,779 | 0.073031 | 0.67 |
| 15-19 years | 87 | 651,309 | 13.4 | 19,819,518 | 0.072167 | 0.96 |
| 20-24 years | 177 | 639,159 | 27.7 | 18,257,225 | 0.066478 | 1.84 |
| 25-29 years | 290 | 676,354 | 42.9 | 17,722,067 | 0.064530 | 2.77 |
| 30-34 years | 657 | 736,557 | 89.2 | 19,511,370 | 0.071045 | 6.34 |
| 35-39 years | 1,072 | 724,826 | 147.9 | 22,179,956 | 0.080762 | 11.94 |
| 40-44 years | 1,691 | 700,200 | 241.5 | 22,479,229 | 0.081852 | 19.77 |
| 45-49 years | 2,428 | 617,437 | 393.2 | 19,805,793 | 0.072117 | 28.36 |
| 50-54 years | 2,931 | 516,541 | 567.4 | 17,224,359 | 0.062718 | 35.59 |
| 55-59 years | 2,881 | 361,170 | 797.7 | 13,307,234 | 0.048454 | 38.65 |
| 60-64 years | 2,817 | 259,440 | 1,083.8 | 10,654,272 | 0.038794 | 42.12 |
| 65-69 years | 2,817 | 206,204 | 1,366.1 | 9,409,940 | 0.034264 | 46.81 |
| 70-74 years | 2,744 | 172,087 | 1,594.5 | 8,725,574 | 0.031772 | 50.66 |
| 75-79 years | 2,634 | 142,958 | 1,842.5 | 7,414,559 | 0.026998 | 49.74 |
| 80-84 years | 1,884 | 99,654 | 1,890.5 | 4,900,234 | 0.017843 | 33.73 |
| 85+ years | 1,701 | 92,692 | 1,839.4 | 4,259,173 | 0.015509 | 28.53 |
| All Ages | | | | 274,633,642 | 1.000000 | 400.3 |

Log Transformation

Select a linear or log-linear model, based on whether or not you would like the dependent variable to be log-transformed.

- No - Linear Model: $y = x'\beta + e$
- Yes - Log-linear Model: $\ln(y) = x'\beta + e$

For a recommendation on which model to select for your analysis, please reference this [FAQ](#).

Heteroscedastic Errors Option (Weighted Least Squares)

Select one of the three Heteroscedastic Errors Options. You must then specify the variable(s) required for that option.

The error random variable in a model is homoscedastic if the variance of the error is constant; otherwise, the error is heteroscedastic. Often the homoscedastic assumption is violated, particularly when the variance of the error varies with time. The option allows the user to choose between a model where errors are assumed to have constant variance (homoscedasticity) and a model where nonconstant variance (heteroscedasticity) is assumed.

Heteroscedasticity is handled by Joinpoint using **weighted least squares** (WLS). The weights in WLS are the reciprocal of the variance and can be specified in several ways. The “Standard Error (Provided)” option allows the user full flexibility to specify the standard error at each time period. These standard errors can come from the output of SEER*Stat.

The last option, “Poisson variance”, estimates the non-constant variance by assuming the dependent variable counts follow a Poisson distribution. If “Crude Rate” is selected, you must choose to have Joinpoint calculate the rate and therefore provide both a count and a population.

Note: The Constant Variance (Homoscedasticity) option is automatically selected if your input data file contains only 2 fields.

Related content

- [Constant Variance \(Homoscedasticity\)](#)
- [Standard Error \(Provided\)](#)
- [Poisson Variance](#)

Constant Variance (Homoscedasticity)

This selection assumes the random errors in the regression model are homoscedastic (have constant variance) and estimates the regression coefficients by ordinary least squares for both model $\ln(y) = xb$ and model $y = xb$.

Note: This option is automatically selected if your input data file contains only 2 fields.

Adding 0.5 to Zero Counts

When analyzing counts or computing crude rates, Joinpoint will add 0.5 to any zero count under the circumstances listed [here](#).

Related content

- [Standard Error \(Provided\)](#)
- [Poisson Variance](#)

Standard Error (Provided)

This is the default selection. It assumes that the random errors are heteroscedastic (have non-constant variance) and estimates the regression coefficients by weighted least squares, where weights at each point are:

- For model $y = xb$

$w = 1/v$, where v is the square of the std dev that has been input for that point.

- 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. (Motivated by delta method.)

Related content

- [Constant Variance \(Homoscedasticity\)](#)
- [Poisson Variance](#)

Poisson Variance

With Provided Count

This selection assumes the dependent variable is $y = c$, where c is the adjusted count, which equals either:

- The count, if the count is greater than zero.
- The count plus one half, if the count is equal to zero.

Assume the random errors are Poisson, and estimate the regression coefficients by weighted least squares, where weights at each point are:

- For model $y = xb$
 $w = 1/c$, where c is the adjusted count for that point.
- For model $\ln(y) = xb$
 $w = c$, where c is the adjusted count for that point. (Motivated by delta method.)

With Calculated Crude Rate

This selection assumes the dependent variable is $y = c/p$, where c is the adjusted count and p is the population. The adjusted count will be equal to either:

- The count, if the count is greater than zero.
- The count plus one half, if the count is equal to zero.

Assume the random errors are Poisson, and estimate the regression coefficients by weighted least squares, where weights at each point are:

- For model $y = xb$
 $w = p^2/c$, where c is the adjusted count and p^2 is the square of the population for that point..
- For model $\ln(y) = xb$
 $w = c$, where c is the adjusted count for that point. (Motivated by delta method.)

Adding 0.5 to Zero Counts

When analyzing counts or computing crude rates, Joinpoint will add 0.5 to any zero count under the circumstances listed [here](#).

Related content

- [Constant Variance \(Homoscedasticity\)](#)
- [Standard Error \(Provided\)](#)

Delay Adjusted Statistics

Joinpoint has the ability to process both delay and non-delay adjusted statistics exported from SEER*Stat. If a new Joinpoint session is created using a SEER*Stat dictionary (DIC) file that contains delay adjusted statistics, two new "Run Types" will be available within the Dependent Variable section on the Input File tab. The following options will be provided within this control:

Delay Adjusted Data - by selecting this option, Joinpoint will only read in and analyze the selected delay adjusted statistic.

Delay & Non-Delay Data - by selecting this option, Joinpoint will read in and analyze both the delay and non-delay statistics for each cohort. The delay and non-delay model graphs and associated Joinpoint statistics will appear together for each cohort.

For information about delay-adjusted adjusted incidence data, please go to the NCI [website](#).

Data Grid

The Data Grid provides users with a snapshot of their data file and shows them how Joinpoint will parse the file.

Once a user selects their input data file, the Data Grid will be populated with the contents of the file. The contents of the grid are updated whenever the user adjusts the Delimiters, Missing Character, or File Contains Column Headers controls.

Users should use the Delimiters, Missing Character, and File Contains Column Headers controls until their data is correctly displayed in the Data Grid. Once the appropriate settings have been made, then users can go and set the other variables (independent, dependent, and by) and controls on the tab.

Users can select to display only a set number of records from their input file. The choices to display the following number of lines: 5, 10, 20, 50, or 100.

Method and Parameters Tab

The purpose of the **Method and Parameters** tab is to specify the:

- Modeling method (Grid Search or Hudson's).
- MADWD - Minimum APC Difference Worth Detecting.
- Constraints on the location(s) of the joinpoints.
- Number of joinpoints.
- Autocorrelated errors options.
- Model selection method.
 - Permutation Test
 - Bayesian Information Criterion
 - BIC3
 - Modified BIC
 - Weighted BIC
 - Weighted BIC Alternative
 - Data Dependent Selection
- APC/AAPC/Tau Confidence Intervals.
- AAPC Segment Ranges.

Related content

- [Number of Joinpoints](#)
- [Method – Grid Search or Hudson's](#)
- [MADWD Minimum APC Difference Worth Detecting in Joinpoint Fitting](#)
- [Autocorrelated Errors Option](#)
- [Minimum Number of Observations Required](#)
- [Model Selection Method](#)
- [APC/AAPC/Tau Confidence Intervals](#)
- [AAPC Segment Ranges](#)

Data Files Containing Cohorts with an Unequal Number of Observations

As of version 4.4.0.0, Joinpoint can process data files that contain cohorts with an unequal number of observations.

Reading the Input File and Populating the Independent Variable:

When all of the input variables on the Input File tab of the session have been specified, Joinpoint will read the data file. The Independent variable format will be populated with all the unique variable values found in the data file and the Maximum Number of Joinpoints (on the Method and Parameters tab) will be assigned based on the cohort with the maximum number of observations. If cohorts with a varying number of observations are located, Joinpoint will not initially indicate that the file contains cohorts of this type. Only when the session is executed will Joinpoint indicate if there are cohorts with an unequal number of observations.

Executing the Session

When the session is executed, a dialog will appear indicating that cohorts with fewer observations have been found. Joinpoint will display a total count of all cohorts whose maximum number of Joinpoints will follow the session maximum. All other cohorts will be organized by their total number of observations and will be assigned a lower maximum number of Joinpoints. That maximum number of Joinpoints will be based off of this table: Default Maximum Number of Joinpoints. The dialog will allow users to cancel the job or to continue.

What happens when the maximum number of Joinpoints is enough for all cohorts?

If you have specified a Maximum Number of Joinpoints that is low enough to satisfy the default maximum requirement for all cohorts, then (upon execution of the session) a message indicating that cohorts with an unequal number of observations were located and that no adjustment to the Maximum Number of Joinpoints for any cohort will be made.

Example:

For example, suppose you have a data file with the following three cohorts:

1. Male and Female 24 observations (data lines)
2. Male 21 observations (data lines)
3. Female 12 observations (data lines)

In this example, Joinpoint will default the Maximum Number of Joinpoints to 4 (please see the Default Maximum Number of Joinpoints). When the session is executed Joinpoint will indicate that cohorts with few observations than required to run with a setting of 4 Joinpoints were located. The following will be displayed on the warning dialog:

1 cohort with 24 observations. Maximum Joinpoints will remain set to 4.

1 cohort with 21 observations. Maximum Joinpoints will be set to 3.

- Male

1 cohort with 12 observations. Maximum Joinpoints will be set to 2.

- Female

Method – Grid Search or Hudson's

Joinpoint allows two different methods for model fitting - [Grid Search](#) or [Hudson's](#). The Grid Search has a finite number of discrete locations that are tested to find the best model fit, while Hudson's allows for continuous fitting.

As of Joinpoint version 4.5.0.0, Hudson's method is unavailable.

Minimum APC Difference Worth Detecting (MADWD) - In some situations Joinpoint may determine that there are one or more pairs of segments where the difference in the Annual Percent Change (APC) between some consecutive segments is quite small. Users can enter a minimum APC difference worth detecting as a minimum percentage point difference (Min. PPD). Joinpoint will start with the final selected model that was chosen using the specified analysis criteria and lower the number of Joinpoints until all consecutive segments pass the Min. PPD criteria.

Related content

- [Grid Search Method - Details](#)
- [Hudson's Method - Details](#)
- [MADWD Minimum APC Difference Worth Detecting in Joinpoint Fitting](#)

Grid Search Method - Details

New for Version 4.5.0.0

For version 4.5.0.0, we have made some changes to the layout and values for Grid Search. But don't worry, everything still works much as you remember from older versions. The old layout was a little confusing with parameters being shared between methods when they really shouldn't be. The new interface and values are more precise and specific to each method. Here's what's been changed and why:

- The "Number of Observations" controls have been moved under Grid Search. In the old interface, the number of observations controls were used for both Grid Search and Hudson's. But this resulted in some ambiguity for the parameter values. So we've moved the controls under Grid Search to make it clear that these are specific only to the Grid Search Method. Hudson's will have its own parameter controls (in time).
- The description for the "Number of Observations" controls has changed. The old descriptions used for these controls were not accurate or clear for all possible cases. We've reworded the descriptions for these controls so that they are accurate in all situations.
- The default values for "Number of Observations" have changed. You may notice that the default values for the number of observations are different than in previous versions. That is correct, but the new values mean the same thing. Due to the change in the descriptions for each control, the old default values were no longer correct. The new default values are accurate and obtain the same results as the old default values. If you load a session from an older version of Joinpoint, these parameters will automatically be updated to the new values. In this way, if you re-run an old session, you'll get the exact same results.
- The "Number of Joinpoints" controls have been moved after the method selection. The controls are now in the correct order. The maximum number of joinpoints is actually dependent on the settings for "Number of Observations", so it makes sense for the "Number of Joinpoints" controls to come after the "Number of Observations" controls.

How Grid Search Works

Using the default settings, the Grid Search Method only allows the joinpoints to occur exactly at observations. This does not, however, 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. So, the Grid Search Method creates a "grid" of all possible locations for joinpoints specified by the settings, and calculates the SSE at each one to find the best possible fit. With lower values for "Number of points to place between...", this method is computationally more efficient.

Min number of observations from a joinpoint to either end of the data (excluding the first or last joinpoint if it falls on an observation)

Use this parameter to keep joinpoints from being placed too close to the end points. This value must be at least one (1) and no more than nine (9), but the default is set to two (2). So, for example if this value is set to 2, there is annual data from 1990-2014, and joinpoints are restricted to fall on the data points, then the first possible joinpoint is 1992.

Min number of observations between two joinpoints (excluding any joinpoint if it falls on an observation)

Use this parameter to keep joinpoints from getting too close together. This value can be set as low as zero (0) and as high as nine (9), but the default is set to two (2). For example, if this value is set to 2, there is annual data from 1990-2014, joinpoints are restricted to fall on the data points and there was a joinpoint at 2000, then the closest possible other joinpoints would be at either 1997 or 2003.

The permutation tests are valid for any of the allowable choices for the above two settings (the minimum number of observations from a joinpoint to either end of the data, and the minimum number of observations between two joinpoints); however, some statistics (the standard error of the slope parameters and the associated *p*-values) cannot always be calculated when there are not at least three observations on a line segment (excluding observations at the joinpoints). See the statistical notes about [Estimated Regression Coefficients \(Beta\)](#). Similarly, some statistics cannot be calculated if a line segment in the joinpoint model is an exact fit.

Number of points to place between adjacent observed x values in the grid search

To allow joinpoints to occur between observations, specify *the number of points to place between adjacent observed x-values in the grid search* as 1-9. So, for example, if this value is set to 3, and the data is annual from 1990-2014, then the joinpoints could occur at values like 1995.25, 2000.5, or 2010.75, as well as on the observations. The minimum allowed value is 0 and the maximum is 9. Note that a fine grid search with this value set at >5 may take a long time to process. An alternative to this would be to use [Hudson's Method](#).

Related content

- [Hudson's Method - Details](#)
- [New default settings for Grid Search](#) Why did the default settings for Grid Search change?
- [Minimum Number of Observations Required](#)

Hudson's Method – Details

As of Joinpoint version 4.5.0.0, Hudson's method is unavailable.

In version 4.5.0.0, Hudson's Algorithm (used to estimate the exact best fitting location of joinpoints in continuous time) was temporarily disabled. It was determined that the algorithms for setting how close joinpoints can be to each other and to the ends of the series were not working properly when using the Hudson's Algorithm. When examining this issue in detail, it was determined that there were some complicated interactions between these settings and the use of the Hudson's Algorithm. We were not fully able to work out these issues in time for this release. Since version 4.5.0.0 had some important other updates, we decided to temporarily disable Hudson's Algorithm. We apologize for any inconvenience this may cause.

In previous versions, the placement of joinpoints was more constrained than it should have been given the settings. Also the description of the settings was incorrect for how they were used by Hudson's Method. We have now separated the settings for Grid Search and Hudson's Method to make this clearer.

Related content

- [Early Stopping Options](#)
- [Grid Search Method - Details](#)
- [Hudson's Method Disabled](#) Why is the Hudson's Method disabled?

MADWD

Minimum APC Difference Worth Detecting in Joinpoint Fitting

In some situations Joinpoint may determine that there are one or more pairs of segments where the difference in slopes (i.e. difference in the Annual Percent Change) between some consecutive segments is quite small. For cancer trends, this tends to occur in the larger cancer sites where there is increased statistical power to detect small differences. Often the analyst is focused on trying to interpret only larger differences, and would like to eliminate joinpoints where the APC difference falls below a specified threshold because they are too small to be of practical importance and/or may be difficult to interpret. It is important to note that this is a subjective criteria that overlays the statistical criteria which is applied first. We call this procedure Minimum APC Difference Worth Detecting (MADWD).

The user enters a minimum APC difference worth detecting as a minimum percentage point difference (Min PPD). For cancer incidence and mortality trends we have generally not found any APC differences less than 0.25 percentage points. A reasonable choice might be a Min PPD of 0.5 percentage points or 0.75 percentage points, although it is completely subjective and depends on the context of the analysis. The Min PPD should generally be selected before analyzing the data. This will guard against an analyst post hoc systematically varying the MIN PPD until they obtain results which conform with pre-conceived notions of what the trends should look like.

The procedure first uses the specified statistical criteria to determine the number of joinpoints. Each pair of segments is then tested to determine if the APC difference is less than the specified cutoff level. If any pairs of segments fails (i.e. APC difference less than the specified level) then a new joinpoint model with one less segment is fit, and that model is tested to see if all segments pass the MADWD criteria. This process is continued until we find the joinpoint model where all consecutive segments pass the criteria.

More technically consider the joinpoint regression model with $\kappa + 1$ segments: for $i = 1, \dots, n$,

$$y_i = \log(r_i) = \beta_0 + \beta_1 x_i + \delta_1(x - \tau_1)^+ + \dots + \delta_\kappa(x - \tau_\kappa)^+$$

where $a^+ = \max(a, 0)$ and κ is the unknown number of joinpoints. Define the APC difference between the j -th and $(j + 1)$ -st segments as

$$|APC_{j+1} - APC_j| = \begin{cases} 100|e^{\beta_1 + \delta_1} - e^{\beta_1}| \approx |\delta_1| & \text{for } j = 1; \\ 100|e^{\beta_1 + \sum_{i=1}^j \delta_i} - e^{\beta_1 + \sum_{i=1}^{j-1} \delta_i}| \approx |\delta_j| & \text{for } j = 2, \dots, \kappa \end{cases}$$

and we estimate it as

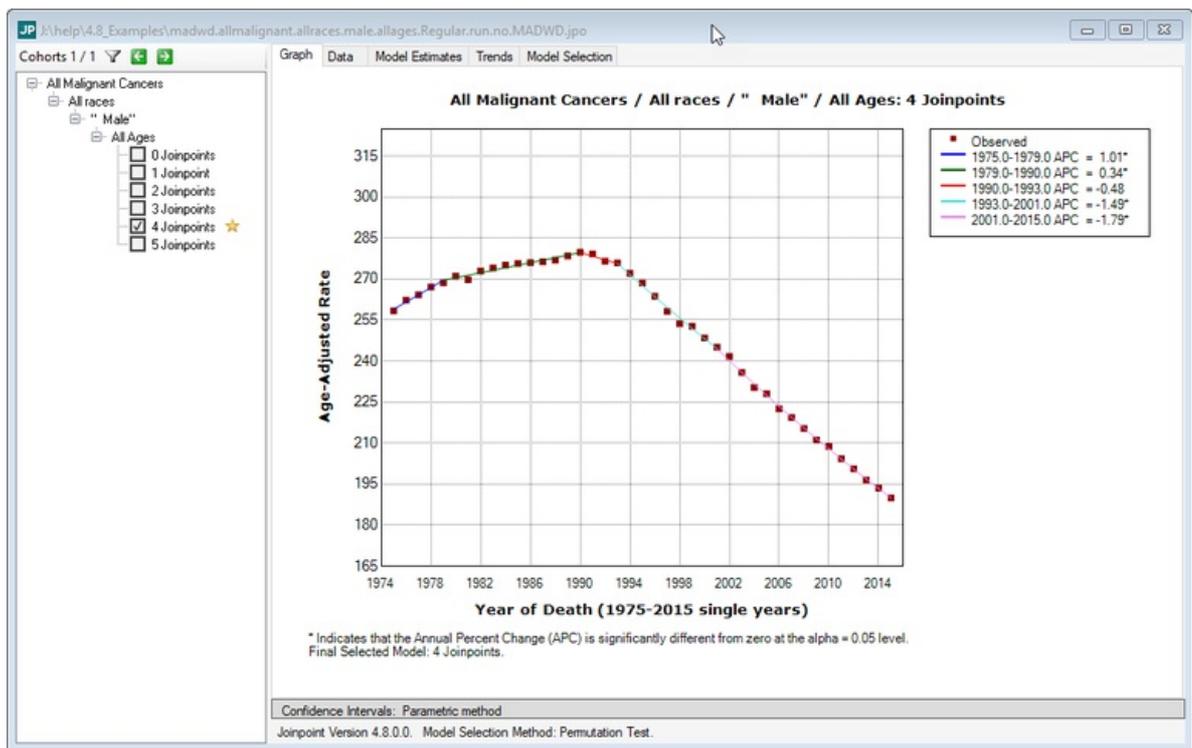
$$d_{j,j+1} = \begin{cases} 100|e^{\hat{\beta}_1 + \hat{\delta}_1} - e^{\hat{\beta}_1}| & \text{for } j = 1; \\ 100|e^{\hat{\beta}_1 + \sum_{i=1}^j \hat{\delta}_i} - e^{\hat{\beta}_1 + \sum_{i=1}^{j-1} \hat{\delta}_i}| & \text{for } j = 2, \dots, \hat{\kappa} \end{cases}$$

| | |
|---------|--|
| Step 1. | Set the MADWD, say at ψ . |
| Step 2. | For each $k = 0, \dots, \hat{\kappa}$, consider the best model fitted by Joinpoint. |
| Step 3. | For $k = \hat{\kappa} (> 0)$, examine if all the adjacent segments meet the MADWD criteria. That is, check if $d_{j,j+1} \geq \psi$ for all $j = 1, \dots, k$. |
| Step 4. | If $k = 0$ or all the neighboring segments meet the MADWD criteria, then stop. |

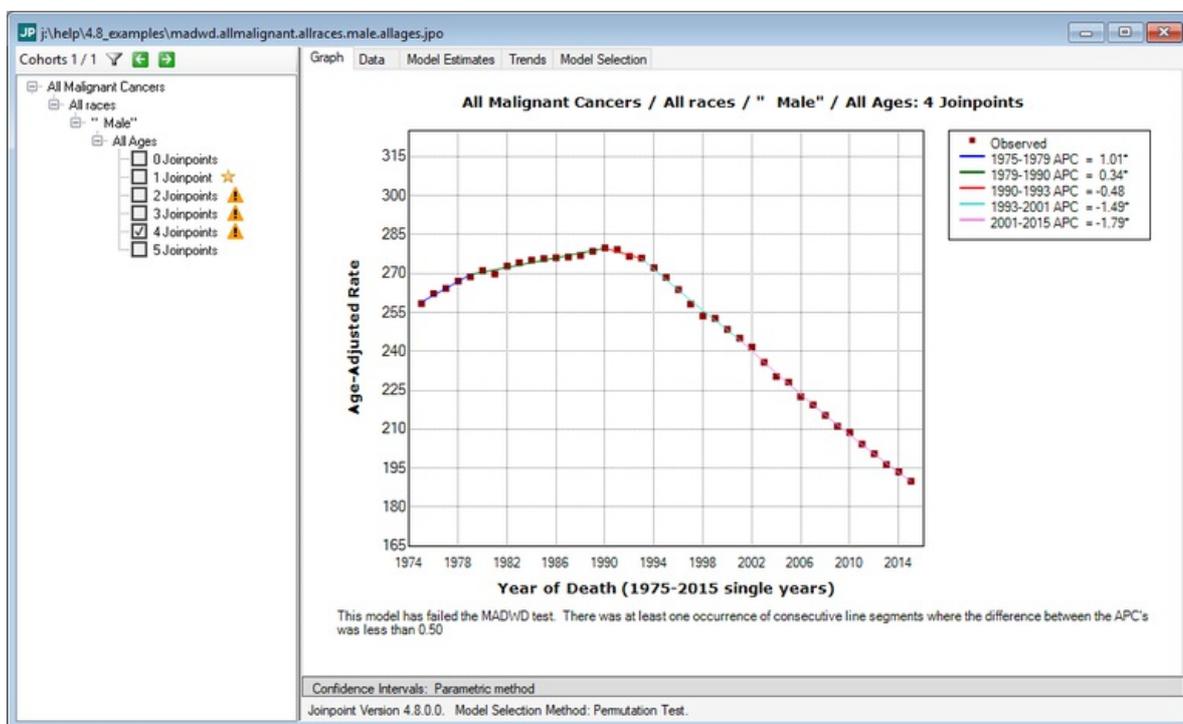
Step 5. If there are at least one neighboring segments not meeting the MADWD criteria, reduce k to $\hat{k} - 1$ and repeat Steps 3-5.

Example

Below is an example of the final selected model for US mortality 1975-2016 for all malignant cancers, all races, male, all ages made from a Joinpoint analysis not using the MADWD criteria. A 4 joinpoint model was selected, but some of the changes in the APC between consecutive segments are small.



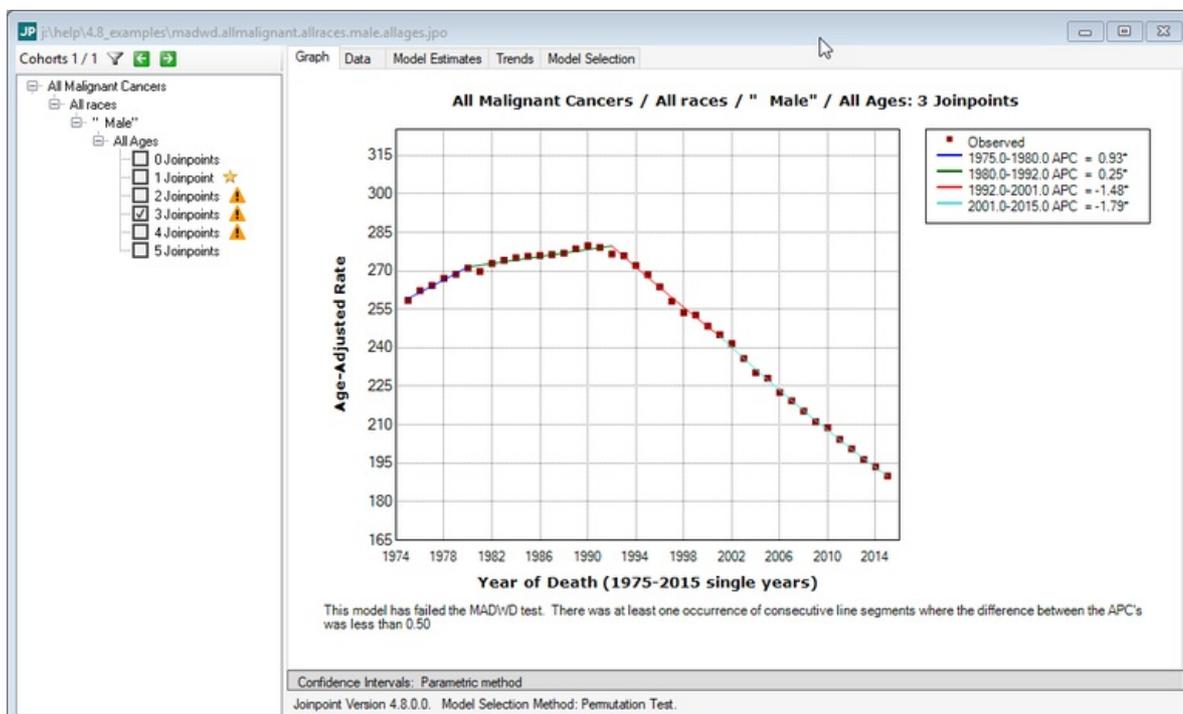
Now running the same session again using MADWD with a Minimum Percentage Point Difference (Min PPD) set to 0.5, we get the following:



The 4 Jointpoint model failed the MADWD test because the percentage point difference between the 1993-2001 and 2001-015 segments is 0.30.

| Minimum APC Difference Worth Detecting (MADWD) | | | | | | | |
|--|-------|----------------------|-----------------------|-----------------------|-------------------------|------------------|--------------|
| Cohort | Model | Number of Joinpoints | Segment 1 Range (APC) | Segment 2 Range (APC) | Absolute APC Difference | Min. PPD Allowed | MADWD Result |
| All Ages | #5 | 4 Jointpoint(s) | 1975-1979 (1.01) | 1979-1990 (0.34) | 0.68 | 0.50 | Pass |
| All Ages | #5 | 4 Jointpoint(s) | 1979-1990 (0.34) | 1990-1993 (-0.48) | 0.82 | 0.50 | Pass |
| All Ages | #5 | 4 Jointpoint(s) | 1990-1993 (-0.48) | 1993-2001 (-1.49) | 1.01 | 0.50 | Pass |
| All Ages | #5 | 4 Jointpoint(s) | 1993-2001 (-1.49) | 2001-2015 (-1.79) | 0.30 | 0.50 | Fail |

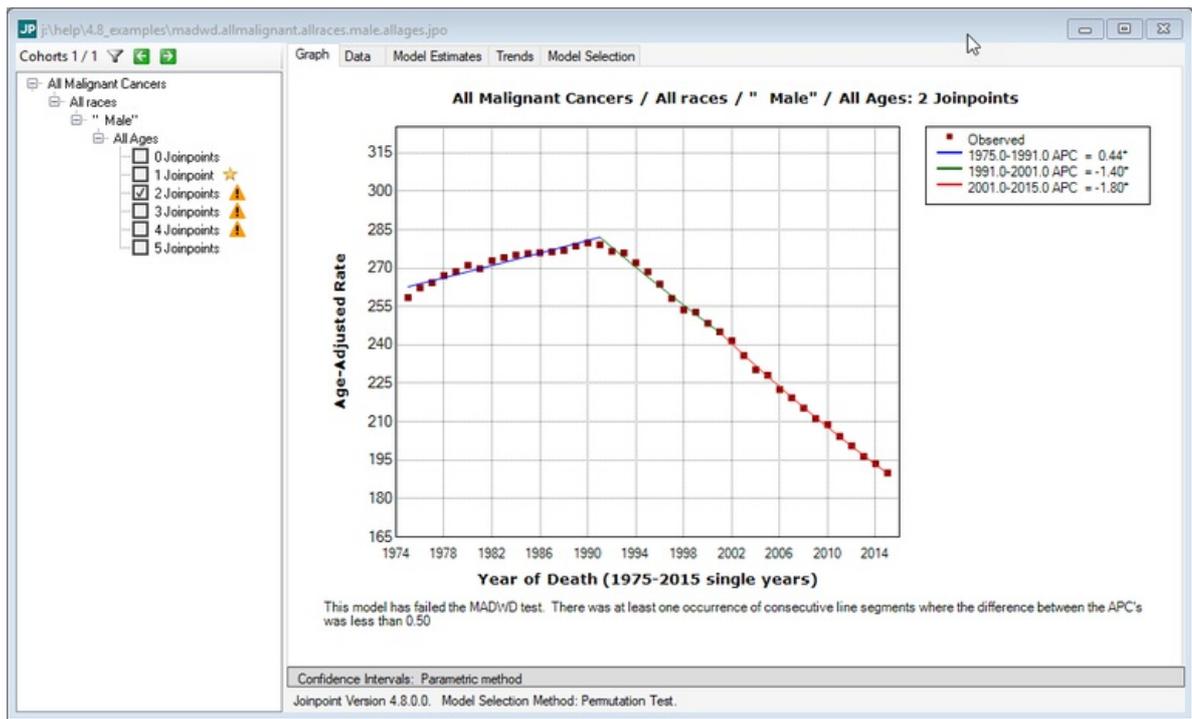
Below is the fit using 3 jointpoints.



The 3 Jointpoint model failed the MADWD test because the percentage point difference between the 1992-2001 and 2001-2015 segments is 0.31.

| Minimum APC Difference Worth Detecting (MADWD) | | | | | | | |
|--|-------|----------------------|-----------------------|-----------------------|-------------------------|------------------|--------------|
| Cohort | Model | Number of Joinpoints | Segment 1 Range (APC) | Segment 2 Range (APC) | Absolute APC Difference | Min. PPD Allowed | MADWD Result |
| All Ages | #4 | 3 Jointpoint(s) | 1975-1980 (0.93) | 1980-1992 (0.25) | 0.68 | 0.50 | Pass |
| All Ages | #4 | 3 Jointpoint(s) | 1980-1992 (0.25) | 1992-2001 (-1.48) | 1.73 | 0.50 | Pass |
| All Ages | #4 | 3 Jointpoint(s) | 1992-2001 (-1.48) | 2001-2015 (-1.79) | 0.31 | 0.50 | Fail |

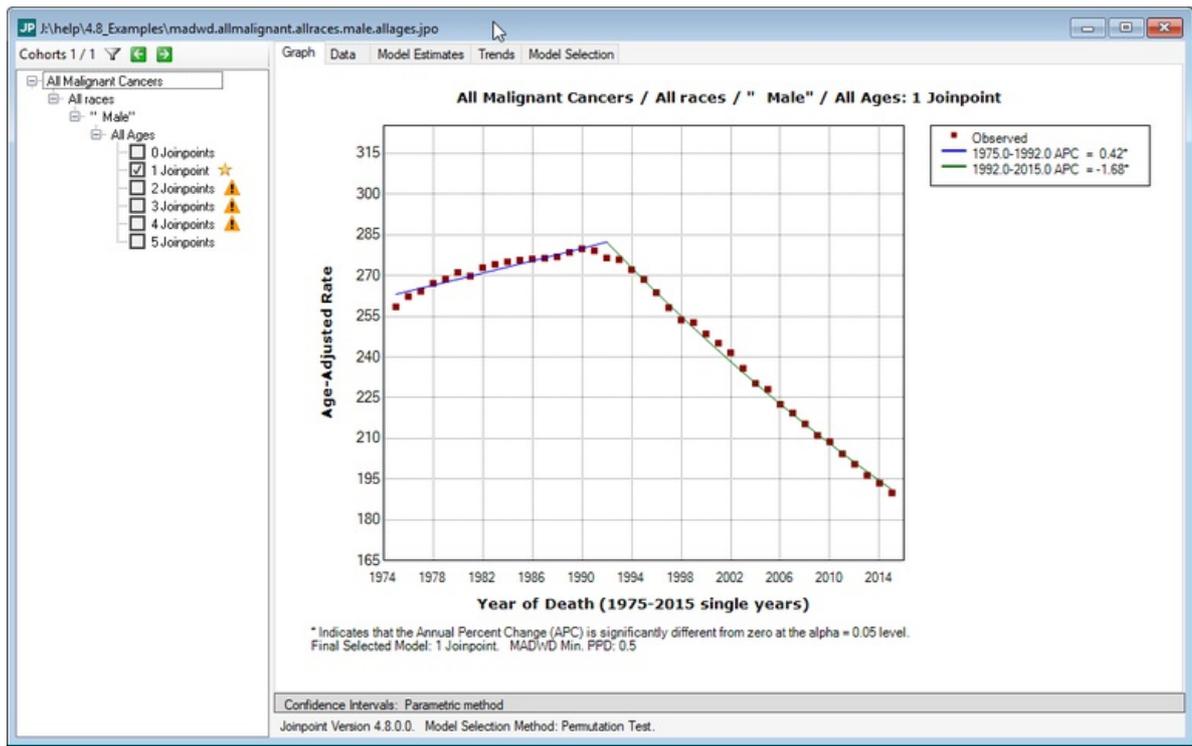
Below is the fit using 2 joinpoints.



The 2 Joinpoint model failed the MADWD test because the percentage point difference between the 1991-2001 and 2001-2015 segments is 0.40.

| Minimum APC Difference Worth Detecting (MADWD) | | | | | | | |
|--|-------|----------------------|-----------------------|-----------------------|-------------------------|------------------|--------------|
| Cohort | Model | Number of Joinpoints | Segment 1 Range (APC) | Segment 2 Range (APC) | Absolute APC Difference | Min. PPD Allowed | MADWD Result |
| All Ages | #3 | 2 Joinpoint(s) | 1975-1991 (0.44) | 1991-2001 (-1.40) | 1.85 | 0.50 | Pass |
| All Ages | #3 | 2 Joinpoint(s) | 1991-2001 (-1.40) | 2001-2015 (-1.80) | 0.40 | 0.50 | Fail |

Below is the fit using 1 joinpoint.



The one joinpoint model passes the MADWD test.

| Minimum APC Difference Worth Detecting (MADWD) | | | | | | | |
|--|-------|----------------------|-----------------------|-----------------------|-------------------------|------------------|--------------|
| Cohort | Model | Number of Joinpoints | Segment 1 Range (APC) | Segment 2 Range (APC) | Absolute APC Difference | Min. PPD Allowed | MADWD Result |
| All Ages | #2 | 1 Joinpoint(s) | 1975-1992 (0.42) | 1992-2015 (-1.68) | 2.10 | 0.50 | Pass |

Table

For any MADWD analysis, a table indicating the MADWD tests conducted will be displayed on the Model Selection tab. The summary table for this MADWD analysis is shown below:

| Minimum APC Difference Worth Detecting (MADWD) | | | | | | | |
|--|-------|----------------------|-----------------------|-----------------------|-------------------------|------------------|--------------|
| Cohort | Model | Number of Joinpoints | Segment 1 Range (APC) | Segment 2 Range (APC) | Absolute APC Difference | Min. PPD Allowed | MADWD Result |
| All Ages | #5 | 4 Joinpoint(s) | 1975-1979 (1.01) | 1979-1990 (0.34) | 0.68 | 0.50 | Pass |
| All Ages | #5 | 4 Joinpoint(s) | 1979-1990 (0.34) | 1990-1993 (-0.48) | 0.82 | 0.50 | Pass |
| All Ages | #5 | 4 Joinpoint(s) | 1990-1993 (-0.48) | 1993-2001 (-1.49) | 1.01 | 0.50 | Pass |
| All Ages | #5 | 4 Joinpoint(s) | 1993-2001 (-1.49) | 2001-2015 (-1.79) | 0.30 | 0.50 | Fail |
| All Ages | #4 | 3 Joinpoint(s) | 1975-1980 (0.93) | 1980-1992 (0.25) | 0.68 | 0.50 | Pass |
| All Ages | #4 | 3 Joinpoint(s) | 1980-1992 (0.25) | 1992-2001 (-1.48) | 1.73 | 0.50 | Pass |
| All Ages | #4 | 3 Joinpoint(s) | 1992-2001 (-1.48) | 2001-2015 (-1.79) | 0.31 | 0.50 | Fail |
| All Ages | #3 | 2 Joinpoint(s) | 1975-1991 (0.44) | 1991-2001 (-1.40) | 1.85 | 0.50 | Pass |
| All Ages | #3 | 2 Joinpoint(s) | 1991-2001 (-1.40) | 2001-2015 (-1.80) | 0.40 | 0.50 | Fail |
| All Ages | #2 | 1 Joinpoint(s) | 1975-1992 (0.42) | 1992-2015 (-1.68) | 2.10 | 0.50 | Pass |

It is important to note the Min PPD is a subjective criteria which overlays the statistical criteria that is applied first. It is also generally important that the analyst does not keep modifying the Min PPD criteria after viewing results until they obtain a result which meets some preconceived notion of what the trends should look like.

Related content

- [Method](#)
- [Annual Percent Change \(APC\) and Confidence Interval](#) A description of APC and Confidence Interval.

Number of Joinpoints

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

Min = 0, Max = 5, when the Grid Search 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: 2) of data points from the beginning or end of a series.
- There must be at least a user-specified number (default: 2) 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 two 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 - 26 | 4 |
| 27+ | 5 |

Note: The Grid Search Method allows a maximum of 9 joinpoints, but those runs could take quite a long time to complete.

Related content

- [Data Errors That Stop An Analysis](#)

Minimum Number of Observations Required

The minimum number of observations required to run a Joinpoint Analysis (using the Grid Search method) is determined by the following formula:

$$\text{Min. Num. Obs.} = (2 * \text{NumObsEnd}) + (\text{MaxJP} - 1) * \text{NumObsBetween} + \text{MaxJP}$$

NumObsEnd: The minimum number of observations from a Joinpoint to either end of the data, excluding the first or last Joinpoint if it falls on an observation.

NumObsBetween: The minimum number of observations between two joinpoints, excluding any Joinpoint if it falls on an observation.

MaxJP: The maximum number of joinpoints, determined by subtracting the user defined minimum number of joinpoints from the user defined maximum number of joinpoints.

If you receive an error message stating that there are too few observations to fit the specified model, check these settings on the Method and Parameters tab and adjust them to fit this formula.

Autocorrelated Errors Option

If you select "Fit an uncorrelated errors model", the program assumes the random errors in the regression model are uncorrelated and estimates the regression coefficients by ordinary least squares (unless the errors are heteroscedastic; see [Heteroscedastic Errors Option](#)).

If you select "Fit an autocorrelated errors model based on the data", the autocorrelation parameter will be estimated separately for each by-group using the method described in Section 2.3 of Kim et al. (2000). Under this option, the autocorrelation parameter is estimated for the model with the default maximum number of joinpoints or the maximum number of joinpoints set by a user.

If you select "Fit an autocorrelated errors model with parameter =", you must input an autocorrelation parameter (usually between -1 and 1) which represents the correlation between adjacent points. The program then assumes the random errors are autocorrelated and estimates the regression coefficients by weighted least squares. The autocorrelation model is based on the assumption that $\text{corr}(e_i, e_j) = \phi^{|i-j|}$ where e_i and e_j are the errors corresponding to the i^{th} and j^{th} points and ϕ is the autocorrelation parameter chosen. This option makes sense only with equally spaced points.

Although the autocorrelation may be estimated from the data, correcting for autocorrelation with this estimate may seriously reduce the power to detect joinpoints (see Section 3 of Kim et al. (2000)). We found in our simulations in Table IV of that paper that adjusting for autocorrelation was helpful in maintaining proper size of the tests of joinpoints when there was large autocorrelation. We also found that if there was no autocorrelation then the adjustment seriously affected the power of the test to detect joinpoints. For example we see in Table IV (b) with $\phi = 0$, the power goes from 90% to 68%. This is because it is difficult to differentiate between autocorrelation and joinpoints in a model.

If you suspect that your data are positively autocorrelated, we suggest using the "Fit an autocorrelated errors model with parameter =" option to see how sensitive your results are to changes in autocorrelation. The option should be used as follows:

1. Fit the model with the uncorrelated errors option.
2. If the user suspects that there is positive autocorrelation in the data, then repeat the analysis trying several values of the autocorrelation parameter, say for example 0.1, 0.2, and 0.3. If the results are very similar with different values of the autocorrelation parameter, then the user knows their results will still hold if there is autocorrelation present. If the results change as the autocorrelation parameter changes, then the user may end up presenting the series of results, to show how the results depend on different assumptions about the autocorrelation.

If you suspect negative autocorrelation, the uncorrelated errors model will suffice (see Kim et al., 2000).

Note: only the uncorrelated errors model can be used with the Pairwise Comparison.

How Joinpoint Selects the Final Model

Selecting the model, that is, the number of joinpoints, can be done by using one of the following options in Joinpoint:

- Permutation Test

Traditional BIC Methods

- Bayesian Information Criterion (BIC)
- BIC3
- Modified BIC

Data Driven BIC Methods

- Weighted BIC (WBIC)
- Weighted BIC Alternative (WBIC-Alt)
- Data Dependent Selection (DDS)

The Permutation test, that was proposed in Kim et al. (2000) and implemented in the original version of Joinpoint, is computationally intensive. In addition to the Permutation test method, there are six methods based on the information criteria. These methods are much more computationally efficient than the Permutation test. These information based criteria methods can be grouped into two categories, Traditional Bayesian Information Criteria (BIC) methods and Data Driven BIC methods. The Traditional BIC methods select the model for which the object function, which is either the sum of the model fit error and the penalty term (BIC and BIC3) or an asymptotic approximation of the Bayes factor (MBIC), is minimized. The Data Driven BIC methods include the DDS method introduced in Joinpoint v4.6.0.0 and two new methods, Weighted BIC and Weighted BIC-Alt introduced in v4.7.0.0.

The motivation of the Data driven BIC methods is to internally determine the model selection method, BIC or BIC3, based on the characteristics of data. The basic idea is to use BIC if change sizes are relatively small, and to use BIC3 otherwise. Through simulations, the permutation test has been shown to produce reasonable results, (i.e. predicts the correct number of joinpoints), when the effect size, a function of the size of slope changes adjusted for the variability in data, is large. The BIC3 method has been shown to produce results similar to the permutation test in situations when the permutation test is performing well relative to the traditional BIC method. For situations with smaller effect sizes, the BIC performs better. However, unlike in simulations, the analyst does not know the true effect sizes in practice, and the data driven methods were developed with an aim to use the data to approximate the effect sizes. The data driven methods either choose between BIC and BIC3 (Data Driven Selection-DDS) or use a weighted average of the BIC and BIC3 (Weighted BIC or Weighted BIC-Alt). Simulations have shown that among the Data Driven BIC methods, the DDS and Weighted BIC methods perform similarly, and the Weighted BIC is preferred because it has a stronger conceptual justification. The Weighted BIC-Alt performs worst of the Data Driven BIC Methods.

Therefore our overall recommendations are:

- Use the permutation test if the user prefers the method that has the longest track record and generally produces parsimonious results (i.e. will detect fewer joinpoints than other approaches, especially if the slope changes are small).
- Use the BIC3 method if the user would like to produce results similar to the permutation test, but computation time is an issue.

- Use the Weighted BIC method if the user prefers a method that on average performs best across a wide range of situations. While the permutation test, BIC, and BIC3 might perform better in specific situations, the Weighted BIC is the most flexible in adapting to different situations. The Weighted BIC is being considered as the default for use in the Joinpoint software after we gain more experience with it. The DDS method is a worthy competitor to the Weighted BIC, and the Weighted BIC was selected as preferred based on conceptual reasons, rather than on the results of our simulations.
- The Weighted BIC-Alt and Modified BIC are generally not recommended for most users, and are included for very specialized purposes.

Details of each method are given below:

Permutation Test

This method uses the sequence of permutation tests to ensure that the approximate probability of overall Type I error is less than the specified significance level (also called the alpha level, default = .05). Assuming that the default value of the minimum number of joinpoints is 0, "the overall Type I error" is the probability of incorrectly concluding that the underlying model has one or more joinpoints when, in fact, the true underlying model has no joinpoints.

Bayesian Information Criteria (BIC)

The value of BIC is the loglikelihood value with penalizing the cost of extra parameters. The model with the minimum value of BIC is selected as the optimal model.

BIC3

A modification of traditional BIC with a harsher penalty.

Modified BIC

A modification of traditional BIC proposed to improve its performance.

Weighted BIC (WBIC) [Beta Version]

While the Data Dependent Selection (DDS) internally uses BIC or BIC3 based on the empirically determined cut-off values for the selection statistics, the weighted BIC combines BIC and BIC3 using a weighted penalty term based on the data characteristic.

Weighted BIC Alternative (WBIC-Alt) [Beta Version]

A modification of BIC3 that is less conservative than WBIC.

Data Dependent Selection (DDS) [Beta Version]

This procedure internally determines the model selection method, BIC or BIC3, based on the characteristics of data, and its basic idea is to use BIC if change sizes are relatively small and BIC3 otherwise.

Related content

- Permutation Test
- Bayesian Information Criterion (BIC)
- BIC3
- Modified BIC
- Weighted BIC (WBIC)
- Weighted BIC Alternative (WBIC-Alt)
- Data Dependent Selection (DDS)
- Alpha and Beta Versions Features in Joinpoint which are labeled Alpha or Beta versions

Permutation Test

The program performs multiple tests to select the number of joinpoints, using the Bonferroni correction for multiple testing. Set the overall significance level for multiple testing.

The program performs permutation tests to select the number of joinpoints. Since fitting all $N!$ possible permutations of the data would take too long, the program takes a Monte Carlo sample of these $N!$ data sets, using a random number generator. Specify the size of the Monte Carlo sample of permuted data sets.

Related content

- [Early Stopping Options](#)
- [Permutation Test Details](#)
- [Joinpoint p-values](#)
- [How Joinpoint Conducts Permutation Testing](#)
- [Model Selection Method](#)

How Joinpoint Conducts Permutation Testing

In the Joinpoint Regression Program, the permutation test is used repeatedly for testing between two different joinpoint models, a simpler model with fewer joinpoints called the null model, and a more complicated model called the alternative model. The alternative model fits better because it is more complicated.

The question for the test is: does it fit much better than would be expected by chance. To test this statistically, we calculate a ratio, SSE_N/SSE_A , where SSE_N is the sum of squared errors (SSE) from the null model and SSE_A is the SSE from the alternative model. Values of the ratio close to 1 mean that the alternative is not much better than the null model, while larger values mean that the alternative is much better.

In order to decide how much larger a ratio needs to be to be statistically significant, we use the permutation method. In this method we randomly permute (that is, shuffle) the errors (also called the residuals) from the null model and add them back onto the modeled values from the null model to create a permutation data set. Then we calculate the ratio for the permutation data set.

- If the true model was the null model, we would expect that about half of the ratios calculated from the permutation data set would be less than the original one.
- If the true model were the alternative model, we would expect that after permuting the errors most of the new ratios would be less than the original ratio. In other words the permuted data set would look less like the alternative model than the original data.

So we reject the null model (or null hypothesis) if less than a certain proportion of the ratios are greater than or equal to the original ratio.

Permutation Test Details

First, the user specifies MIN as the minimum number of joinpoints and MAX as the maximum number of joinpoints on the Method and Parameters tab. Then the program uses a sequence of permutation tests to select the final model. Each one of the permutation tests performs a test of the null hypothesis H_0 : number of joinpoints= k_a against the alternative H_a : number of joinpoints= k_b where $k_a < k_b$. The procedure begins with $k_a = \text{MIN}$ and $k_b = \text{MAX}$. If the null is rejected, then increase k_a by 1; otherwise, decrease k_b by 1. The procedure continues until $k_a = k_b$ and the final value of $\hat{k} = k_a = k_b$ is the selected number of joinpoints.

Significance level of each individual test in a sequential testing procedure

Because multiple tests are performed, the significance level of each test is adjusted to control the overall type I error at specified α level (0.05). Before Version 3.0, Bonferroni adjustment was used, i.e., $\alpha_1 = \alpha / (\text{MAX} - \text{MIN})$ and if the individual test p-value is less than α_1 , the null is rejected.

The Bonferroni adjustment is conservative because the actual overall significance level is usually less than the nominal level α . Starting with Version 3.0, the new adjustment procedure controls the overall over-fitting error probabilities,

$$P(\hat{k} > K_a), K_a = \text{MIN}, \dots, \text{MAX} - 1,$$

under α . Let k denote the number of joinpoints and $\alpha(k_a; k_b)$ be the significance level of each individual test $H_0 : k = k_a$ vs. $H_1 : k = k_b$. The new procedure set $\alpha(k_a; k_b) = \alpha / (\text{MAX} - k_a)$. Notice that the individual significance level depends on the number of joinpoints k_a under the null. Consider an example where $\text{MIN} = 0$ and $\text{MAX} = 4$. The new procedure has the following properties:

$$P(\hat{k} > 0 | k = 0) \leq \alpha(0, 4) + \alpha(0, 3) + \alpha(0, 2) + \alpha(0, 1);$$

$$P(\hat{k} > 1 | k = 1) \leq \alpha(1, 4) + \alpha(1, 3) + \alpha(1, 2);$$

$$P(\hat{k} > 2 | k = 2) \leq \alpha(2, 4) + \alpha(2, 3);$$

$$P(\hat{k} > 3 | k = 3) \leq \alpha(3, 4).$$

If we like to bound these over-fitting probabilities by α , then we can assign different values for each $\alpha(k_a; k_b)$. That means, we can achieve a better power by setting

$$\alpha(0, 4) = \alpha(0, 3) = \alpha(0, 2) = \alpha(0, 1) = \alpha/4;$$

$$\alpha(1, 4) = \alpha(1, 3) = \alpha(1, 2) = \alpha/3;$$

$$\alpha(2, 4) = \alpha(2, 3) = \alpha/2;$$

$$\alpha(3, 4) = \alpha.$$

Overall Significance Level

Set the significance level to be used for the permutation test.

Number Of Randomly Permuted Data Sets

The minimum allowed number of permutations is 1000. The maximum number is 10,000. For

greater consistency in the p-values obtained if one were to change the seed for each run, we strongly recommend running the program for at least 4499 permutations. You may use a smaller number of permutations to speed up the calculations, but the permutation test may produce less consistent results. Unless you have a good understanding of the implications of changing the number of permutations, we recommend against it.

Joinpoint p-values

The Joinpoint Regression Program performs a series of hypothesis tests that test the null hypothesis of k_a joinpoints against the alternative hypothesis of k_b joinpoints, where k_a and k_b change for each hypothesis test. Each p -value corresponds to this type of test. The p -value is an estimate of the probability, under the assumption that there are only k_a joinpoints, of observing data that look more like a joinpoint model with greater than k_a joinpoints than the data that we have in fact observed.

As the permutation test is a randomization test, it depends on the random number generator. For greater consistency in the permutation test p -values obtained if one were to change the seed for each run, we suggest running the program for at least 4499 permutations. For this reason, the default number of permutations is now 4499 in the current version of the Joinpoint Regression Program. Choice of the number of permutations selected by the user is a trade off between computer time and consistency of the p -values obtained.

Related content

- [Random Number Generation](#)

Random Number Generation

The program performs permutation tests to select the number of joinpoints. Since fitting all $N!$ possible permutations of the data would take too long, the program takes a Monte Carlo sample of these $N!$ data sets, using a random number generator to calculate p -values for a series of permutation tests.

Here we discuss the implications of the choice of the number of permutation data sets, say N . The program runs faster with smaller values of N , but it gives better precision for the p -value with larger values of N . In addition, a larger N reduces the probability that another analysis of the same data might get a different answer when run with different random number generator seeds.

Computer programs can produce pseudo-random numbers through algorithms that mimic randomness, which we use to shuffle or permute the errors. The algorithms use a seed or seeds to start the algorithm. These seeds can be used to produce repeatable pseudo-random numbers. The problem of two analyses obtaining different answers from the same data is addressed by this program by specifying default random number generator seeds. Thus, as long as no parameters are changed (including the random number generator seed and N), repeats of the analyses will produce the same results. Otherwise, two runs of the same analysis using different seeds could get different answers.

To get an idea how results would change for someone using different random number generator seeds, we list some confidence intervals for p -values below. For example, with $N=999$ Monte Carlo samples if you obtained a p -value of .04 from the program there is an approximately 99% chance that another researcher repeating the analysis with N very large (i.e., an ideal situation with $N \rightarrow \text{infinity}$) would obtain a p -value between .025 and .0577.

N=99
lower 99% ci estimate upper 99% ci

| | | |
|--------|------|--------|
| 0.0000 | 0.01 | 0.0521 |
| 0.0034 | 0.04 | 0.1065 |
| 0.0069 | 0.05 | 0.1218 |
| 0.0111 | 0.06 | 0.1364 |
| 0.0325 | 0.10 | 0.1910 |
| 0.2702 | 0.40 | 0.5281 |

N=999
lower 99% ci estimate upper 99% ci

| | | |
|--------|------|--------|
| 0.0031 | 0.01 | 0.0199 |
| 0.0250 | 0.04 | 0.0577 |
| 0.0331 | 0.05 | 0.0694 |
| 0.0415 | 0.06 | 0.0810 |
| 0.0762 | 0.10 | 0.1259 |

| | | |
|--------|------|--------|
| 0.3595 | 0.40 | 0.4402 |
|--------|------|--------|

N=9999

lower 99% ci estimate upper 99% ci

| | | |
|--------|------|--------|
| 0.0075 | 0.01 | 0.0127 |
| 0.0350 | 0.04 | 0.0452 |
| 0.0445 | 0.05 | 0.0558 |
| 0.0540 | 0.06 | 0.0663 |
| 0.0923 | 0.10 | 0.1079 |
| 0.3873 | 0.40 | 0.4127 |

N=99999

lower 99% ci estimate upper 99% ci

| | | |
|--------|------|---------|
| 0.0092 | 0.01 | 0.0108 |
| 0.0384 | 0.04 | 0.04116 |
| 0.0482 | 0.05 | 0.0518 |
| 0.0581 | 0.06 | 0.0620 |
| 0.0978 | 0.10 | 0.1025 |
| 0.3960 | 0.40 | 0.4040 |

Typically, you should allow the computer to use the default seeds for the random number generator. By using these default seeds, one can duplicate results from a previous Joinpoint session even though the program uses Monte Carlo sampling. Although not recommended in general, the Joinpoint Regression Program allows one to change the default seed by selecting **Session > Preferences** from the **File** menu.

Early Stopping Options

Since Hudson's Method is computationally intensive, Early Stopping Options have been added so that not all permutations need to be analyzed. The options are:

- Fixed - all permuted data sets are analyzed, with the default being 4499.
- B-Value - the maximum number of permuted data sets to be analyzed is determined by the significance level specified, using a less conservative approach. See Fay et al. for details (Fay, MP, Kim, H-J, and Hachey, M. (2007) "On using Truncated Sequential Probability Ratio Test Boundaries for Monte Carlo Implementation of Hypothesis Tests" (*Journal of Computational and Graphical Statistics* 16, 946-967).
- Curtailed - the maximum number of permuted data sets to be analyzed is determined by the significance level specified, using a more conservative approach.

This option is not currently available for the Grid Search Method but may be added in a future version of the software.

Related content

- [References](#)

Bayesian Information Criterion (BIC)

The equation for computing the BIC for a k-joinpoint model is:

$$\text{BIC}(k) = \ln\{\text{SSE}(k)/\#\text{Obs}\} + \{\#\text{Parm}(k) / \#\text{Obs}\} * \ln(\#\text{Obs}),$$

where SSE is the sum of squared errors of the k-joinpoint regression model, $\#\text{Parm}(k)=2*(k+1)$ is the number of parameters of the k-joinpoint model and $\#\text{Obs}$ is the number of observations.

The k-joinpoint model with the minimum value of BIC(k) is selected as the final model.

Related content

- [Data Errors That Stop An Analysis](#)
- [Model Selection Method](#)

BIC3

BIC3 uses a harsher penalty than the traditional Bayesian Information Criterion (BIC). The equation for computing the BIC3 for a k-joinpoint model is:

$$\text{BIC3}(k) = \ln\{\text{SSE}(k) / \#\text{Obs}\} + \{\text{PenaltyCoefficient}(k) / \#\text{Obs}\} * \ln(\#\text{Obs}),$$

where SSE is the sum of squared errors of the k-joinpoint regression model, PenaltyCoefficient(k) = 3k + 2 is the penalty coefficient of the k-joinpoint model, and #Obs is the number of observations. The k-joinpoint model with the minimum value of BIC3(k) is selected as the final model.

Related content

- [Model Selection Method](#)
- [Bayesian Information Criterion \(BIC\)](#)

Modified BIC

Zhang and Siegmund (2007, Biometrics) discussed that in the context of change-point problems, the traditional BIC does not satisfy the technical assumptions of Schwarz (1978, Annals of Statistics) and proposed a modification to improve its performance. The MBIC in Joinpoint regression is derived as an asymptotic approximation of the Bayes factor and is of the form:

$$MBIC(k) = BIC(k) + \frac{\ln|X'_k(\hat{\tau})X_k(\hat{\tau})|}{n} - \frac{2}{n}\ln\Gamma\left(\frac{n-k-3}{2}\right) - \frac{k+3}{n}\ln(SSE(k)),$$

where n is the number of observations, $\Gamma(z)$ is the gamma function:

$$\Gamma(z) = \int_0^{\infty} t^{z-1}e^{-t}dt$$

and

$$X_k(\hat{\tau}) = \begin{pmatrix} 1 & x_1 & (x_1 - \hat{\tau}_1)^+ & \dots & (x_1 - \hat{\tau}_k)^+ \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & (x_n - \hat{\tau}_1)^+ & \dots & (x_n - \hat{\tau}_k)^+ \end{pmatrix}$$

for a k -joinpoint model with the values of the independent variable (x_1, \dots, x_n) and the joinpoints estimated as $(\hat{\tau}_1, \dots, \hat{\tau}_k)$. Note that $a^+ = \max(a, 0)$.

Related content

- [Model Selection Method](#)

Data Dependent Selection (DDS)

NOTES:

- **The Data Dependent Selection (DDS) Method is available in Joinpoint as of version 4.6.0.0, but is labeled as a "beta" feature and considered to be still in development.**

Data Dependent Choice of Model Selection Methods

Selecting the model, that is, the number of joinpoints, can be done by using one of the following options in Joinpoint: Permutation test, Bayesian Information Criterion (BIC), BIC3, and Modified BIC. The Permutation test was proposed in Kim et al. (2000) and implemented in the original version of Joinpoint. The use of BIC in Joinpoint is described in Kim et al. (2009), and the use of BIC3, which applies a harsher penalty than the traditional BIC, is introduced in Kim and Kim (2016). The modified BIC, originally proposed by Zhang and Siegmund (2007) and applied to the joinpoint regression model, was implemented in Joinpoint V 3.5.

The Permutation test is computationally intensive, and the information based criteria, BIC, BIC3 and Modified BIC, are much more computationally efficient than the Permutation test. Regarding their performances, simulation studies indicated that (i) BIC performs well to detect a change with a small effect size and has a tendency of over-estimating the number of joinpoints, (ii) Modified BIC is the most conservative among these selection methods and performs well to detect a change with a large effect size, (iii) the performance of BIC3 is comparable to that of the Permutation test.

A new feature that combines BIC and BIC3 to improve the performance of BIC3 and is computationally more efficient than the Permutation test is implemented in Joinpoint version 4.6.0.0-Alpha. The new procedure internally determines the model selection method, BIC or BIC3, based on the characteristics of data, and its basic idea is to use BIC if change sizes are relatively small and BIC3 otherwise.

Consider a joinpoint regression model,

$y = \beta_0 + \beta_1 x + \delta_1(x - \tau_1)^+ + \dots + \delta_k(x - \tau_k)^+ + \epsilon$, where K is an unknown number of joinpoints, and $a^+ = a$ if $a > 0$, and 0 otherwise. Suppose that with a pre-specified k_{max} , a model with k joinpoints is selected by BIC or BIC3 ($0 \leq k \leq k_{max}$), for which the parameters are estimated as $\hat{\tau}_1, \dots, \hat{\tau}_k, \hat{\beta}_0, \hat{\beta}_1, \hat{\delta}_1, \dots, \hat{\delta}_k$. For the observations in the i th and $(i+1)$ st segments estimated ($i=1, 2, \dots, k$), that is, the observations whose x -values are in $(\hat{\tau}_{i-1}, \hat{\tau}_{i+1}]$, where $\hat{\tau}_0 = \min x_i - 1$ and $\hat{\tau}_{k+1} = \max x_i$, call their x -values in ascending order as $x_{j_1+1}, \dots, x_{j_2}$, and let

$$Z_i = ((x_{j_1+1} - \hat{\tau}_i)^+, \dots, (x_{j_2} - \hat{\tau}_i)^+)^T \text{ and}$$

$$X_0 = \begin{pmatrix} 1 & x_{j_1+1} \\ \vdots & \vdots \\ 1 & x_{j_2} \end{pmatrix}. \text{ Also let}$$

$\Delta_{i,i+1} = \hat{\sigma}_i^2 Z_i^T (I - H_0) Z_i / \hat{\sigma}^2$, where $H_0 = X_0 (X_0^T X_0)^{-1} X_0^T$ and $\hat{\sigma}^2$ is the mean squared error of the model with the maximum number of joinpoints, k_{max} , and define

$$\Delta(k) = \min_{i=1, \dots, k} \Delta_{i,i+1}.$$

Note that the measure $\Delta_{i,i+1}$ is motivated from the consideration of a quantity related to the power of a test to detect a slope change of δ from a simple linear regression model.

Given two pre-specified values, c and d , as cutoff values, the number of joinpoints is estimated

as \hat{K} according to the following steps:

Step 1: Estimate the number of joinpoints using both BIC and BIC3 and call them \hat{K}_{BIC} and \hat{K}_{BIC3} , respectively.

Step 2: If $\hat{K}_{BIC} = \hat{K}_{BIC3}$, then report it as \hat{K} .

Step 3: If $\hat{K}_{BIC} \neq \hat{K}_{BIC3}$, compute $\Delta_{max} = \max(\Delta(\hat{K}_{BIC}), \Delta(\hat{K}_{BIC3}))$ and $\Delta_{min} = \min(\Delta(\hat{K}_{BIC}), \Delta(\hat{K}_{BIC3}))$.

Step 4: Use BIC if $\Delta_{max} \leq c$ or $\Delta_{diff} = \Delta_{max} - \Delta_{min} > d$ (that is, $\hat{K} = \hat{K}_{BIC}$), and use BIC3 otherwise (that is, $\hat{K} = \hat{K}_{BIC3}$).

Based on a simulation study where the performance of the new model selection procedure with various choices of c and d was examined, we recommend to use c=10 and d=200. Among the values of c and d considered, the selection procedure with these choices of c and d was observed to perform best for the goal of it being at least as good as BIC3 and improving BIC3 when BIC performs better than BIC3. Further details can be found in a technical report that is available upon request.

References

- [Consistent Model Selection in Segmented Line Regression](#)
- [A Modified Bayes Information Criterion with Applications to the Analysis of Comparative Genomic Hybridization Data](#)

Related content

- [Model Selection Method](#)
- [Alpha and Beta Versions](#) Features in Joinpoint which are labeled Alpha or Beta versions

Weighted BIC (WBIC)

NOTES:

- **The Weighted BIC Selection Method is available in Joinpoint as of version 4.7.0.0, but is labeled as a "beta" feature and considered to be still in development.**

While the Data Dependent Selection (DDS) internally uses BIC or BIC3 based on the empirically determined cut-off values for the selection statistics, the weighted BIC combines BIC and BIC3 using a weighted penalty term based on the data characteristic. That is, it assigns a harsher penalty, making the selection rule close to BIC3, when change sizes are relatively large and a less penalty, making the selection rule close to BIC, otherwise.

In addition to the notations used for the description of the DDS, let $y_i = (y_{j_1+1}, \dots, y_{j_2})^T$ and consider the partial R^2 as

$$R_{i,i+1}^2 = \frac{\{y_i^T(I - H_0)z_i\}^2}{\{y_i^T(I - H_0)y_i\}\{z_i^T(I - H_0)z_i\}}.$$

Then, define the weighted BIC as

$$WBIC(k) = \ln\{SSE(k)/\#Obs\} + (2 + R_{max}^2)(k/\#Obs) * \ln(\#Obs),$$

where $R_{max}^2 = \max_{j=1, \dots, k} R_{i,i+1}^2$.

Related content

- [Bayesian Information Criterion \(BIC\) Details](#) What is the Bayesian Information Criterion method for selecting the best model?
- [BIC3](#)
- [Data Dependent Selection \(DDS\)](#)
- [Model Selection Method](#)
- [Alpha and Beta Versions](#) Features in Joinpoint which are labeled Alpha or Beta versions

Weighted BIC Alternative (WBIC-Alt)

NOTES:

- **The Weighted BIC Alternative (WBIC-Alt) Selection Method is available in Joinpoint as of version 4.7.0.0, but is labeled as a "beta" feature and considered to be still in development.**

A modification of BIC3 that is less conservative than WBIC. WBIC-Alt selects the model that minimizes

$$WBIC - Alt(k) = \ln\{SSE(k)/\#Obs\} + (2 + R_{min}^2)(k / \#Obs) * \ln(\#Obs),$$

where $R_{min}^2 = \min_{i=1, \dots, k} R_{i,i+1}^2$.

Related content

- [Model Selection Method](#)
- [BIC3](#)
- [Weighted BIC \(WBIC\)](#)
- [Alpha and Beta Versions](#) Features in Joinpoint which are labeled Alpha or Beta versions

APC/AAPC/Tau Confidence Intervals

A description of the methods used for calculating APC, AAPC, and Tau Confidence Intervals.

Relevant FAQ's:

- [What is an APC?](#)
- [What is an AAPC?](#)
- [What is Tau?](#)

Parametric

In the Joinpoint software, the parametric confidence interval for the true AAPC is based on the normal distribution, and the APC confidence interval is based on a t distribution. If an AAPC lies entirely within a single joinpoint segment, the AAPC is equal to the APC for that segment. To obtain consistency between the APC and AAPC confidence intervals in this situation, the confidence interval for the AAPC has been modified to be identical to that used for the APC using the t distribution instead of the normal distribution.

- For details regarding the confidence intervals for the APC, see [Annual Percent Change \(APC\) and Confidence Interval](#).
- For details regarding the confidence intervals for the AAPC, see [Average Annual Percent Change \(AAPC\) and Confidence Interval](#).
- For details regarding the confidence intervals for the Tau's (Joinpoint Locations), see [Estimated Joinpoint \(Estimated Tau\) and Confidence Interval](#).

Empirical Quantile

Motivated by a conservative tendency of the asymptotic confidence interval for the true AAPC, a new method called the empirical quantile method was introduced as an improved confidence interval. The idea of the empirical quantile method is to generate resampled data by (i) generating resampled residuals as the inverse function values of the uniform random numbers over (0,1) where the function is the empirical distribution function of the original residuals and then (ii) adding resampled residuals to the original fit. For each resampled data, the model fit was made and the AAPCs are estimated. Then, the $100(\alpha/2)$ th and $100(1-\alpha/2)$ th percentiles of the resampled AAPC values are obtained as the lower and upper limits of the $100(1-\alpha)\%$ empirical quantile confidence interval for the true AAPC. For details regarding this method please see the [Improved Confidence Interval for Average Annual Percent Change in Trend Analysis](#) article.

For a more detailed description of the Empirical Quantile methodology, please go [here](#).

This method is also available for the APC and Tau Confidence Intervals.

Alpha Levels: To learn how to adjust the alpha levels for any of the confidence intervals, please go to the [Preferences](#) help section .

Annual Percent Change (APC) and Confidence Interval

A description of APC and Confidence Interval.

Annual Percent Change (APC) is one way to characterize trends in cancer rates over time. When the Log Transformation option on the Input File tab is $\ln(y)=xb$, then the output calculates the estimated annual percentage rate change (APC). With this approach, the cancer rates are assumed to change at a constant percentage of the rate of the previous year. For example, if the APC is 1%, and the rate is 50 per 100,000 in 1990, the rate is $50 \times 1.01 = 50.5$ in 1991 and $50.5 \times 1.01 = 51.005$ in 1992. Rates that change at a constant percentage every year change linearly on a log scale. For this reason, to estimate the APC for a series of data, the following regression model is used:

$\log(R_y) = b_0 + b_1y$ where $\log(R_y)$ is the natural log of the rate in year y .

$$\begin{aligned} \text{The APC from year } y \text{ to year } y + 1 &= \left[\frac{R_{y+1} - R_y}{R_y} \right] \times 100 \\ &= \frac{\{e^{b_0 + b_1(y+1)} - e^{b_0 + b_1(y)}\}}{e^{b_0 + b_1(y)}} \times 100 \\ &= (e^{b_1} - 1) \times 100 \end{aligned}$$

One advantage of characterizing trends this way is that it is a measure that is comparable across scales, for both rare and common cancers. For example, it is reasonable to think that rates for a rare cancer and a common cancer could both change at 1% per year, but it is not reasonable to think that a rare cancer and a common cancer would change in the same increments on an absolute (or arithmetic) scale. That is, a cancer with a rate of 100 per 100,000 could be changing by 2 per 100,000 every year, but a cancer with a rate of 1 per 100,000 would probably not change in the same increments.

It is not always reasonable to expect that a single APC can accurately characterize the trend over an entire series of data. The joinpoint model uses statistical criteria to determine when and how often the APC changes. For cancer rates, it is fit using joined log-linear segments, so each segment can be characterized using an APC. For example, cancer rates may rise gradually for a period of several years, rise sharply for several years after that, and then drop gradually for the next several years. Finding the joinpoint model that best fits the data allows us to determine how long the APC remained constant, and when it changed.

Confidence Intervals - Parametric Method

For any segment with slope β , the APC is $100\{ \exp(\beta) - 1 \}$. The $100(1-\alpha)\%$ **confidence limits** are:

$$\text{Lower} = 100\{ \exp(\beta - s * t_d^{-1}(1 - \alpha/2)) - 1 \} \text{ Upper} = 100\{ \exp(\beta + s * t_d^{-1}(1 - \alpha/2)) - 1 \}$$

where d is the degrees of freedom and s is the standard error for the slope listed in the output (i.e., from the unconstrained linear models), and $t_d^{-1}(q)$ is the q th quantile of a t distribution with d degrees of freedom.

The p -value for a two-sided test that the true APC is zero is calculated based on a t distribution.

Confidence Intervals - Empirical Quantile Method

For a more detailed description of the Empirical Quantile methodology, please go [here](#).

Related content

- [Estimated Regression Coefficients \(Beta\)](#)

Average Annual Percent Change (AAPC) and Confidence Interval

While Joinpoint computes the trend in segments whose start and end are determined to best fit the data, sometimes it is useful to summarize the trend over a fixed predetermined interval. The AAPC is a method which uses the underlying Joinpoint model to compute a summary measure over a fixed pre-specified interval.

Average Annual Percent Change (AAPC) is a summary measure of the trend over a pre-specified fixed interval. It allows us to use a single number to describe the average APCs over a period of multiple years. It is valid even if the joinpoint model indicates that there were changes in trends during those years. It is computed as a weighted average of the APCs from the joinpoint model, with the weights equal to the length of the APC interval.

What is an AAPC?

How is the AAPC Computed?

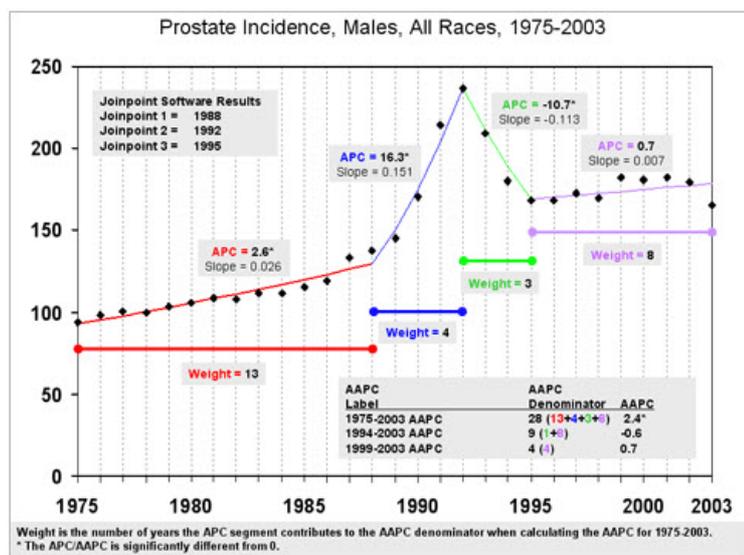
How are the AAPC Confidence Intervals Computed?

Can We Compare AAPCs for Two Independent Groups?

What are the Relative Advantages and Disadvantages of Reporting an AAPC over APCs?

What is the Advantage of Reporting an AAPC over an APC Computed by Fitting a Single Line (on a log Scale) to the Data?

How is the AAPC Computed?



AAPC is derived by first estimating the underlying joinpoint model that best fits the data. The accompanying figure shows the joinpoint model for prostate cancer incidence from 1975-2003 (from the 2005 data submission), which found joinpoints in 1988, 1992, and 1995. (This model is fit under the default joinpoint parameters). The AAPC over any fixed interval is calculated using a weighted average of the slope coefficients of the underlying joinpoint regression line with the weights equal to the length of each segment over the interval. The final step of the calculation transforms the weighted average of slope coefficients to an annual percent change. If we denote b_j s as the slope coefficients for each segment in the desired range of years, and

the w_i s as the length of each segment in the range of years, then:

$$APC_i = \{ \text{Exp}(b_i) - 1 \} \times 100$$

and

$$AAPC = \left\{ \exp\left(\frac{\sum w_i b_i}{\sum w_i}\right) - 1 \right\} \times 100$$

In the prostate cancer example, to compute the AAPC from 1994 to 2003, we first note that an APC of -10.7 runs for 1 year (with a slope coefficient of -0.113), while an APC of 0.7 runs for 8 years (with a slope coefficient of 0.007). Thus, the AAPC is computed as:

$$\{ \text{Exp}((1 \times (-0.113) + 8 \times 0.007) / 9) - 1 \} \times 100 = -0.6.$$

The AAPC for a year range which is entirely within a single joinpoint segment is equal to the APC for that segment, e.g., the AAPC for 1999 to 2003 is computed as:

$$\{ \text{Exp}((4 \times 0.007) / 4) - 1 \} \times 100 = 0.7.$$

The AAPC for the entire range of data is computed as:

$$\{ \text{Exp}((13 \times 0.026 + 4 \times 0.151 + 3 \times (-0.113) + 8 \times 0.007) / 28) - 1 \} \times 100 = 2.4.$$

How are the AAPC Confidence Intervals Computed?

Parametric Confidence Interval for True AAPC

Denote the normalized weight as

$$\tilde{w}_i = w_i / \sum w_j$$

and rewrite the AAPC as

$$AAPC = \left\{ \exp\left(\sum \tilde{w}_i b_i\right) - 1 \right\} \times 100$$

An approximate $100(1 - \alpha)\%$ confidence interval for the true *average annual percent change* is $(AAPC_{L(\alpha)}, AAPC_{U(\alpha)})$, where

$$AAPC_{L(\alpha)} = \left\{ \exp\left[\log((AAPC/100) + 1) - z_{1-\alpha/2} \sqrt{\sum \tilde{w}_i^2 \hat{\sigma}_i^2}\right] - 1 \right\} \times 100$$

$$AAPC_{U(\alpha)} = \left\{ \exp\left[\log((AAPC/100) + 1) + z_{1-\alpha/2} \sqrt{\sum \tilde{w}_i^2 \hat{\sigma}_i^2}\right] - 1 \right\} \times 100$$

are the lower and upper confidence limits of the interval, Z_α is the α^{th} quantile of the standard normal distribution, and $\hat{\sigma}_i^2$ denotes the estimate variance of b_i obtained from the fit of the joinpoint model.

In the Joinpoint software, the AAPC confidence interval is based on the normal distribution, and the APC confidence interval is based on a t distribution. If an AAPC lies entirely within a single joinpoint segment, the AAPC is equal to the APC for that segment. To obtain consistency between the APC and AAPC confidence intervals in this situation, the confidence interval for the AAPC has been modified to be identical to that used for the APC using the t distribution instead of the normal distribution. For details on the confidence intervals for the APC, see: [Annual Percent Change \(APC\) and Confidence Interval](#).

Empirical Quantile Confidence Interval for True AAPC

Motivated by a conservative tendency of the asymptotic confidence interval for the true AAPC, a new method called the empirical quantile method is implemented in V 4.2. as an improved confidence interval. The idea of the empirical quantile method is to generate resampled data by (i) generating resampled residuals as the inverse function values of the uniform random

numbers over (0,1) where the function is the empirical distribution function of the original residuals and then (ii) adding resampled residuals to the original fit. For each resampled data, the model fit was made and the AAPCs are estimated. Then, the 100(α/2)th and 100(1-α/2)th percentiles of the resampled AAPC values are obtained as the lower and upper limits of the 100(1-α)% empirical quantile confidence interval for the true AAPC. For more information on the empirical quantile confidence interval, please go [here](#).

Confidence Interval and Hypothesis Test

If the confidence interval contains zero, then there is no evidence to reject the null hypothesis that the true AAPC is zero at the significance level of α otherwise, we reject the null hypothesis in favor of the alternative hypothesis that the true AAPC is different from zero.

[Return to the top]

Can we compare AAPCs for two independent groups?

The true AAPCs for two independent groups, for example, males and females, can be compared by using the following approximate 100(1-α)% confidence interval with the estimated difference of $AAPC^{(1)} - AAPC^{(2)}$, where

$$AAPC^{(1)} = \{ \exp(\sum \tilde{w}_i^{(1)} b_i^{(1)}) - 1 \} \times 100 = \{ \exp(\hat{\phi}^{(1)}) - 1 \} \times 100 \text{ and}$$

$$AAPC^{(2)} = \{ \exp(\sum \tilde{w}_i^{(2)} b_i^{(2)}) - 1 \} \times 100 = \{ \exp(\hat{\phi}^{(2)}) - 1 \} \times 100$$

Note that an approximate 100(1-α)% confidence interval for $\phi^{(1)} - \phi^{(2)}$ is $(d_{L(\alpha)}, d_{U(\alpha)})$, where

$$d_{L(\alpha)} = \hat{\phi}^{(1)} - \hat{\phi}^{(2)} - z_{1-\alpha/2} \sqrt{\widehat{Var}(\hat{\phi}^{(1)}) + \widehat{Var}(\hat{\phi}^{(2)})} \text{ and}$$

$$d_{U(\alpha)} = \hat{\phi}^{(1)} - \hat{\phi}^{(2)} + z_{1-\alpha/2} \sqrt{\widehat{Var}(\hat{\phi}^{(1)}) + \widehat{Var}(\hat{\phi}^{(2)})},$$

Then, using a Taylor series expansion, an approximate 100(1-α)% confidence interval for the difference between the two true average annual percent change rates can be obtained as $(\hat{d}_{L(\alpha)}, \hat{d}_{U(\alpha)})$, where

$$\hat{d}_{L(\alpha)} = AAPC^{(1)} - AAPC^{(2)} - z_{1-\alpha/2} 100 \sqrt{e^{2\hat{\phi}^{(1)}} \widehat{Var}(\hat{\phi}^{(1)}) + e^{2\hat{\phi}^{(2)}} \widehat{Var}(\hat{\phi}^{(2)})}$$

$$\hat{d}_{U(\alpha)} = AAPC^{(1)} - AAPC^{(2)} + z_{1-\alpha/2} 100 \sqrt{e^{2\hat{\phi}^{(1)}} \widehat{Var}(\hat{\phi}^{(1)}) + e^{2\hat{\phi}^{(2)}} \widehat{Var}(\hat{\phi}^{(2)})}$$

[Return to the top]

What are the relative advantages and disadvantages of reporting an AAPC over APCs?

Reporting an APC for each joinpoint segment provides a complete characterization of the trend over time. However, sometimes a summary measure over a fixed interval may be desirable. The statistical power to determine if an APC is different from 0 is a function of the length of the interval. Thus, a short segment rising at a steep rate may not be statistically significant. Comparing the last segment of two series (e.g. males and females) sometimes yields seemingly contradictory results when the segments are of very different lengths. Comparing AAPCs of equal lengths from both series is usually a more meaningful comparison. For example, the delay-adjusted thyroid cancer incidence trend (using data from 1975-2002) was rising for males at 2.2% per year from 1980-2000 (and is characterized as rising since it is statistically significant) and at 11.4% per year from 2000 to 2002 (characterized as a non-significant change because the APC is not statistically significant). However, for females it was rising at 5.3% per year from 1993-2002 (characterized as rising since it is statistically significant). Because the last segment for males is relatively short, it introduces uncertainty, and we arrive at the conclusion that recent rates for females are rising, while for males we are uncertain if the rates are rising. To make the comparison between males and females more

comparable, it is useful to compute the AAPC over the same fixed interval for both series. The AAPC for 1993-2002 is 4.2% for males and 5.3% for females (each characterized as rising since they are both statistically significant). The AAPC for 1998-2002 is 6.8% for males and 5.3% for females (each characterized as rising since they are both statistically significant).

Rather than reporting the APC for the final segment for a long list of cancer sites, there may be advantages to reporting the AAPCs over specified fixed intervals. If space permits, reporting both the AAPC and the final segment APC gives an even more complete picture, since each give a somewhat different perspective.

[\[Return to the top\]](#)

What is the advantage of reporting an AAPC over an APC computed by fitting a single line (on a log scale) to the data?

Prior to the development of the joinpoint and AAPC methodology, to characterize a trend over a fixed interval, a single regression line (on a log scale) over the fixed interval was fit, and the slope coefficient was then transformed to an APC. This older methodology has two disadvantages over the AAPC. First, the older methodology assumes linearity of the trend (on a log scale) over the interval, while the AAPC does not. Secondly, the AAPC can be used to characterize a short segment based on a joinpoint model fit over a much longer series. This is especially advantageous for situations when the data are sparse (e.g. a rare cancer or data from a small geographic area). In these cases, because of the variability of the underlying data, the older method might estimate an APC for the period 1996-2005 with a very wide confidence interval. The AAPC from 1996-2005 might be based on an underlying joinpoint model using data from 1975-2005, and thus the resulting AAPC would usually be more stable.

For any additional questions about the AAPC, please contact [Joinpoint Technical Support](#). More technical details about the AAPC are available in the following journal article:

Clegg LX, Hankey BF, Tiwari R, Feuer EJ, Edwards BK. Estimating average annual percent change in trend analysis. *Statistics in Medicine* 2009; 28(29): 3670-82. [\[Abstract\]](#)

[\[Return to the top\]](#)

Estimated Joinpoint (Estimated Tau) and Confidence Interval

What is Tau?

The tau parameters are the unknown joinpoints in the joinpoint regression model described by Kim et al. (2000). For fitted models with one or more joinpoints, the Joinpoint software displays the estimates of the tau parameters and estimates of other joinpoint model parameters

Confidence Intervals - Parametric Method

If $k > 0$ then the output lists the estimated joinpoints. The associated confidence intervals (CI) come from Lerman (1980) equation 6 using C_2^α . Explicitly, the $100(1-\alpha)\%$ confidence interval for the j th of k joinpoints includes all values of x from the grid such that $Q_{x,j,k} \leq C_2^\alpha$, where

$$C_2^\alpha = Q_k\left(1 + \frac{k}{n-p}\right)F_{k,n-p}^{-1}(1 - \alpha)$$

Confidence Intervals - Empirical Quantile Method

For a more detailed description of the Empirical Quantile methodology, please go [here](#).

Related content

- [Parameterizations](#)

Empirical Quantile Confidence Interval

A description of Empirical Quantile Confidence Interval.

NOTES:

- **In version 4.6.0.0, the Empirical Quantile Confidence Interval (CI) calculation was changed and can now use Method 1 or Method 2. It is labeled as an "alpha" feature and considered to be still in development. Please see [Why is the Empirical Quantile Confidence Interval option labeled as "Alpha"?](#) for more details.**

In Joinpoint version 4.2, a new method called the empirical quantile method was implemented to construct a confidence interval for the true AAPC. The motivation behind this was a conservative tendency of the asymptotic confidence interval for the AAPC. As of Joinpoint version 4.6.0.0, the empirical quantile method was implemented to construct a confidence interval for the true APC and tau (a location of a joinpoint). Along with that, two separate methods of the empirical quantile confidence interval are provided: Method 1 and Method 2. In version 4.2, only Method 1 was provided.

Recommendation on When to Use Empirical Quantile Method 1 vs. Method 2 for APC Confidence Intervals

Our simulation study of the APC empirical quantile confidence intervals suggests Method 2 when the number of observations is as small as 10 and Method 1 when the number of observations is greater than or equal to 20. More details on each method are below.

Description of the Empirical Quantile Methods

The idea of the empirical quantile method is to generate resampled data by (i) generating resampled residuals as the inverse function values of the uniform random numbers over (0,1) where the function is the empirical distribution function of the original residuals and then (ii) adding resampled residuals to the original fit. For each resampled data, the model is fit and the AAPC, APC and tau values are estimated. Then, the $100(\alpha/2)$ th and $100(1-\alpha/2)$ th percentiles of the resampled AAPC, APC and tau values are obtained as the lower and upper limits of the $100(1-\alpha)\%$ empirical quantile confidence intervals for the true AAPC, true APC and tau.

For the empirical quantile method implemented in Joinpoint version 4.2, the inverse function values of the uniform random numbers in Step 2 described in Section 3 of Kim et al. (2017) are calculated using u and u' , where u and u' are independent random numbers from the uniform distribution on (0,1). This method is called Method 1. Kim et al. (2017) also describe another way to calculate the inverse function values with $u' = (u - \frac{i}{n+1}) * (n+1)$ where u is a random number from the uniform distribution on (0,1), and we call this Method 2. Kim et al. (2017) provide simulation study results for the AAPC empirical quantile confidence intervals. Additional simulation studies for the APC and tau empirical quantile confidence intervals, show that (i) the coverage probabilities of the Method 2 confidence intervals for the true APC and tau are usually greater than those of the Method 1 confidence intervals, (ii) the coverage probabilities of the Method 1 confidence interval for the true APC were often much below the nominal level, especially when the number of observations is small, and (iii) the Method 2 confidence interval for the true APC is very conservative sometimes, although its coverage probabilities varied less than those of the Method 1 APC confidence interval.

Starting in Joinpoint version 4.6.0.0, both Method 1 and Method 2 are provided. Our simulation study of the APC empirical quantile confidence intervals suggests Method 2 when the number of observations is as small as 10 and Method 1 when the number of observations

is greater than or equal to 20.

The empirical quantile confidence interval for the true APC is shown to be more robust in terms of the segment length than the asymptotic parametric interval. In our simulation study, it was shown that the asymptotic parametric interval for the true APC is often very liberal when a segment is very short, but the empirical quantile confidence interval maintains the coverage probability to the nominal level in most cases. For the confidence interval for tau, the asymptotic confidence interval and the empirical quantile confidence interval were observed to perform comparably.

For details regarding this method please see the [Improved Confidence Interval for Average Annual Percent Change in Trend Analysis](#) article.

To adjust the random number seed involved with producing the resampled residuals, please go to the [Preferences](#) help section.

Related content

- [Preferences](#)
- [APC/AAPC/Tau Confidence Intervals](#) A description of the methods used for calculating APC, AAPC, and Tau Confidence Intervals.

AAPC Segment Ranges

From the AAPC Segment Ranges section, you can select to display the AAPC for the entire range of data, the last N observations, or other specified ranges.

The [AAPC Confidence Intervals](#) can be calculated using the **Parametric** option or the **Empirical Quantile** option.

When the **AAPC Confidence Intervals Parametric Method** option is selected, the AAPCs can be computed after the Joinpoint calculations are performed, using the APCs. With this option, the AAPCs can be modified in the output by following the following procedure, without having to rerun the session.

1. When viewing the Joinpoint output, to specify which ranges to display use the **Output > Specify AAPC Ranges...** menu item from the top of the main Joinpoint Output window.
2. From the **Specify AAPC Ranges** window, you can select to display the AAPC for the entire range of data, the last N observations, or other specified ranges.

When the **AAPC Confidence Intervals Empirical Quantile** option is selected, the AAPCs cannot be computed after the Joinpoint calculations are performed.

Related content

- [APC/AAPC/Tau Confidence Intervals](#) A description of the methods used for calculating APC, AAPC, and Tau Confidence Intervals.

Advanced Analysis Tools Tab

This is where you can select your:

- Pairwise Comparison options
- Jump Model and Comparability Ratio Model options
- Multi-group clustering (will be available in the future).

Related content

- [Pairwise Comparison](#)
- [Multi-Group Clustering](#)
- [Jump Model / Comparability Ratio Model](#) The Jump Model / Comparability Ratio Model in the Joinpoint software provides a direct estimation of trend data (e.g. cancer rates) where there is a systematic scale change, which causes a “jump” in the rates, but is assumed not to affect the underlying trend.

Pairwise Comparison

The main goal of the comparability test is to compare two sets of trend data whose mean functions are represented by joinpoint regression. Specific interests are testing:

1. whether two Joinpoint regression functions are identical (test of coincidence) or
2. whether the two regression mean functions are parallel (test of parallelism)

and the details can be found in Kim et al. (2004):

H-J. Kim, M. P. Fay, B. Yu, M. J. Barrett and E. J. Feuer (2004), Comparability of Segmented Line Regression Models, *Biometrics*, 1005-1014.

Consider the Joinpoint regression mean function at the j^{th} year of the i^{th} group, $x = x_{ij}$,

$$\mu_i(x) = \beta_{i,0} + \beta_{i,1}x + \delta_{i,1}(x - \tau_{i,1})^+ + \dots + \delta_{i,k_i}(x - \tau_{i,k_i})^+,$$

where k_i is the unknown number of joinpoints, the T's are the unknown joinpoints, the β 's and the δ s are the regression coefficients, and $a^+ = a$ for $a > 0$ and 0 otherwise. In general, the two groups may have different numbers of joinpoints, say k_i for the i^{th} group, but the program fits both groups with a larger model with k_{max} joinpoints.

Computation Notes:

- These tests are selected using the Pairwise Comparison option under the Advanced Analysis Tools tab.
- Using these tests could substantially increase the computation time.
- Autocorrelated errors models are not available with these tests.
- All 2-way combinations of the innermost by-variable are tested.

What are the hypotheses?

The null hypotheses for the test of coincidence and the test of parallelism are

$$H_0 : (\tau_{1,1}, \dots, \tau_{1,k_{\text{max}}}, \beta_{1,0}, \beta_{1,1}, \delta_{1,1}, \dots, \delta_{1,k_{\text{max}}}) = (\tau_{2,1}, \dots, \tau_{2,k_{\text{max}}}, \beta_{2,0}, \beta_{2,1}, \delta_{2,1}, \dots, \delta_{2,k_{\text{max}}})$$

and

$$H_0 : (\tau_{1,1}, \dots, \tau_{1,k_{\text{max}}}, \beta_{1,1}, \delta_{1,1}, \dots, \delta_{1,k_{\text{max}}}) = (\tau_{2,1}, \dots, \tau_{2,k_{\text{max}}}, \beta_{2,1}, \delta_{2,1}, \dots, \delta_{2,k_{\text{max}}}),$$

respectively.

What is the test static?

The test statistic with $k = k_{\text{max}}$ is

$$F_k = \frac{(RSS_{H_0,k} - RSS_{H_1,k})/d_{1,k}}{RSS_{H_1,k}/d_{2,k}},$$

where RSS denotes the residual sum of squares obtained from the least squares fitting, and $d_{1,k}$ and $d_{2,k}$ are appropriate degrees of freedom.

How is the k_{max} chosen?

The recommended data driven choice of k_{max} is the largest number of joinpoints estimated under the null and alternative models:

$$\hat{K}^+ = \max(\hat{K}_0, \hat{K}_1, \hat{K}_2),$$

where \hat{K}_0 is the number of joinpoints estimated for the two groups combined under the null model, and \hat{K}_1 and \hat{K}_2 are the numbers of Joinpoints estimated separately.

How is the p-value estimated?

The p-value of the test is estimated using the permutation distribution of the test statistic. The residuals, obtained under the null model, are permuted to generate the permutation distribution of the test statistic, and the p-value is estimated as the proportion of the permutation data sets whose test statistic values are greater than or equal to the original test statistic value.

Significance Level

Set the significance level for the test of coincidence or the test of parallelism.

Max Number Of Randomly Permuted Data Sets

The program performs a permutation procedure for the test of coincidence or the test of parallelism. Since fitting all $N!$ possible permutations of the data would take too long, the program takes a Monte Carlo sample of these $N!$ data sets, using a random number generator. Specify the size of the Monte Carlo sample of permuted data sets.

The minimum allowed number of permutations is 1000. The maximum number is 10,000. For greater consistency in the p-values obtained if one were to change the seed for each run, we strongly recommend running the program for at least 4499 permutations. You may use a smaller number of permutations to speed up the calculations, but the permutation test may produce less consistent results. Unless you have a good understanding of the implications of changing the number of permutations, we recommend against it.

Related content

- [Data Errors That Stop An Analysis](#)

Jump Model / Comparability Ratio Model

The Jump Model / Comparability Ratio Model in the Joinpoint software provides a direct estimation of trend data (e.g. cancer rates) where there is a systematic scale change, which causes a “jump” in the rates, but is assumed not to affect the underlying trend.

NOTES:

- **The Jump and Comparability Ratio models are available in Joinpoint as of version 4.4.0.0, but are labeled as an "beta" feature and considered to be still in development.**

Models

The jump and comparability ratio model in the Joinpoint software provide a direct estimation of trend data (e.g. cancer rates) where there is a coding change, which causes a “jump” in the rates, but is assumed not to affect the underlying trend. The user supplies the location of the last data point before the coding change occurs. The software automatically locates the discontinuity or “jump” halfway between this last data point and the next one. The software currently only allows for a single “jump”. For example, the ICD coding changes (Anderson et al. 2001) from ICD-9 to ICD-10 for classification rules for selecting underlying causes of death. The last year that ICD-9 was used was 1998, and ICD-10 was implemented starting in 1999. If the user had entered annual data, they would enter 1998, and the software would place the “jump” at 1998.5.

There are two different options for how a model can be implemented to allow for a coding change:

Comparability Ratio Model: In some cases a “double coding” study has been conducted, where a certain number of cases have been coded under both the old and new systems. In such cases a “comparability ratio” and its standard error can be externally entered, where the comparability ratio is defined as:

$$\frac{\text{count under new code}}{\text{count under old code}}$$

For example, the National Center for Health Statistics (NCHS) maintains a [web page on comparability ratios derived from double coding studies](#), and has a report ([Anderson et al.\(PDF\)](#)) which contains the comparability ratios and associated standard errors for a long list of underlying causes of death. In other cases (for example, changes in how stage of disease is coded for cancer), there may be a number of years where both staging systems were used simultaneously, and the comparability ratio can be derived from these years of overlap. In the comparability ratio model, the data before the jump is multiplied by the comparability ratio (and new standard errors are computed using the delta method utilizing both the standard error of the data point and the standard error of the comparability ratio). The new series is fit using a standard joinpoint model, and the transformed data points and fitted values are transformed back to the original coding (by dividing by the comparability ratio) prior to graphing.

Jump Model: In other situations, there may be no “double coding” studies available. In these situations, the “jump” is a parameter in the model (rather than entered externally) and is estimated directly. It is the underlying assumption of this model (i.e. that the same trend continues before and after the jump) that allows this type of model to be estimated. For the

jump model, the user only has to enter the location of the last data point before the coding change occurs. The model can be described as below. For $y = \log r$, where r is the rate at a given time, and x is the time, assume:

$$y_i = \log(r_i) = \beta_0 + \beta_1 x_i + \delta_1(x_i - \tau_1)^+ + \dots + \delta_K(x_i - \tau_K)^+ + \gamma I(x_i \geq s) + \varepsilon_i,$$

where τ 's are unknown change-points, s is the known location of the coding change, ε_i are independent errors, the symbol a^+ represents the a for $a > 0$ and 0 if $a \leq 0$, I is an indicator function equaling to 1 if the condition is satisfied, β 's, δ 's, τ 's and γ are the model parameters to be estimated. The parameter γ represents the jump and $\exp(\gamma)$ represents the ratio of rates coded under the new coding scheme divided by rates coded under the old coding scheme (i.e. the comparability ratio estimated from the jump model).

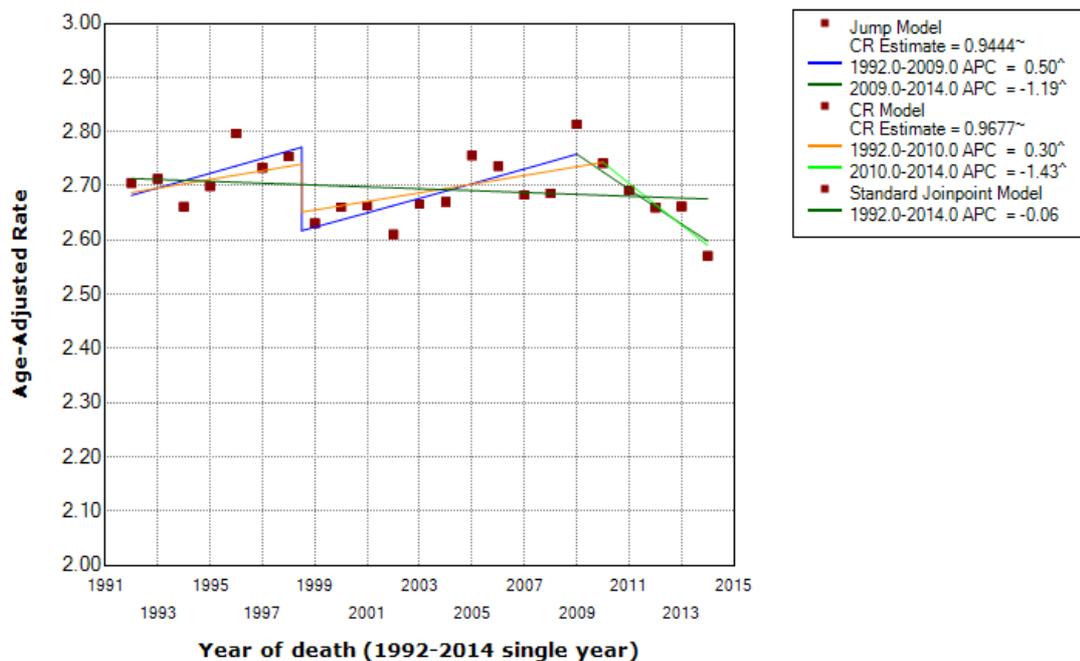
An example is provided to demonstrate how incorporating a coding change (even one that is relatively modest) can change the overall conclusions about the trends. For melanoma, the published ICD-9 to ICD-10 comparability ratio is 0.9677 (SE = 0.0032, 95% CI (0.9614, 0.9741)). Figure 1 shows US melanoma mortality for all races and both genders from 1992 through 2014 using the standard joinpoint model, the comparability ratio model, and the jump model. The standard joinpoint model found no joinpoints, and shows a flat trend with a non-significant APC of -0.06% per year. A comparability ratio less than one (i.e. 0.9677) forces a sudden drop in the trend line between 1998 and 1999. With this shift, the comparability ratio model shows a joinpoint in 2010 with a significant rise of 0.30% per year prior to 2010 and a significant fall from 2010 to 2014 of 1.43% per year. The jump model estimates a similar comparability ratio of 0.9444, and finds a joinpoint at 2009 with a significant rise of 0.50% per year prior to the joinpoint, and a significant decline of 1.19% after. These are qualitatively different results when the coding change from ICD-9 to ICD-10 is taken into account.

Motivated by this example, the complete set of US melanoma mortality results in NCI's Cancer Statistics Review, with results from the standard joinpoint model, jump model and comparability ratio model, can be found on the Melanoma example.

Figure 1. Standard joinpoint model, jump model, and comparability ratio model for all races and both genders US mortality for melanoma, 1992-2014. The estimate of the comparability ratio estimated from the jump model is 0.9444 with standard error = 0.0116 (the estimate of the comparability ratio is statistically different than 1). The comparability ratio (input from a double coding study) is 0.9677 with standard error = 0.0032 (the comparability ratio is statistically different than 1).

All Races / Male and female

Jump Model: 1 Joinpoint, CR Model: 1 Joinpoint, Standard Joinpoint Model: 0 Joinpoints



[^] Indicates that the Annual Percent Change (APC) is significantly different from zero at the alpha = 0.05 level.

[~] Significant at level 0.05 for test of Comparability Ratio = 1.

Final Selected Model: Jump Model - 1 Joinpoint, CR Model - 1 Joinpoint, Standard Joinpoint Model - 0 Joinpoints.

Input Parameters

- For the jump model, the user supplies:
 - The location where the coding changes start to occur.
 - All other parameters set in the standard joinpoint regression model such as minimum and maximum number of joinpoints, permutation methods, etc.
- For the comparability ratio model, the user supplies:
 - The location where the coding changes start to occur.
 - The size of the comparability ratio. The ratio is limited to [0.01, 100].
 - The variance of the comparability ratio. The variance is limited to [0, 100]. If you use the homoscedastic error option for Joinpoint (i.e. no standard errors) the variance of the comparability ratio will be set to "NA" (not applicable) and will not be able to be changed. If the heteroscedastic error option is used, the user may enter zero if they do not have an estimate of the variance of the comparability ratio, but they are encouraged to enter the variance if they have one. All other parameters set in the standard joinpoint regression model such as minimum and maximum number of joinpoints, permutation methods, etc.
- **Restrictions:**
 - The Jump Location must be 2 or more observations from either end of the data.
 - If the user selects to put the jump location 2 observations from either end of the data, the following warning message will be displayed: "In general, it is not recommended to place a jump too close to the ends because that tends to cause unstable results resulting from a confounding of the locations of the joinpoints and size of the jumps. For example, instead of finding a joinpoint, the model makes an large jump. Even though the software allows the jump location to be as close as to 2 data points away, the users need to be cautious about whether the jump size is due to confounding."
 - The jump and comparability ratio models can only be run with the Grid Search method.

- The jump and comparability ratio models can only be run when the Log Transformation is set to Yes.

Output

- Both the jump and comparability ratio models will output all the items/statistics produced by the standard joinpoint model.
- The jump model will produce additional statistics related to the jump value and also provide an estimate of the comparability ratio: $\text{comparability ratio} = \exp(\text{jump value})$. While in almost all situations the comparability ratio and the jump values should both be either statistically different from 0 and 1 respectively (since they are related by a transformation, occasionally, the comparability ratio and jump value do not show the same statistical significance. This minor discrepancy rarely happens. It is due to an approximation from the delta method applied to finding the confidence interval of the comparability ratio of the jump model.

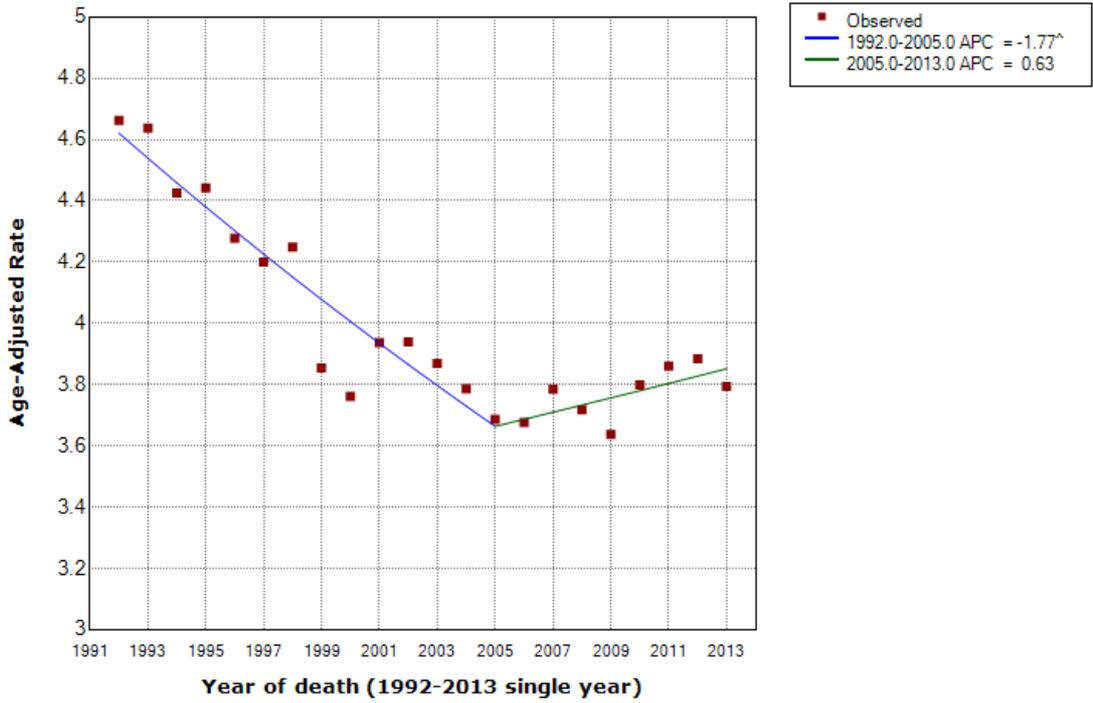
Which model to use?

Considerations of which model to use could include:

- No “double coding” study may be available, in which case the jump model is the only option.
- The “double coding” study on which the comparability ratio is estimated usually is conducted using data from calendar years close to when the coding change occurred. However, the actual ratio may vary as one gets further from the year the coding change occurred. The jump model implicitly uses all of the years before and after the coding change to estimate a best fitting jump.
- The population for which the “double coding” study was conducted may differ from the population for your data series (e.g. the double coding study may have been conducted for all races and your data series may be for blacks, or the double coding study may have been conducted in one cancer registry, but the data series is for a different registry). The jump model has some advantages in this case because it is estimated directly using the data series of interest.
- A joinpoint may be close to the location of the jump. In this case, the estimate of the size of the “jump” in the jump model may be partially confounded with the slope before and after the joinpoint. For example, a series for non-Hispanic white males for oral cavity and pharynx cancer mortality is shown in Figures 2 thru 4. The Standard Joinpoint model is shown in Figure 2 and displays an annual percent change (APC) of -1.77% from 1992 through 2005 and a non-statistically significant APC of 0.63% from 2005 to 2013. The comparability ratio (estimated from a double coding study) is 0.9603 and shows an APC of -1.36% from 1992 through 2005, and a non-statistically significant APC of 0.53% from 2005 to 2013 (Figure 3). The Jump model (Figure 4) estimated a comparability ratio of 0.8844 which is further from the null value of 1 as compared to the value from the comparability ratio model (0.9603). An examination of this model shows a joinpoint at 1997 which is very close to the coding change at 1998.5. The upward APC segment from 1997 through 2002 is only made possible by the large compensating downward jump, and appears to be a spurious result.

Figure 2. Standard joinpoint model for White non-Hispanic Male US Mortality for Oral Cavity and Pharynx Cancers, 1992-2013.

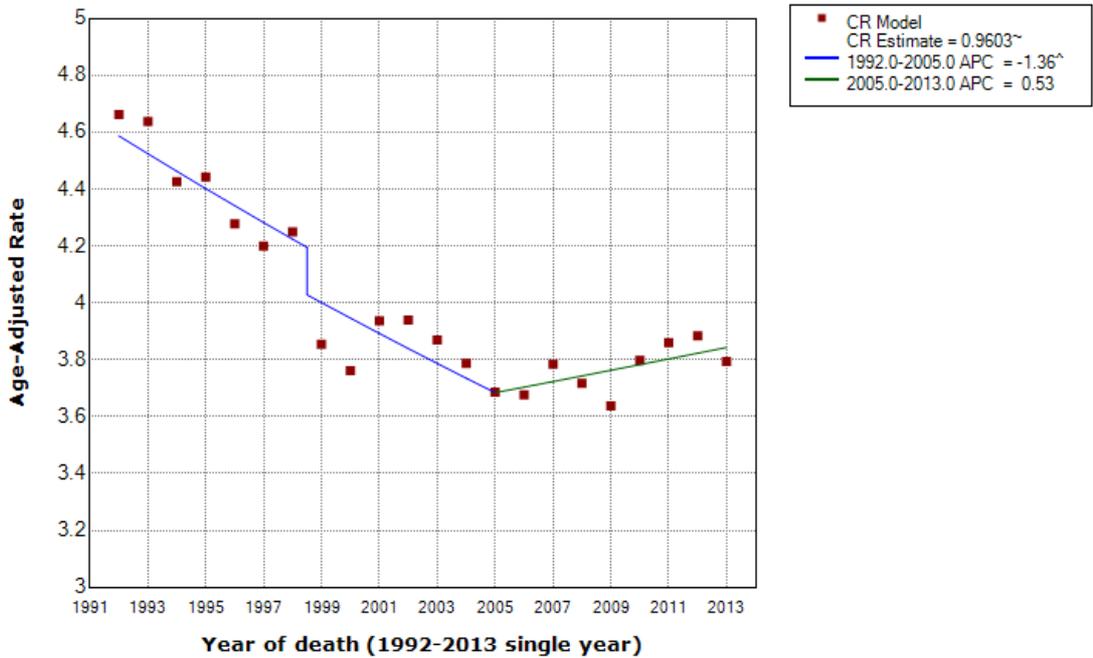
Oral Cavity and Pharynx / White Non-Hispanic / Male: 1 Joinpoint



^ Indicates that the Annual Percent Change (APC) is significantly different from zero at the alpha = 0.05 level.
Final Selected Model: 1 Joinpoint.

Figure 3. Comparability ratio model for White non-Hispanic Male US Mortality for Oral Cavity and Pharynx Cancers, 1992-2013. The comparability ratio (input from a double coding study) is 0.9603 with standard error = 0.0039.

White Non-Hispanic / Male
CR Model: 1 Joinpoint

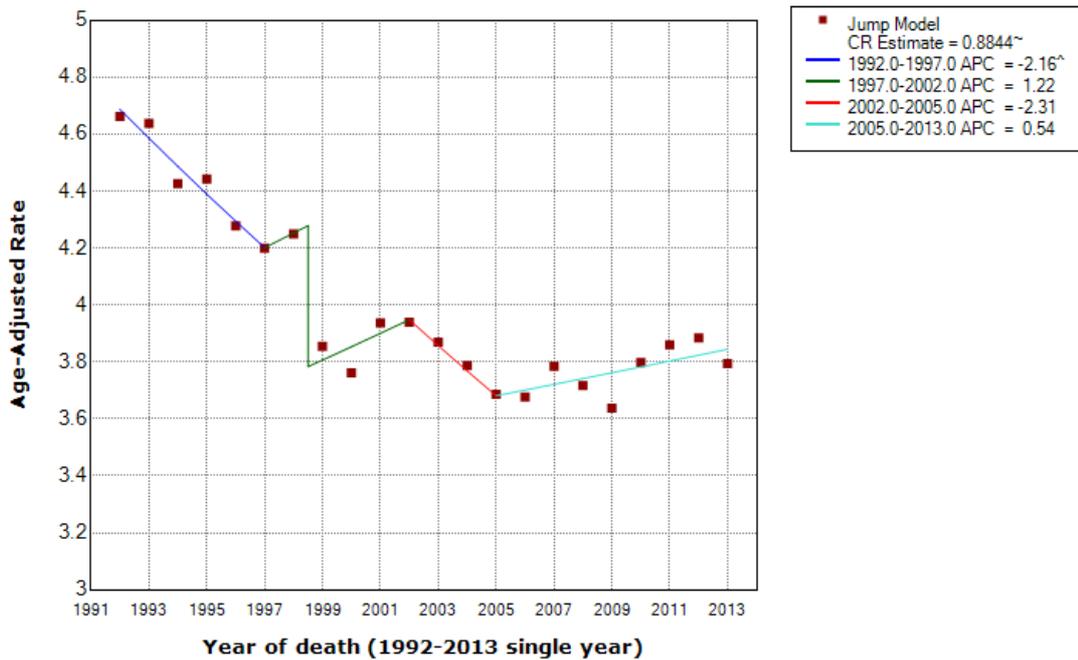


^ Indicates that the Annual Percent Change (APC) is significantly different from zero at the alpha = 0.05 level.
~ Significant at level 0.05 for test of Comparability Ratio = 1.
Final Selected Model: CR Model - 1 Joinpoint.

Figure 4. Jump model for White non-Hispanic Males US Mortality for Oral Cavity and

Pharynx Cancers, 1992-2013. The estimate of the comparability ratio is 0.8844 with standard error = 0.0311.

**Oral Cavity and Pharynx / White Non-Hispanic / Male
Jump Model: 3 Joinpoints**



^ Indicates that the Annual Percent Change (APC) is significantly different from zero at the alpha = 0.05 level.
 ~ Significant at level 0.05 for test of Comparability Ratio = 1.
 Final Selected Model: Jump Model - 3 Joinpoints.

- The underlying variability of the data may make estimation of a small or modest jump size impossible. In small sub-populations (e.g. API, AI/AN, rare cancer sites, or small geographic areas), such situations may occur. Since the jump is a parameter estimated in the jump model, a test can be conducted if the jump is statistically greater than zero. In cases where the jump size is insignificant, the comparability ratio model may be the better choice. Even in situations where the jump is statistically significant, if there is large variability, and the comparability ratio is small, one should be wary of estimates of the jump which differ widely from the comparability ratio.
- Fitting both models is usually a good idea. The “safer bet” is usually the comparability ratio model since the jump model can occasionally produce anomalous results. However, the jump model can offer better estimates if the double coding study was estimated from a limited range of years or from a population which differs from the data series being modeled. In many cases where the estimates from the two models are similar, the jump may be preferred.
- In cases where no double coding studies exist, one should be cautious in accepting the results of the jump model. The analyst should evaluate the size of the underlying variability of the data, and should be suspicious of joinpoint segments which start or end close to the jump location, and the slope of the segment seems to be “compensating” for the size of the jump.

In general, it is best to do a careful examination of the models using the criteria above before deciding which model to select. An algorithmic approach may be desired in the case of many data sequences to be analyzed. See Melanoma for an example of an algorithmic approach.

Display

The software allows viewing of a graph for each joinpoint model, from the model with the

minimum number of joinpoints to the model with the maximum number of joinpoints. The graph will show a “jump” at the location where the coding change occurs. The software also allows for graphs of different models and/or cohorts to be combined.

Citation & Reference

- For more technical details on the Jump Model / Comparability Ratio Model, please contact [Joinpoint Technical Support](#).
- Anderson RN; Miniño AM, H.D.R.H., *Comparability of Cause of Death Between ICD-9 and ICD-10: Preliminary Estimates*. National Vital Statistics Reports, 2001. **49**(2): p. 2001-1120.

Multi-Group Clustering

Multi-Group Clustering is a future enhancement. In version 4.6.0.0, the clustering controls are no longer available.

Preferences

File Preferences

Access the File Preferences window by selecting Preferences from the File menu. Make changes to the fields as necessary. When you are finished, click **OK**.

Default Data Directory

The default directory to be used when browsing for your data from a Joinpoint session can be set here. In order to set the directory, use the browse button and select the directory to be used as the default.

Display Splashscreen on startup

When the Joinpoint program is started, a splashscreen will automatically appear for 5 seconds. This option can be used to display or hide the screen.

Maximum Number of Threads

With the introduction of processing multiple cohorts simultaneously (threading), an option to adjust the number of threads is provided. The number of threads that Joinpoint will try to use when running a session can be set here. There number cannot be set above the maximum for the computer that Joinpoint is installed on.

A reason to lower the number of threads is to provide other concurrently running Windows programs more resources under Windows. The more threads that Joinpoint tries to use, the slower other concurrently running Windows programs might run.

Turn on all Joinpoint warning messages

Joinpoint has various 'nag' warning messages that appear to users to warn them of potential problems. A 'nag' warning is a warning that users can choose to turn off and when turned off it will remain off until this option is adjusted. When this button is clicked, all 'nag' warning messages previously turned off will be turned back on.

Session Preferences

Access the Session Preferences window by selecting Session Preferences from the Session menu (only available when a session is open). Make changes to the fields as necessary. When you are finished, click **OK**.

Alpha Levels

Joinpoint Locations - The significance level used to test the Type I error when selecting a model. The default value for this field is 0.05. Alpha level values must be greater than 0 and less than 1.

APCs - The significance level used to test whether an APC value is significantly different from zero. The default value for this field is 0.05. Alpha level values must be greater than 0 and less than 1.

AAPCs - The significance level used to test whether an AAPC value is significantly different from zero. The default value for this field is 0.05. Alpha level values must be greater than 0 and less than 1.

Jump Value and Comparability Ratio - The significance level used to test whether a Jump Value or Comparability Ratio is significantly different from zero. The default value for this field is 0.05. Alpha level values must be greater than 0 and less than 1.

Permutation Test Seed

Typically, you should allow the computer to use the default seeds for the random number generator. By using these default seeds, one can duplicate results from a previous Joinpoint session even though the program uses Monte Carlo sampling. Although not recommended in general, the Joinpoint Regression Program allows one to change a default seed here. The default value is 7160.

Empirical Quantile Seed

For the Empirical Quantile confidence interval calculation method, you can adjust the value of the seed used for the random number generator involved with producing the resampled residuals for each cohort. To set the initial seed value, use the "Seed Value" text box. The default value is 10,000.

The seed can be kept constant or changed for each cohort. The seed options are:

Constant Seed - this option will keep the seed the same for all cohorts in your data file. The seed used will always be the number in the "Seed Value" text box.

Varying Seed - this option will change the seed for each cohort in your data file. The first cohort in your data file will use the seed value set in the "Seed Value" text box. For each cohort after that, the seed is computed using the following formula: $(\text{Initial Seed Value} * \text{Cohort Number}) + (\text{Cohort Number} - 1)$. The "Initial Seed Value" is the value set in the "Seed Value" text box. The "Cohort Number" is the number of the cohort that is being processed. The first cohort in the data file is 1, the second 2, etc.

Joinpoint Session Template

A file type that allows users to save a Joinpoint Session and reopen it with a different data file.

The Joinpoint Session Template (.jpt) is a file type that allows users to save a Joinpoint Session and reopen it with a different data file. The template should only be used when you have multiple similar input files that you want to analyze in Joinpoint. The new input files should have the same “structure” as the file the template was originally saved with. “Structure” meaning that all the fields should be in exactly the same order as the original input data file and the only difference is in the values of the Dependent Variable.

The new input file(s) CANNOT HAVE:

- A different number of fields in it than the input file it was originally saved with. All new input files MUST have the same number of columns/fields.

The new input file(s) CAN HAVE:

- Different Independent Variable and By Variable definitions than it was saved with. BUT, you will have to update those variable formats through the variable “Define” buttons. The template will contain the original formats for these variables. It will NOT adjust them automatically.

When you open a Session Template file, the first thing that will happen is that you will be prompted to select a data file. As previously mentioned, this data file should have the same general structure as the original data file. If the data file varies too much in its fields, you will have numerous problems to fix in the Session window. Once a data file has been selected, the Session Window will be displayed with all of the settings of the original session. If the new input file has the same By Variable and Dependent Variable values as the original input file and are in the same field location, then you should be able to just execute the session at this point.

To save a Joinpoint Session Template:

You must first set all the required session variables and settings. Then choose “Save Joinpoint Session Template...” from the “File” menu. You will be prompted for a filename and location, and then a JPT file will be created on your hard drive.

Executing the Joinpoint Program

Once you have set all controls in the Joinpoint Regression Program, execute the program by clicking  on the toolbar or clicking Run on the menu bar.

A progress meter will be shown on the screen while the Joinpoint calculation engine processes the data and generates the output. Click Cancel if you would like to halt the calculations before they are complete.

Once the processing is complete, an output window will automatically be displayed. You will need to save this output if you would like to access it again at a later time.

Related content

- [Processing Multiple Cohorts Simultaneously \(Threading\)](#)
- [Statistical Notes](#)
- [Viewing Results](#)
- [Data Files Containing Cohorts with an Unequal Number of Observations](#) As of version 4.4.0.0, Joinpoint can process data files that contain cohorts with an unequal number of observations.

Processing Multiple Cohorts Simultaneously (Threading)

Because the processing of each data cohort in Joinpoint is independent of the others, the ability to process multiple cohorts simultaneously was introduced into the program starting with Version 4.1.0. This new functionality could greatly decrease the time it takes to execute sessions in Joinpoint. Also due to this change, the job progress window has been adjusted to display the status of each cohort being processed at any given time. As in the past, there will also be an overall job progress meter displayed during the execution of any Joinpoint session.

Please note that this feature will not work when executing Joinpoint under Windows XP. Windows XP users will only be able to process one cohort at a time (similar to previous versions of Joinpoint).

Input Data – Errors and Warnings

When a session is executed, Joinpoint will first conduct a series of edits on the data. Some edit failures will cause the job to stop, while others will produce warnings.

Related content

- [Data Errors That Stop An Analysis](#)
- [Data Warnings Displayed by Joinpoint](#)

Data Errors That Stop An Analysis

- If there are too few data columns in the input data file.
Error Message: Too few data columns in Input Data File.
Reason(s) triggered: This is triggered when the variable positions set in the session go beyond the number of variables found in any given data line from the input data file. Most likely the input data file (or data file name) was changed and now it does not match up with what was used to set up the original session.
- If Joinpoint is computing age-adjusted rates and encounters a by-group with a different number of observations from the previous by-group.
Error Message: You cannot calculate age-adjusted rates from a data file with differing number of observations per cohort!
- If the shift-value specified by the user is greater than the largest interval between independent variable values.
Error Message: Error with shift value: The shift value cannot be larger than the largest independent variable interval.
- If the Independent variable has a format and Joinpoint encounters a value that is NOT in the format.
Error Message: "???" is not a valid value for the independent variable.
- If the independent variable has a value that is non-numeric.
Error Message: Error in Data File: Joinpoint cannot process by-groups where independent variable values are non-numeric.
- If two or more independent variable values are found to be equal (within a given cohort).
Error Message: Error in Data File: Joinpoint cannot process by-groups where two or more independent variable values are equal.
- For comparison runs, if the inner-most by variable has only one grouping.
Error Message: You cannot test for pair-wise differences when the inner-most by variable only has one category.
- When age-adjusted rates are computed, if the standard population value used to compute an age-adjusted rate is zero.
Error Message: Standard Population denominator is zero.
- If there are one or more cohorts that do not have enough data points for the specified job parameters. The following formula is used to determine the minimum number of required data points for any given cohort:

Min Required Observations = $(2 * \text{Min Number of Obs from a Joinpoint to either end of data}) + ((\text{Max Joinpoint} - 1) * (\text{min num of observations between two joinpoints})) + \text{Max Joinpoint}$;
Error Message: Too few observations to fit specified model. Please verify that your data is sorted correctly.

Data Warnings Displayed by Joinpoint

When a session is executed, users will first be notified if any of the following problems exist in the cohort data:

General Warnings

When analyzing counts or computing crude rates, Joinpoint will add 0.5 to any zero count under the circumstances listed [here](#). The following WARNING will be displayed once per analysis execution:

- Warning Message: There are records with a value of zero for count, 0.5 will be added to each zero count value.

Cohort-Specific Warnings

If any of the following problems exists, a special window will appear which will list all the errors for each cohort. From the window, users can either end the job or continue it. If they choose to continue the job, any cohort with one or more of the following problems will not be analyzed.

If a record is found with the dependent variable value missing.

- Warning Message: Joinpoint cannot process records with missing dependent variable values. Joinpoint will not analyze by-groups that have this error.

If the dependent variable value is 0 and it is a log linear run, then the dependent variable cannot have records with the value of zero.

If a record is found with the independent variable value missing.

- Warning Message: Joinpoint cannot process records with no independent variable. Joinpoint will not analyze by-groups that have this error

If a record is found where the dependent variable value is non-numeric.

- Warning Message: Error in Data File: Joinpoint cannot process records with independent variable values that are non-numeric.

For analyses where standard errors are read in/used:

- If the SE is missing - Warning Message: Joinpoint cannot process records with missing weight values. Joinpoint will not analyze by-groups that have this error.
- If the SE is less than or equal to zero - Warning Message: Joinpoint cannot process records with weight variable ≤ 0 . Joinpoint will not analyze by-groups that have this error.

When crude rates are calculated by Joinpoint, the following will trigger a warning and the

cohort will not be analyzed by Joinpoint:

- There were one or more count values that were non-numeric, missing, or less than zero.
- There were one or more population values that were non-numeric, missing, or less than zero.
- There were one or more population values of zero.
- If the standard error is required for the analysis and one or more count values are zero, then the associated SEs will be zero. A zero SE is not allowed.
- If it is a Log Linear analysis and a count value of zero is found.

When age-adjusted rates are calculated by Joinpoint, the following will trigger a warning message and the cohort will not be analyzed by Joinpoint:

- There were one or more count values that were non-numeric, missing, or less than zero.
- There were one or more population values that were non-numeric, missing, or less than zero.
- There were one or more standard population values that were non-numeric, missing, or less than zero.
- The total of the standard population does not match the total obtained from the first cohort.
- If it is a Log Linear analysis and the final age-adjusted rate value is zero.
- One or more of the following adjustment variable problems exist:
 - The adjustment variable only has 1 category/group.
 - The records for the cohort are not evenly divisible by the number of adjustment variable groups.

When proportions are calculated by Joinpoint, the following will trigger an error:

- There were one or more numerator or denominator values that were non-numeric or negative.
- There are one or more records where the denominator was zero.
- There are one or more proportions that are less than zero.
- There are one or more proportions that are greater than 1.0 (numerator greater than denominator).

When percentages are calculated by Joinpoint, the following will trigger a warning:

- There were one or more numerator or denominator values that were non-numeric or negative.
- There were one or more records where the denominator was zero.
- There are one or more percentages that are less than zero.
- There are one or more percentages that are greater than 100.00 (numerator greater than denominator).

Related content

- [Dependent Variable](#)
- [Independent Variable](#)

Viewing Results

This window displays the Output results of your joinpoint analysis. The various tabs (Graph, Data, Model Estimates, Trends, Model Selection, or Comparison) allow you to view graphs, view the graph data, or view the other analysis results. You may also access both the graph data and the report information by exporting these for use by other software packages (Excel, SAS, etc.).

The Cohort Tree control on the left of the Output Window selects which By-Group and Model (i.e. the 1 joinpoint model or the 3 joinpoint model) to display. The bottom of this window displays the Rate setting and the version of Joinpoint that this output was created in.

The Joinpoint calculation engine displays the output window but doesn't automatically save or create a permanent output file. You can save the output as a single binary file and/or you can export the graph data and the report information as text files. JPO (Joinpoint Output) is the extension Joinpoint uses for the output files.

Navigating through the different cohorts: Cohort Tree

How different session types are displayed: Viewing Output from the Different Analysis Types

Tabs you can view on this window are:

- Graph
- Data
- Model Estimates
- Trends
- Model Selection
- Comparison

Possible options from this window are (most are accessible from the **Output** menu):

- Saving the results - From the menu, select **File > Save As...** Output files are saved with a JPO file extension.
- Retrieve the Session - See Retrieving a Session.
- View Multiple Joinpoint Models - primarily used to view multiple cohort model results on one graph. See View Multiple Joinpoint Models
- Specify AAPC ranges - See Specify AAPC Ranges.
- Display Options - See Display Options.
- Display Numeric Values - This option will display cohort numbers instead of names in the cohort selection control at the top of the window.
- Exporting the results to a Text file - See Exporting Results to Text.
- Exporting the results to an Excel file - See Exporting Results to Excel.
- Printing the results - See Print Options.
- Viewing the Output properties - See Output Properties.

Related content

- [Exporting Results to Excel](#)
- [Retrieving a Session](#)

- Specify AAPC Ranges
- Display Options
- Print Options
- Exporting Results to Text
- Graph Tab
- Data Tab
- Model Estimates Tab
- Trends Tab
- Model Selection Tab
- Comparison Tab
- Output Properties

Cohort Tree

The Cohort Tree allows you to select the cohort to view in the Output Window.

On the left-hand side of the Output window is the Cohort Tree interface. The Cohort Tree interface displays the number of join point models generated for each unique combination of By-Variable values (which is called a cohort). By selecting check boxes in this list, you can display the associated graph and data for that joinpoint model.

Tree List

The Cohort Tree allows you to select which joinpoint model you wish to display for a given cohort. Just select a check box to view the model results for that cohort. Only one item can be selected at a time. Several icons may also be displayed:

- ★ This is the Final Selected Model.
- ☆ This is the selected model for a grouping not chosen.
- ⊗ This cohort does not have any joinpoints to display or has invalid data.
- ✕ This cohort has been excluded by the user.

Tree Tool Bar

The bar at the top of the Cohort Tree gives you access to navigation and filtering options. The number of cohorts displayed is shown first, followed by a Filter button, and then 2 green arrow buttons. The Filter button will display the Filter Cohorts pop-up window to allow you to configure the display of cohorts in the Cohort Tree. The green arrow buttons will allow you to move through the cohorts that are available.

Filter Cohorts Window

This Filter Cohorts pop-up window allows you to configure the cohorts that are displayed in the Cohort Tree.

- **Search** - When you type text in this box, the Cohort Tree will be filtered to only show cohorts which have this text. Next to this search box, the X button will clear your search.
- **Cohorts** - The check boxes allow you to select the types of cohort you want to display. Your choices are Valid, Invalid (⊗), and Excluded (✕). The Valid option allows you to pick the types of models to display: All Models, Final Selected Model (★), Specific Model. Invalid cohorts are those which could not be processed successfully by Joinpoint. Excluded cohorts are those that you chose to exclude from this session in the Session window.
- **Restrict Cohorts By** - This option allows you to restrict the cohorts that will be displayed by specific By-Var values. A selection window will allow you to pick those values you want to display.
- **View Multiple Models Simultaneously** - This option will allow you to select more than one model for display at the same time. This is useful if you want to compare the graphs of multiple joinpoint models within the same cohort, or even to compare model results from different cohorts. For more information, please see [View Multiple Joinpoint Models](#).

Related Content

[Graph Tab](#)

[Data Tab](#)

[Model Estimates Tab](#)

[Trends Tab](#)

[Model Selection Tab](#)

[Comparison Tab](#)

Graph Tab

This tab displays a graph of the output data. It is a scatter plot graph with dots for each data point, and lines between join points.

Possible options from this tab are:

- Change graph settings (Titles, Legend, X-Axis, Y-Axis, Colors, Symbols) - See [Display Options](#).
- Export the graph to a file - See [Exporting Results to Text](#).
- Print the graph - See [Print Options](#).

Related content

- [Data Tab](#)
- [Model Estimates Tab](#)
- [Trends Tab](#)
- [Specify AAPC Ranges](#)
- [Model Selection Tab](#)
- [Comparison Tab](#)

Data Tab

This tab displays the actual data that is graphed on the graph tab. All X and Y values are listed including the modeled Y value and the Joinpoint locations.

Possible options from this tab are:

- Change decimal point settings or whether APCs are shown - See [Display Options](#).
- Export the data to a text file - See [Exporting Results to Text](#).
- Export the data to an Excel file - See [Exporting Results to Excel](#).
- Print the data - See [Print Options](#).

Related content

- [Graph Tab](#)
- [Model Estimates Tab](#)
- [Trends Tab](#)
- [Specify AAPC Ranges](#)
- [Model Selection Tab](#)
- [Comparison Tab](#)

Model Estimates Tab

This tab displays the specifications for each of the models run in the analysis. The final, selected model is first displayed.

Possible options from this tab are:

- Change which statistics are displayed, or to change the number of decimal places - See [Display Options](#).
- Export the data to a text file - See [Exporting Results to Text](#).
- Export the data to an Excel file - See [Exporting Results to Excel](#).
- Print the data - See [Print Options](#).

Related content

- [Graph Tab](#)
- [Data Tab](#)
- [Trends Tab](#)
- [Specify AAPC Ranges](#)
- [Model Selection Tab](#)
- [Comparison Tab](#)

Trends Tab

This tab displays the estimated joinpoints, APCs, and AAPCs for each of the models run in the analysis.

Note: this window is only displayed if the Log Transformation is Yes { $\ln(y) = xb$ }.

Possible options from this tab are:

- Change which statistics are displayed, or to change the number of decimal places - See [Display Options](#).
- Change which AAPCs are displayed - See [Specify AAPC Ranges](#).
- Export the data to a text file - See [Exporting Results to Text](#).
- Export the data to an Excel file - See [Exporting Results to Excel](#).
- Print the data - See [Print Options](#).

Related content

- [Graph Tab](#)
- [Data Tab](#)
- [Model Estimates Tab](#)
- [Specify AAPC Ranges](#)
- [Model Selection Tab](#)
- [Comparison Tab](#)
- [Average Annual Percent Change \(AAPC\) and Confidence Interval](#)

Model Selection Tab

Results for the specified Model Selection Method will be displayed on this tab. The information presented here provides details on how the best, selected model was chosen.

If selected, results for the Minimum APC Difference Worth Detecting (MADWD) will also be displayed on this tab.

Please note that when you are on this tab, the model (# Joinpoints) controls are disabled. This is because the information about all the models are displayed side-by-side on this tab.

Possible options from this tab are:

- Change the number of decimal places displayed - See [Display Options](#).
- Export the data to a text file - See [Exporting Results to Text](#).
- Export the data to an Excel file - See [Exporting Results to Excel](#).
- Print the data - See [Print Options](#).

Related content

- [Graph Tab](#)
- [Data Tab](#)
- [Model Estimates Tab](#)
- [Trends Tab](#)
- [Specify AAPC Ranges](#)
- [Comparison Tab](#)
- [MADWD Minimum APC Difference Worth Detecting in Joinpoint Fitting](#)

Comparison Tab

The Comparison Tab of the Joinpoint Output Window displays the results from the Pairwise Comparison Test.

Please note that the Test for Parallelism or Coincidence is only applicable for the model with the maximum number of joinpoints. The values on this tab will remain the same for the entire cohort, regardless of the specific joinpoint model you have chosen.

Possible options from this tab are:

- Change the number of decimal places displayed - See [Display Options](#).
- Export the data to a text file - See [Exporting Results to Text](#).
- Export the data to an Excel file - See [Exporting Results to Excel](#).
- Print the data - See [Print Options](#).

Related content

- [Model Estimates Tab](#)
- [Trends Tab](#)
- [Specify AAPC Ranges](#)
- [Model Selection Tab](#)
- [Pairwise Comparison](#)
- [Graph Tab](#)
- [Data Tab](#)
- [Parallel Pairwise Comparison and Number of Joinpoints](#) For a statistical test of parallelism between two groups, how is the assumed number of joinpoints (which is displayed on the comparison tab) determined?

Retrieving a Session

The Joinpoint session file is saved as part of the output, so you can select **Output > Retrieve Session** from the top of the main window to retrieve the session that was used to create the output, even if that session file wasn't specifically saved. This is a convenient way to see exactly what settings were used to create a particular output file.

Specify AAPC Ranges

When the **AAPC Confidence Intervals - Parametric** option is selected for the session, the AAPCs can be computed after the Joinpoint calculations are performed, using the APCs. With this option, the AAPCs can be modified in the output without having to rerun the session. When viewing the Joinpoint output, to specify which ranges to display use the **Output > Specify AAPC Ranges...** menu item from the top of the main Joinpoint Output window. On this window you can select to display the AAPC for the entire range of data, the last N observations, or other specified ranges.

When the **AAPC Confidence Intervals - Empirical Quantile** option is selected for the session, the AAPCs cannot be computed after the Joinpoint calculations are performed. The Specify AAPC Ranges window will be unavailable in this case.

Related content

- [Average Annual Percent Change \(AAPC\) and Confidence Interval](#)

Display Options

While displaying the Output Results from a joinpoint run, you may want to configure how the information is displayed. From the Main Menu, select **Output > Options...** to show the Display Options window.

The Display Options window allows you to customize the output that is being displayed in the Output Results window. Each of the tabs on this window correspond to a tab on the Output Results window. Options include:

Graph Tab

- The Titles and Labels for the graph.
- Whether to show the legend and where to place it.
- Show APC Ranges.
- Sizing for the X and Y Axis.
- The colors and symbols to use when displaying the graph.

Data Tab

- Which statistics to display.
- The number of decimal places to show for values.

Model Estimates

- Which statistics to display.
- The number of decimal places to show for values.

Trends

- Which statistics to display.
- The number of decimal places to show for values.

Model Selection

- The number of decimal places to show for values.

Comparison

- The number of decimal places to show for values.

Note, the options you select will be saved with that particular output file (if you save the file after changing the options). If you would like your preferences to always be the default for any new Joinpoint runs, click on **Set as Default** after you have set the options the way you desire. In the future, you can select **Load Defaults** to bring back your saved defaults. On the Graph tab, only the Legend and Colors and Symbols sections are saved as defaults.

Related content

- [Graph Tab](#)
- [Data Tab](#)
- [Model Estimates Tab](#)

- Trends Tab
- Model Selection Tab
- Comparison Tab

Exporting Results to Text

You may access both the graph data and the report information by exporting these as text files for use with other software packages (Excel, SAS, etc.).

Notes:

- Only what is displayed in the Output Window will be exported. For example, if only 1 decimal place is chosen, only that many decimal places will be in the exported file. Or, if the "Regression Coefficients" box is unchecked, those statistics will not be included in the exported file.
- This option will not be available if the "View Multiple Joinpoint Models" option on the output viewer is selected.

To export information (data) from a Joinpoint run, first select **Output > Export > Text...** from the main menu and then, depending on the data you desire, select one or more of the following:

- **Export All** - Choose this option to automatically select all of the options below.
- **Graphs** - Choose this option to export the graphs as an image file (.bmp or .jpg). Each graph is saved as a separate file using the base name and location you specify, plus the specific cohort name for that graph. You do have the option of exporting graphs in high resolution. Please see details below.
- **Data** - Choose this option to export the observed and predicted values for each observation.
- **Model Estimates** - Choose this option to export the model estimates and related information.
- **Jump Model** - This feature will be available in a future version of Joinpoint.
- **APCs** - Choose this option to export the APCs.
- **AAPCs** - Choose this option to export the AAPCs.
- **Model Selection** - Choose this option to export information about the model selection process.
- **Selection Results** - Choose this option to export information about the data dependent selection method process.
- **MADWD** - Choose this option to export information about the minimum apc difference worth detecting results.
- **Comparison** - Choose this option to export information about the Pairwise Comparison test.
- **Final Selected Model** - Choose this option to export the number of Joinpoints in the final selected model for each cohort.

Once you have selected one or more of the above options, you will need to address the following:

- **Output By-variables as** - When exporting By-Variables, Joinpoint can export the numeric representation of the By-variable or the By-variable label (either in quotes or not).
- **Line Delimiter** - Select the appropriate line delimiter for the platform on which you will be working with the exported file.
- **Field Delimiter** - Each variable field in an export line will be separated by a delimiter. Joinpoint offers three types of field delimiters: Tab, Comma, Semicolon.

- **Graphs** - When exporting graphs, you have the option of exporting the currently displayed graph, the graph for the final selected model of each cohort, or the graph for each cohort/model combination. There is also an option to export the graphs in High Resolution. Once the high resolution export option is checked, you have the ability to adjust the number of pixels involved with the width and height of each graph image.
- **Export Models** - Joinpoint can export all of the joinpoint models per cohort or it can export only the final selected joinpoint model.
- **Missing Characters** - Sometimes data values can be missing. Joinpoint offers several export options for missing values. For SAS users, a period is considered a missing value.
- **Remove Flags Marking the Joinpoints** - Whether to include a column with Joinpoint flags.
- **Remove Flags Marking the Best Models** - Whether to include an asterisk indicating which model is the best one.
- **Output All Models in the Same Column** - Whether to include a column indicating the model.
- **Output Variable Names Before Each By-Group** - Redisplay the column headers for each new By-Group.
- **Load Defaults** - This option will load your default settings for this window (your defaults are automatically loaded when the window is displayed). The selected options and their filenames are not included in the default settings.
- **Set As Default** - This option will save all of your settings in this window (excluding the options and filenames). This button is only enabled after you make changes to your settings.

Related content

- [Graph Tab](#)
- [Data Tab](#)
- [Model Estimates Tab](#)
- [Trends Tab](#)
- [Model Selection Tab](#)
- [Comparison Tab](#)

Exporting Results to Excel

You may export your report information to an Excel file. Each section of data will be present in the Excel file as a separate worksheet.

Notes:

- You must have the Excel application installed on your computer for this function to work.
- This option will not be available if the "View Multiple Joinpoint Models" option on the output viewer is selected.

To export information (data) from a Joinpoint run, first select **Output > Export > Excel...** from the main menu and then, depending on the data you desire, select one or more of the following:

- **Filename** - The name of the excel file to create. You can press the Browse button to choose a location for the file.
- **Export All** - Choose this option to automatically select all of the options below.
- **Data** - Choose this option to export the observed and predicted values for each observation.
- **Model Estimates** - Choose this option to export the model estimates and related information.
- **Jump Model** - This feature will be available in a future version of Joinpoint.
- **APCs** - Choose this option to export the APCs.
- **AAPCs** - Choose this option to export the AAPCs.
- **Model Selection** - Choose this option to export information about the model selection process.
- **Comparison** - Choose this option to export data about your Pairwise Comparison Test or Multi-Group Clustering.
- **Selection Results** - Choose this option to export information about the data dependent selection method process.
- **MADWD** - Choose this option to export information about the minimum apc difference worth detecting results.

Once you have selected one or more of the above options, you will need to address the following:

- **Output By-Variables As** - When exporting By-Variables, Joinpoint can export the numeric representation of the By-variable or the By-variable label (either in quotes or not).
- **Export Models** - Joinpoint can export all of the joinpoint models per cohort or it can export only the final selected joinpoint model.
- **Missing Characters** - Sometimes data values can be missing. Joinpoint offers several export options for missing values. For SAS users, a period is considered a missing value.
- **Remove Flags Marking the Joinpoints** - Whether to include a column with Joinpoint flags.
- **Remove Flags Marking the Best Models** - Whether to include an asterisk indicating which model is the best one.
- **Output All Models in the Same Column** - Whether to include a column indicating the model.
- **Output Variable Names Before Each By-Group** - Redisplays the column headers for

each new By-Group.

- **Load Defaults** - This option will load your default settings for this window (your defaults are automatically loaded when the window is displayed). The Filename and Sections to Export are not included in the default settings.
- **Set As Default** - This option will save all of your settings in this window (excluding the Filename and the Sections to Export). This button is only enabled after you make changes to your settings.

Related content

- [Comparison Tab](#)
- [Graph Tab](#)
- [Data Tab](#)
- [Model Estimates Tab](#)
- [Trends Tab](#)
- [Model Selection Tab](#)

Print Options

While viewing the Output Results from a joinpoint run, you may want to print out this information. From the Main Menu, select **Output > Print...** to display the Print Options window. From this window, you can configure what information you want to display in your report.

Please note that this option will not be available if the "**View Multiple Joinpoint Models**" option is selected.

Options to choose from are:

- **Cohort Selection** - Whether to print the currently displayed cohort, or all of the result cohorts.
- **Model Selection** - Whether to print the currently displayed model, the final selected model, or all models. All models is not available for Comparison analyses. Only the final selected model is available when all cohorts are selected.
- **Sections to Print**
 - Most options are related to which output tab information you want to display from the results.
 - **Session Parameters** - Includes all of the session parameters used to generate the output results. The parameters will be printed at the beginning of the report.
- **Graphs** - If you select the "Graph" option from the "Sections to Print", you can choose to output high resolution graphs.

Clicking the **Print** button on this window will display a report viewer window. This window will allow you to view your report before you actually print it. Options from this window are to scroll through the pages, send the report to a printer, setup the print page, or to save the report to an Excel, PDF, or Word file.

Output Properties

Select **Output > Properties** from the main menu to see the Output Properties window. This displays information on the settings used to create the active output file. Information such as the Joinpoint version, the Input Data Filename, By Groups, Creation Time, Date Submitted, Output Filename, and more is displayed.

If you want to re-run the models with different settings, a convenient way is to retrieve the session from the output file.

Related content

- [Retrieving a Session](#)

Viewing Output from the Different Analysis Types

There are a number of different analysis types you can execute in Joinpoint. Within the Output window, each is displayed slightly differently.

Standard Joinpoint Regression Analysis

A standard regression analysis is one where no special options are selected and is the most common type of Joinpoint analysis. Within the Output Window, a standard analysis is displayed in the Cohort Tree as a list of By-Variable combinations where each resultant cohort contains the model results for the joinpoint numbers requested by the user.

Pairwise Comparison Analysis

A Pairwise Comparison is performed by selecting the “Pairwise Comparison” option on the Advanced Analysis Tools tab of the Session window. The results of a Pairwise Comparison analysis will be displayed in a similar fashion to a Standard Analysis except that an additional By-Variable level named “Coincident / Not Coincident” or “Parallel / Not Parallel” will be present. This added level will be above the last By-Variable. Within Coincident or Parallel you can select only one model. However, under Not Coincident or Not Parallel, you can select one model from each of the lowest level By-Variables. For either selection, you will get two graphs and two groups of data displayed.

Details on the Pairwise Comparison Analysis may be found [here](#).

Delay / Non-Delay Data Analysis

A Delay / Non-Delay analysis is performed when “Delay & Non-Delay Data” is selected for the Run Type on the Input File tab of the Session window. In addition to the normal By-Variables, a new lowest level By-Variable is created for Delay and Non-Delay. You can select one model from Delay, and one model from Non-Delay. Both models will be displayed on the graph at the same time, and in each of the data grids.

More information on Delay Adjusted Statistics may be found [here](#).

Jump Model / Comparability Ratio / Standard Joinpoint Analysis

A Jump Model / Comparability Ratio analysis is performed when “Jump Model” or “Comparability Ratio” is selected for the Jump Model / Comparability Ratio on the Advanced Analysis Tools tab of the Session window. On the Output Window, a new lowest level By-Variable will be created for each of your selections. For example, if you selected Jump Model and Include Standard Joinpoint Analysis, each of the joinpoint models would be broken out by Jump Model and Standard Joinpoint Model. You can select one model from each different type to display.

Details on the Jump Model / Comparability Ratio Analysis may be found [here](#).

Delay / Non-Delay with Jump Model / Comparability Ratio / Standard Joinpoint Analysis

The Jump Model / Comparability Ratio / Standard Joinpoint Analysis options can be used with Delay / Non-Delay data. When this occurs, each Delay / Non-Delay cohort data set will have results for each of the selected analysis types (Jump Model, Comparability Ratio, and Standard Joinpoint Analysis). The cohort tree will reflect this.

Viewing Multiple Joinpoint Models

As of Version 4.4.0.0, users can select to consolidate the results of multiple Joinpoint models. This option is particularly useful for viewing multiple cohort model results on one graph.

For most analysis types produced by Joinpoint, the Cohort Tree will provide this option (inside the filter cohorts option accessed at the top of the tree control). The View Multiple Joinpoint Models enables a different display mode in the Output Window. This mode allows you to select as many models as you want to display on the graph and associated output tabs. When this option is selected, the following rules apply:

1. When viewing multiple joinpoint models from the same cohort, all line segments from the same model will have the same color. Each model will be assigned a unique color.
2. If you select model results from multiple cohorts, then all model segment lines from the same cohort will get the same color.
3. You cannot Export or Print when viewing multiple joinpoint models.

Statistical Notes

Let K_{\min} and K_{\max} be the minimum and maximum for the number of joinpoints, respectively. First the program goes through each of the k -joinpoint models, $K_{\min} \leq k \leq K_{\max}$. For each of the models, the program estimates the regression parameters with the smallest sum of squared error (SSE, or smallest weighted SSE). Statistics related to each of the k -joinpoint models are discussed in the selections under **Related Content**. The sequential permutation test procedure to choose the best joinpoint model is discussed in detail in Kim et al. (2000), and only briefly described here.

Statistical Information Pages

[Notation](#)

[Parameterizations](#)

[SSE and MSE](#)

[Degrees of Freedom](#)

[Estimated Joinpoints](#)

[Estimated Regression Coefficients \(Beta\)](#)

[Annual Percent Change \(APC\) and Confidence Interval](#)

[AAPC Comparison](#)

[Significance Level of each Individual Test in a Sequential Testing Procedure](#)

[MADWD - Minimum APC Difference Worth Detecting](#)

Statistics Related to the k-joinpoint Model

Several statistics related to the k-joinpoint model are described here. All statistics come directly from Lerman (1980).

Related content

- [Notation](#)
- [Parameterizations](#)
- [SSE and MSE](#)
- [Estimated Joinpoints](#)
- [Estimated Regression Coefficients \(Beta\)](#)
- [AAPC Comparison](#)

Notation

- n is the total number of data points.
- k is the number of joinpoints in the model.
- p is the total number of parameters in the model, including the joinpoint parameters. For our models $p=2k+2$.
- Q_k is the sum of squared errors (SSE) from the model that minimizes SSE with k joinpoints. If a weighted analysis is done (anything but heteroscedastic errors option= constant variance) then Q_k represents the minimum weighted SSE.
- $Q_{x,j,k}$ is the (weighted) SSE from the model that minimizes (weighted) SSE with k joinpoints and with the j th joinpoint occurring at x .
- $F^{-1}_{a,b}(p)$ is the p th quantile of the F distribution with a and b degrees of freedom.

Parameterizations

The program outputs parameter estimates from two different parameterizations: the "general changepoint" parameterization (GCP), and the "standard" parameterization (SP) of Kim et al. (2000).

The standard parameterization is (see Kim, et al. 2000, equation 1),

$$E[y|x] = \beta_0 + \beta_1 x + \delta_1(x - \tau_1)^+ + \dots + \delta_k(x - \tau_k)^+ \quad (1)$$

where $(a)^+ = a$ if $a > 0$ and 0 otherwise.

The general changepoint parameterization is,

$$E[y|x] = \sum_{j=1}^{k+1} (\beta_{0,j} + \beta_{1,j}x) I(\tau_{j-1} < x \leq \tau_j) \quad (2)$$

where $I(A)$ is the indicator function for $\{A\}$, $\tau_0 = \min(x)$, and $\tau_{k+1} = \max(x)$, and under the constraint that $E[y|x]$ is continuous at τ_j .

For the relationship between the parameterizations see Table 1 and [Appendix A](#).

Table 1: Parameter Transformations for Different Models

| Output Label | Standard | General Changepoint |
|--|-------------------------------|--|
| Intercept 1 | $\beta_{0,1}$ | β_0 |
| Intercept j , $j \geq 2$ | | $\beta_0 - \sum_{h=1}^{j-1} \delta_h \tau_h$ |
| Slope 1 | $\beta_{1,1}$ | β_1 |
| Slope j , $j \geq 2$ | | $\beta_1 + \sum_{h=1}^{j-1} \delta_h$ |
| Slope j - Slope $(j-1)$, $j \geq 2$ | $\beta_{1,j} - \beta_{1,j-1}$ | |

Appendix A

Relationship Between Standard and General Changepoint Parameterizations

Rewrite standard,

$$\begin{aligned}
 E[y|x] &= \beta_0 + \beta_1 x + \delta_1(x - \tau_1)^+ + \dots + \delta_k(x - \tau_k)^+ \\
 &= \beta_0 + \beta_1 x + \sum_{j=1}^k \delta_j(x - \tau_j)I(x > \tau_j) \\
 &= \beta_0 + \beta_1 x + \sum_{j=1}^k \delta_j x I(x > \tau_j) - \sum_{j=1}^k \delta_j \tau_j I(x > \tau_j)
 \end{aligned}$$

Rewrite GCP,

$$\begin{aligned}
 E[y|x] &= \sum_{j=1}^{k+1} (\beta_{0j} + \beta_{1j}x) \{I(x > \tau_{j-1}) - I(x > \tau_j)\} \\
 &= \sum_{j=1}^{k+1} (\beta_{0j} + \beta_{1j}x)I(x > \tau_{j-1}) - \sum_{j=1}^{k+1} (\beta_{0j} + \beta_{1j}x)I(x > \tau_j) \\
 &= \sum_{j=0}^k (\beta_{0,j+1} + \beta_{1,j+1}x)I(x > \tau_j) - \sum_{j=1}^k (\beta_{0j} + \beta_{1j}x)I(x > \tau_j) \\
 &= \beta_{01} + \beta_{11}x + \sum_{j=1}^k I(x > \tau_j) \{(\beta_{0,j+1} - \beta_{0j}) + (\beta_{1,j+1} - \beta_{1j})x\}
 \end{aligned}$$

Standard General Changepoint Parameterizations

$$\beta = \beta_{01}$$

$$\beta = \beta_{11}$$

$$\delta_j = \beta_{1,j+1} - \beta_{1j}$$

$$T_j = -\frac{\beta_{0,j+1} - \beta_{0j}}{\beta_{1,j+1} - \beta_{1j}}$$

SSE and MSE

Sum of squared errors (SSE) is actually the weighted sum of squared errors if the heteroscedastic errors option is not equal to constant variance. The mean squared error (MSE) is the SSE divided by the degrees of freedom for the errors for the constrained model, which is $n-2(k+1)$.

The minimum SSE for a k-joinpoint model is calculated using Lerman's grid-search method (1980) based on Kim et al's standard parametrization ([Equation 1](#)). The corresponding values for (τ_1, \dots, τ_k) and $(\beta_0, \beta_1, \delta_1, \dots, \delta_k)$ are the estimates of joinpoints and regression coefficients, respectively.

Degrees of Freedom

When the Grid Search Method is used, the degree of freedom for the estimated regression coefficient is $d=n^*-2(k+1)$, where k is the number of joinpoints and n^* is the effective number of data points after deleting offending observations, data points that are on joinpoints.

When Hudson's Method is used for a continuous fitting, the degree of freedom is adjusted in a continuous manner instead of subtracting the offending observations. The idea is to delete, with higher probabilities, the observations whose x -values are closer to joinpoints, while no more than one observation is deleted around each joinpoint. See Kim et al. (Kim, H.-J., Yu, B. and Feuer, E.J. (2008) "Inference in segmented line regression: A simulation study", *Journal of Statistical Computation and Simulation* 78(11), 1087-1103) for details.

Estimated Joinpoints

For information see [Estimated Joinpoint \(Estimated Tau\) and Confidence Interval](#).

Estimated Regression Coefficients (Beta)

The output is a combination of the two parameterizations (see Table 1). The estimates of $(\beta_0, \beta_1, \delta_1, \dots, \delta_k)$ come from the grid-search method. The estimates of $(\beta_{01}, \beta_{11}, \dots, \beta_{0,k+1}, \beta_{1,k+1})$ are calculated based on Table 1.

However, the standard errors of the regression coefficients are estimated under the GP model (Equation (2)) without continuity constraints. Following Lerman's implementation, (Lerman; 3rd paragraph, page 79, 1980; Feder, Section 4, page 69, 1975), the data points that are on the joinpoints are deleted. Then conditioned on the partition implied by the estimated joinpoints (τ_1, \dots, τ_k) , the standard errors of $(\beta_{01}, \beta_{11}, \dots, \beta_{0,k+1}, \beta_{1,k+1})$ are calculated using unconstrained least square for each segment. If there are segments with zero or one observation (not including the joinpoints at both ends), then a generalized inverse is used to calculate the covariance matrix. The standard error of the difference in slopes, δ_j , is the square root of the sum of the squared standard errors (variance) for the two consecutive slopes β_{1j} and $\beta_{1,j+1}$.

The test statistic (U) is the parameter estimate divided by the standard error. The test statistic has a t distribution with d degrees of freedom where d is defined as follows. Let n_j be the number of data points that are on joinpoints. The effective number of data points is $n^* = n - n_j$. Let k_0 and k_1 be the number of segments with zero or one observation. The effective number of parameters (rank of the design matrix) is $p^* = 2(k+1) - 2k_0 - k_1$ and the degrees of freedom $d = n^* - p^*$. For the default option where the minimum number of data points between two joinpoints (excluding any joinpoint that falls on an observation) is two, $k_0 = k_1 = 0$ and $d = n^* - 2(k+1)$. For testing $H_0: \beta = 0$ the p-value is calculated as $2\{1 - t_d(|U|)\}$, where t_d is a t distribution with d degrees of freedom.

If a standard error cannot be calculated, then the associated statistics will not be displayed.

Related content

- [Data Errors That Stop An Analysis](#)

AAPC Comparison

When the test for pairwise comparison is selected, and the two groups are not parallel or coincident, the AAPCs for the two groups are tested to determine if they are statistically different. For details, see Average Annual Percent Change (AAPC).

Significance Level of each Individual Test in a Sequential Testing Procedure

The Joinpoint program uses a sequence of "permutation" tests to select the final model. Each one tests the null hypothesis $H_0: k = k_a$ against the alternative hypothesis $H_1: k = k_b$. The procedure begins with $k_a = K_{\min}$ and $k_b = K_{\max}$. If the null is rejected, then increase k_a by 1; otherwise, decrease k_b by 1. Let $\hat{k} = k_a = k_b$ be the final selected number of Joinpoints.

Because multiple tests are performed, Bonferroni adjustment is used to ensure that the approximate overall type I error is less than the specified significance level (significance level is also called the α -level, default $\alpha = .05$). Each of these permutation test are carried out a significance level of $\alpha_1 = \alpha / (K_{\max} - K_{\min})$, i.e., if the p-value $< \alpha_1$, then it rejects the null.

The Bonferroni adjustment is conservative because the actual overall significance level is usually less than the nominal level α . The new adjustment procedure controls the overall over-fitting probabilities.

$$P(\hat{k} > k_a | k = k_a), \quad k_a = K_{\min}, \dots, K_{\max} - 1$$

Let $\alpha(k_a, k_b)$ be the significance level of each individual test $H_0: k = k_a$ vs. $H_1: k = k_b$. The new procedure set $\alpha(k_a, k_b) = \alpha / (K_{\max} - k_a)$. Notice that the individual significance level depends on the number of joinpoints k_a under the null. Consider an example where $K_{\min} = 0$ and $K_{\max} = 4$. The new procedure has the following properties:

$$P(\hat{k} > 0 | k = 0) \leq \alpha(0, 4) + \alpha(0, 3) + \alpha(0, 2) + \alpha(0, 1);$$

$$P(\hat{k} > 1 | k = 1) \leq \alpha(1, 4) + \alpha(1, 3) + \alpha(1, 2);$$

$$P(\hat{k} > 2 | k = 2) \leq \alpha(2, 4) + \alpha(2, 3);$$

$$P(\hat{k} > 3 | k = 3) \leq \alpha(3, 4).$$

If we like to bound these over-fitting probabilities by α , then we can assign different values for each $\alpha(k_a, k_b)$. That means, we can achieve a better power by setting

$$\alpha(0, 4) = \alpha(0, 3) = \alpha(0, 2) = \alpha(0, 1) = \alpha/4$$

$$\alpha(1, 4) = \alpha(1, 3) = \alpha(1, 2) = \alpha/3$$

$$\alpha(2, 4) = \alpha(2, 3) = \alpha/2$$

$$\alpha(3, 4) = \alpha$$

Technical Requirements/Help

[Citation](#)

[Technical Support](#)

[System Requirements](#)

[Frequently Asked Questions](#)

FAQ's

[Joinpoint Revision History](#)

What are the most recent changes to the Joinpoint program?

[Alpha and Beta Versions](#)

Features in Joinpoint which are labeled Alpha or Beta versions

Citation

The joinpoint regression model and permutation tests for identifying changes in trend are described in:

Methods Citation:

Kim HJ, Fay MP, Feuer EJ, Midthune DN. "Permutation tests for joinpoint regression with applications to cancer rates" *Statistics in Medicine* 2000; 19:335-351: (correction: 2001;20:655).

The present document describes the use of the Joinpoint Regression Program for Windows.

Software Citation:

Joinpoint Regression Program, Version 4.8.0.1 - April 2020; Statistical Methodology and Applications Branch, Surveillance Research Program, National Cancer Institute.

Technical Support

For a list of common questions, please check the [Frequently Asked Questions \(FAQs\)](#) list.

For technical support beyond the help system or Web site, please use the [On-line Technical Support Request Form](#).

Note: if you contact Joinpoint technical support for help with a specific analysis, it helps if you can attach a copy of your data file and Joinpoint session (.jps) or output (.jpo) file.

System Requirements

The system requirements to run Joinpoint are a personal computer with at least:

- Pentium-based PC
- 32 MB application RAM
- Approximately 10 MBs hard disk space
- A 32-bit or 64-bit Microsoft Windows Operating System - Windows 8.1 or higher
- Screen resolution set to 800 by 600 pixels or greater
- .NET Framework 4.0 or higher (this will be installed with Joinpoint if necessary)
- The use of Excel features (Excel as an input data file, exporting to Excel) requires Office 2013 or later to be installed

Frequently Asked Questions

- For more help using Joinpoint check the Surveillance Research Program Web site for Joinpoint: Frequently Asked Questions.

Citation & References

1. Is there a suggested citation for Joinpoint?
2. Are there any reference papers that can help me understand the methodology?

Software Updates and Installation

3. What are the most recent changes to the Joinpoint program?
4. How can I replicate previous results with a newer version of Joinpoint?
5. Can Joinpoint run on a Macintosh PC?

Model Estimation and Selection Methods

6. Why doesn't the Joinpoint program give me the best possible fit? I can see other models with more joinpoints that would fit better. Exactly how does the program decide which tests to perform and which joinpoint model is the final model?
7. Describe the permutation test used here.
 - a. How many permuted data sets should I use?
 - b. What does the p -value mean for Joinpoint?
 - c. Why are new significance levels being used in Version 3.0 and succeeding releases?
8. What is the Bayesian Information Criterion method for selecting the best model?
9. Why should I use the Permutation Test (PT) when Bayesian Information Criterion (BIC) is so much faster?
10. Should I use the linear or log-linear model?
11. When running the test for pairwise differences with the Coincident option, should I use a weighted or unweighted model?
12. Does the test of parallelism of two series require the series to be uncorrelated?
13. For a statistical test of parallelism between two groups, how is the assumed number of joinpoints (which is displayed on the comparison tab) determined?
14. Can the Joinpoint software estimate the autocorrelation parameter?
15. Why did the default settings for Grid Search change?
16. Why is the Hudson's method disabled?
17. What is Tau?
18. Why would I choose Data Dependent Selection as a Model Selection Method?
19. What happened to the Clustering controls on the Advanced Analysis Tools Tab?
20. Why does the number of joinpoints for the best fit sometimes decrease when the maximum number of joinpoints is increased?

APC and AAPC

21. What is an APC?
22. What is an AAPC?
23. How is the AAPC Computed?
24. Can we compare AAPCs for two independent groups?
25. What are the relative advantages and disadvantages of reporting an AAPC over APCs?
26. What is the advantage of reporting an AAPC over an APC computed by fitting a single line (on a log scale) to the data?
27. Which Method is better to use for the Empirical Quantile CI calculations?
28. Why is the Empirical Quantile Confidence Interval option labeled as "Alpha"?
29. When I fit a Joinpoint model to find the rate trend, why doesn't the APC for one of the segments match the APC found by SEER*Stat for the corresponding segment?
30. Sometimes, the APC for one segment is significantly different from zero, but when an extra joinpoint in the segment is determined by the Joinpoint software, neither APCs for the two consecutive segments are significant. Why?

Applications of Joinpoint

31. I used the program on SEER data and do not get the same answers as in the SEER Cancer Statistics Review. Why?
32. Sometimes, the change in the slopes between two segments is not statistically significant (p -value >0.05) from the t -test, but there is joinpoint between two segments or vice versa. Why?
33. Can I do Joinpoint regressions where the time intervals are not equally spaced or the time points are not integer?
34. Can Joinpoint be used to model proportions or percents?
35. How does the Joinpoint software deal with zeros in the dependent variable?
36. Can I use Joinpoint to project or extrapolate into the future?
37. What type of SEER*Stat sessions are compatible?
38. How do you handle an independent variable that is not yearly data, but is continuous data and is presented in discrete categories, or is ordinal data?

Using Joinpoint Results

39. How do I import Joinpoint's graphs into other software such as Word, Excel, or PowerPoint?
40. How do I print Joinpoint results?
41. How do I save Joinpoint results to PDF or Word?
42. How do I export Joinpoint results to a text file or to Excel?
43. How do I export high resolution graphs?

Data Sources

44. When I use Excel as my data source, why do I get "####" in the data?

Citations

Is there a suggested citation for Joinpoint?

Answer:

Both a methods and software citation can be found in the Citation section [here](#).

References

Are there any reference papers that can help me understand the methodology?

Answer:

Reference papers are listed [here](#).

Joinpoint Revision History

What are the most recent changes to the Joinpoint program?

Answer:

An outline of the most recent changes to the Joinpoint program can be found [here](#).

Replicating Previous Results

How can I replicate previous results with a newer version of Joinpoint?

Answer:

Because of changes in statistical methodology, computational issues, and the random number generator from one version to the next, the same version of the program should be used again to exactly replicate results. If you would like prior versions of Joinpoint to replicate previous results, please request them by contacting [Joinpoint Technical Support](#).

Mac Installation

Can Joinpoint run on a Macintosh PC?

Answer:

Joinpoint was designed to run under a Microsoft Windows operating system. Some users have reported successfully running Joinpoint on a Macintosh using a windows emulator, but we do not support that operating system.

Related content

- [System Requirements](#)

Selecting the Final Model

Why doesn't the joinpoint program give me the best possible fit? I can see other models with more joinpoints that would fit better. Exactly how does the program decide which tests to perform and which joinpoint model is the final model?

Answer:

As with many statistical models, if you add more parameters you get a better fit. The same is true with joinpoint models. What the program does is to try to choose the smallest number of joinpoints such that if one more joinpoint is added, the improvement is not statistically significant. Thus, in the final model you may interpret each of the joinpoints and its corresponding changes in trend as significant.

Joinpoint selects the final model using two different methods: Permutation Test and Bayesian Information Criterion (BIC). First, the user specifies MIN as the minimum number of joinpoints and MAX as the maximum number of joinpoints on the **Method and Parameters** tab.

Then the program uses a sequence of permutation tests to select the final model. Each one of the permutation tests performs a test of the null hypothesis H_0 : number of joinpoints = k_a against the alternative H_a : number of joinpoints = k_b where $k_a < k_b$. The procedure begins with $k_a = \text{MIN}$ and $k_b = \text{MAX}$. If the null is rejected, then increase k_a by 1; otherwise, decrease k_b by 1. The procedure continues until $k_a = k_b$ and the final value of $\hat{k} = k_a = k_b$ is the selected number of joinpoints.

The second method is based on the Bayesian Information Criterion (BIC). The value of BIC is the loglikelihood value penalized by the cost of extra parameters. The model with the minimum value of BIC is selected as the optimal model.

Related content

- [Method and Parameters Tab](#)
- [How Joinpoint Selects the Final Model](#)

Permutation Test Details

Describe the permutation test used here.

Answer:

In this program, the permutation test is used repeatedly for testing between two different joinpoint models, a simpler model with fewer joinpoints called the null model and a more complicated model called the alternative model. The alternative model fits better because it is more complicated.

The question for the test is: does it fit much better than would be expected by chance. To test this statistically, we calculate a ratio, $T = \text{SSEN} / \text{SSEA}$, where SSEN is the sum of squared errors (SSE) from the null model and SSEA is the SSE from the alternative model. Values of the ratio (T) close to 1 mean that the alternative is not much better than the null model, while larger values mean that the alternative is much better.

In order to decide how much larger a ratio needs to be to be statistically significant, we use the permutation method, since it is not possible to derive this analytically. In this method, we randomly permute (that is, shuffle) the errors (also called the residuals) from the null model and add them back onto the modeled values from the null model to create a permutation data set. Then we calculate the ratio T for the permutation data set and measure how much evidence the data provide against the null hypothesis by estimating the proportion of the permutation data sets whose T values are at least as extreme as the one we observed with the original data set. This proportion is called the *p*-value in statistical testing.

If the true model was the null model (i.e. the truth being the simpler model), we would expect that on average, about half of the T ratios computed from the permutation data sets would be greater than the one derived from the original data (i.e. the distribution of the *p*-value under the null is uniform over the interval from 0 to 1).

If the true model was the alternative model, we would expect that after permuting the errors most of the new T ratios would be less than the original T ratio and thus the *p*-value would be small. In other words, the permuted data set would look less like the alternative model than the original data.

So we reject the null model (or null hypothesis) if less than a certain proportion of the T ratios are greater than or equal to the original T ratio. That is, we reject the null hypothesis if the *p*-value is small.

For more specific details, see Kim HJ, Fay MP, Feuer EJ, Midthune DN. Permutation Tests for Joinpoint Regression with Applications to Cancer Rates. *Stat Med* 2000;19:335-351. To request a reprint, email Mr. Reggie Taborn for a copy at: tabornr@mail.nih.gov

Related content

- [Permutation Test](#)
- [How Joinpoint Conducts Permutation Testing](#)

Number of Permuted Data Sets

How many permuted data sets should I use?

Answer:

For greater consistency in the p -values obtained if one were to change the seed for each run, we suggest running the program for at least 4499 permutations. For this reason, the default number of permutations is now 4499 in the current version of Joinpoint (it was 999 in previous versions).

If the maximum number of possible joinpoints is set to 3 and the minimum is set to 0, then there are 3 statistical tests each conducted at the Bonferonni adjusted cutoff significance level of $.05/3 = .0167$. The value 4499 was chosen so that if you obtained a p -value of $.0167$ using one seed with 4499 permutations, then, assuming the number of possible permutations is large, the complete run using all possible permutations would have approximately a 99% chance of the p -value being between $.0120$ and $.0220$ (length of confidence interval = $.0100$), and approximately a 95% chance of the p -value being between $.0129$ and $.0206$ (length of confidence interval = $.0077$). Choice of the number of permutations selected by the user is a tradeoff between computer time and consistency of the p -values obtained.

The Joinpoint program uses Monte Carlo simulation to calculate p -values for a series of permutation tests. See [Permutation Test Details](#) for details on the permutation tests performed.

Here we discuss the implications of the choice of the number of permutation data sets (N). The program runs faster with smaller values of N , but it gives better precision for the p -value with larger values of N . In addition, a larger N reduces the probability that another analysis of the same data might get a different answer when run with different random number generator seeds. (Computer programs produce pseudo-random numbers through algorithms that mimic randomness, which we use to shuffle or permute the errors. The algorithms use a seed or seeds to start the algorithm. These seeds can be used to produce repeatable pseudo-random numbers.)

The problem of two analyses obtaining different answers from the same data is addressed by this program by specifying default random number generator seeds. Thus, as long as no parameters are changed (including the random number generator seed and N), repeats of the analyses will produce the same results. Otherwise, two runs of the same analysis except with different seeds could get different answers.

To get an idea how results would change for someone using different random number generator seeds, we list some confidence intervals for p -values below.

N=99

| lower 99% ci | p-value | upper 99% ci |
|--------------|---------|--------------|
| 0.0000 | 0.01 | 0.0521 |
| 0.0034 | 0.04 | 0.1065 |
| 0.0069 | 0.05 | 0.1218 |
| 0.0111 | 0.06 | 0.1364 |
| 0.0325 | 0.10 | 0.1910 |
| 0.2702 | 0.40 | 0.5281 |

N=999

| lower 99% ci | p-value | upper 99% ci |
|--------------|---------|--------------|
| 0.0031 | 0.01 | 0.0199 |
| 0.0250 | 0.04 | 0.0577 |
| 0.0331 | 0.05 | 0.0694 |
| 0.0415 | 0.06 | 0.0810 |
| 0.0762 | 0.10 | 0.1259 |
| 0.3595 | 0.40 | 0.4402 |

N=9999

| lower 99% ci | p-value | upper 99% ci |
|--------------|---------|--------------|
| 0.0075 | 0.01 | 0.0127 |
| 0.0350 | 0.04 | 0.0452 |
| 0.0445 | 0.05 | 0.0558 |
| 0.0540 | 0.06 | 0.0663 |
| 0.0923 | 0.10 | 0.1079 |
| 0.3873 | 0.40 | 0.4127 |

N=99999

| lower 99% ci | p-value | upper 99% ci |
|--------------|---------|--------------|
| 0.0092 | 0.01 | 0.0108 |
| 0.0384 | 0.04 | 0.0416 |
| 0.0482 | 0.05 | 0.0518 |
| 0.0581 | 0.06 | 0.0620 |
| 0.0976 | 0.10 | 0.1025 |
| 0.3960 | 0.40 | 0.4040 |

Related content

- [Permutation Test](#)
- [Joinpoint p-values](#)

p-values

What does the p-value mean for Joinpoint?

Answer:

The Joinpoint Regression Program performs a series of hypothesis tests that test the null hypothesis of k_a joinpoints against the alternative hypothesis of k_b joinpoints, where k_a and k_b change for each hypothesis test ($k_a < k_b$). Each p -value corresponds to this type of test. The p -value is an estimate of the probability of observing data at least as extreme as the data that we have, in fact, observed under the assumption that the null hypothesis (i.e. smaller number of joinpoints) is true.

As the permutation test is a randomization test, it depends on the random number generator. For greater consistency in the permutation test p -values obtained if one were to change the seed for each run, we suggest running the program for at least 4499 permutations. For this reason, the default number of permutations is now 4499 in the current version of the Joinpoint Regression Program. Choice of the number of permutations selected by the user is a tradeoff between computer time and consistency of the p -values obtained.

Related content

- [Permutation Test](#)
- [Joinpoint p-values](#)

New Significance Levels

Why are new significance levels being used in Version 3.0 and succeeding releases?

Answer:

The Joinpoint software uses a series of permutation tests to determine the number of joinpoints. Prior to Version 3.0, the software used the Bonferroni adjustment to control the error probability of each of the multiple tests. The Bonferroni adjustment has been shown to be conservative and the procedure tends to select fewer joinpoints than it should. The procedure with new significance levels controls the over-fitting probability and it is superior to the traditional Bonferroni adjustment.

Details: The Bonferroni adjustment is conservative because the actual overall significance level is usually less than the nominal level α . Starting with Version 3.0, the new adjustment procedure controls the overall over-fitting error probabilities, $P(\hat{k} > k_a | k = k_a)$, $k_a = MIN, \dots, MAX-1$, under α .

Let k denote the number of joinpoints and $\alpha(k_a; k_b)$ be the significance level of each individual test $H_0: k = k_a$ vs. $H_a: k = k_b$.

The new procedure set $\alpha(k_a; k_b) = \alpha / (MAX - k_a)$.

Notice that the individual significance level depends on the number of joinpoints k_a under the null. Consider an example where $MIN = 0$ and $MAX = 4$. The new procedure has the following properties:

$$P(\hat{k} > 0 | k = 0) \leq \alpha(0,4) + \alpha(0,3) + \alpha(0,2) + \alpha(0,1);$$

$$P(\hat{k} > 1 | k = 1) \leq \alpha(1,4) + \alpha(1,3) + \alpha(1,2);$$

$$P(\hat{k} > 2 | k = 2) \leq \alpha(2,4) + \alpha(2,3);$$

$$P(\hat{k} > 3 | k = 3) \leq \alpha(3,4).$$

If we like to bound these over-fitting probabilities by α , then we can assign different values for each $\alpha(k_a; k_b)$. That means, we can achieve a better power by setting

$$\alpha(0,4) = \alpha(0,3) = \alpha(0,2) = \alpha(0,1) = \alpha / 4;$$

$$\alpha(1,4) = \alpha(1,3) = \alpha(1,2) = \alpha / 3;$$

$$\alpha(2,4) = \alpha(2,3) = \alpha / 2;$$

$$\alpha(3,4) = \alpha$$

Related content

- [Permutation Test](#)

Bayesian Information Criterion (BIC) Details

What is the Bayesian Information Criterion method for selecting the best model?

Answer:

Permutation test (PT) and BIC are two different approaches to selecting the optimal number of joinpoints. The Permutation test approach uses a sequence of permutation tests to determine the best number of joinpoints. The PT approach controls the error probability of selecting the wrong model at a certain level (i.e. 0.05), whereas the BIC approach finds the model with the best fit by penalizing the cost of extra parameters. The models picked by BIC tend to fit the data well but are less parsimonious. The applications have shown that the PT approach worked well for cancer incidence and mortality data.

The equation for computing the BIC for a k-joinpoint model is:

$$BIC(k) = \ln\{SSE(k)/\#Obs\} + \{\#Parm(k) / \#Obs\} * \ln(\#Obs),$$

where SSE(k) is the sum of squared errors of the k-joinpoint regression model, #Parm(k)=2*(k+1) is the number of parameters of the k-joinpoint model and #Obs is the number of observations.

The k-joinpoint model with the minimum value of BIC(k) is selected as the final model.

Related content

- [Bayesian Information Criterion \(BIC\)](#)

Permutation Test vs. Bayesian Information Criterion

Why should I use the Permutation Test (PT) when Bayesian Information Criterion (BIC) is so much faster?

Answer:

PT and BIC are two different approaches to selecting the optimal number of joinpoints.

The Permutation test approach uses a sequence of permutation tests to determine the best number of joinpoints. The PT approach controls the error probability of selecting the wrong model at a certain level (i.e. 0.05), whereas the BIC approach finds the model with the best fit by penalizing the cost of extra parameters.

The models picked by BIC tend to fit the data well but are less parsimonious. The applications have shown that the PT approach worked well for cancer incidence and mortality data.

Related content

- [Permutation Test](#)
- [Bayesian Information Criterion \(BIC\)](#)

Linear or Log-linear Model

Should I use the linear or log-linear model?

Answer:

The linear or log-linear model can be chosen depending on how linear the observed rates or the logarithm of the observed rates are over time. In order to check the goodness of fit of the chosen model, a user can test for normality of the residuals obtained under the linear or the log-linear fit. Select a model whose residual analysis indicates a better fit, regarding the model assumptions of normality, linearity, equal variance, and independence. One reason for using a log transformation for cancer rates is that they arise from a Poisson distribution which is skewed especially when the cancer is rare or the rates come from a small population. The log transformation is a standard way to make this skewed distribution approximately a normal distribution. Rates for common cancers or which come from a large population can be approximated as arising from a normal distribution without a transformation.

One motivation for using the log-linear model for cancer rates regardless if they are rare or not is the ease of interpretation. Under a log-linear model the rates change at a constant percent per year (i.e. a fixed annual percent change - APC), while for a linear model the rates change at a constant fixed amount per year. When comparing trends across age groups or across cancer sites where the rates are very different, the advantage of a log-linear model is that the APC is a metric which makes sense to compare across widely different scales. For example, a rare cancer and a common cancer may change at the same annual percent per year, but it is highly unlikely that they would change at the same fixed amount per year (e.g. if the rates were declining, the rare cancer rate would quickly become negative!).

Related content

- [Log Transformation](#)

Pairwise Differences with Coincident Option

When running the test for pairwise differences with the Coincident option, should I use a weighted or unweighted model?

Answer:

If the weights for the two cohorts are fairly different, the combined model will be heavily influenced by the cohort with the larger weight, and the results may not be as expected since the combined fit will closely mimic the level and the number and location of joinpoints for that individual cohort. In this case the statistical algorithm is appropriately weighting the series that is more reliable. In some cases, however, it may be more appropriate to weight each series equally, even though one series has a larger variance. For example, it may be appropriate to weight two racial/ethnic groups equally if the goal is to find the best fit for two groups ignoring the fact that one series is more reliable than the other. In such a case, running the unweighted model using the Heteroscedastic Errors Option of "Constant Variance" may be more appropriate. These same considerations are less obvious but can be relevant when fitting a parallel model. In this case, level is not an issue (since each cohort has its own level), but the fit of the number and location of joinpoints will be heavily influenced by the larger cohort, unless an unweighted analysis is used. There is no "correct" answer, but careful consideration should be given to the overall purpose of the analysis in these situations.

For more details, see [Pairwise Comparison](#) in the Joinpoint help system.

Related content

- [Pairwise Comparison](#)
- [Parallel Pairwise Comparison and Number of Joinpoints](#) For a statistical test of parallelism between two groups, how is the assumed number of joinpoints (which is displayed on the comparison tab) determined?

Correlation In Two Series

Does the test of parallelism of two series require the series to be uncorrelated?

Answer:

The permutation test for parallelism assumes exchangeability in the pair of residuals at each time period. When errors of the two series are correlated, this assumption is violated, so the permutation p -value will not be accurate. Errors in series are correlated, for example, when the two series come from the same population or sub-population measured from the same survey (e.g. obesity and diabetes for white males from the same survey) or the same cancer registry (e.g. white male colorectal and prostate cancer rates from the same registry). Errors in series may be considered uncorrelated when the observations in the two series come from different samples of the same population (e.g. obesity for white males measured from one national survey and diabetes for white males measured from a different national survey).

One way to test for parallelism in two series with correlated errors is to compute a new time series, the difference between the original two series. The series are parallel if the difference time series has a constant mean. One way this can be checked is by fitting a simple linear regression and testing whether the slope coefficient is zero.

Related content

- [Permutation Test](#)
- [Parallel Pairwise Comparison and Number of Joinpoints](#) For a statistical test of parallelism between two groups, how is the assumed number of joinpoints (which is displayed on the comparison tab) determined?

Parallel Pairwise Comparison and Number of Joinpoints

For a statistical test of parallelism between two groups, how is the assumed number of joinpoints (which is displayed on the comparison tab) determined?

Answer:

When conducting the test of parallelism between two groups, the true number of joinpoints for each group under the null hypothesis (H_0) of parallelism and alternative hypothesis (H_A) lack of parallelism are unknown. As described in Kim, et al. (Biometrics, 2004) the test of parallelism is conducted under K_{\max} joinpoints, where K_{\max} is greater than or equal to the number joinpoints for both groups under both H_0 and H_A . Simulations presented in Kim et al. have shown that a reasonable choice for K_{\max} is the maximum of the estimated number of joinpoints for each group fit separately, and for the two groups fit together under the assumption of parallelism. For example, if the Joinpoint software selects 2 joinpoints for males and 0 joinpoints for females fit separately, and 1 joinpoint for males and females fit under an assumption of parallelism, then the parallelism test is conducted under an assumption of $\text{Max}(2,0,1) = 2$ joinpoints.

Kim, H-J, Fay, M.P., Yu, Binbing, Barrett, M.J., and Feuer, E.J. (2004) "Comparability of segmented line regression models", Biometrics 60, 1005-1014

Example:

The Comparison tab shows "2 Joinpoints", when the cohort was deemed "parallel" and the parallel best fit was 1 joinpoint.

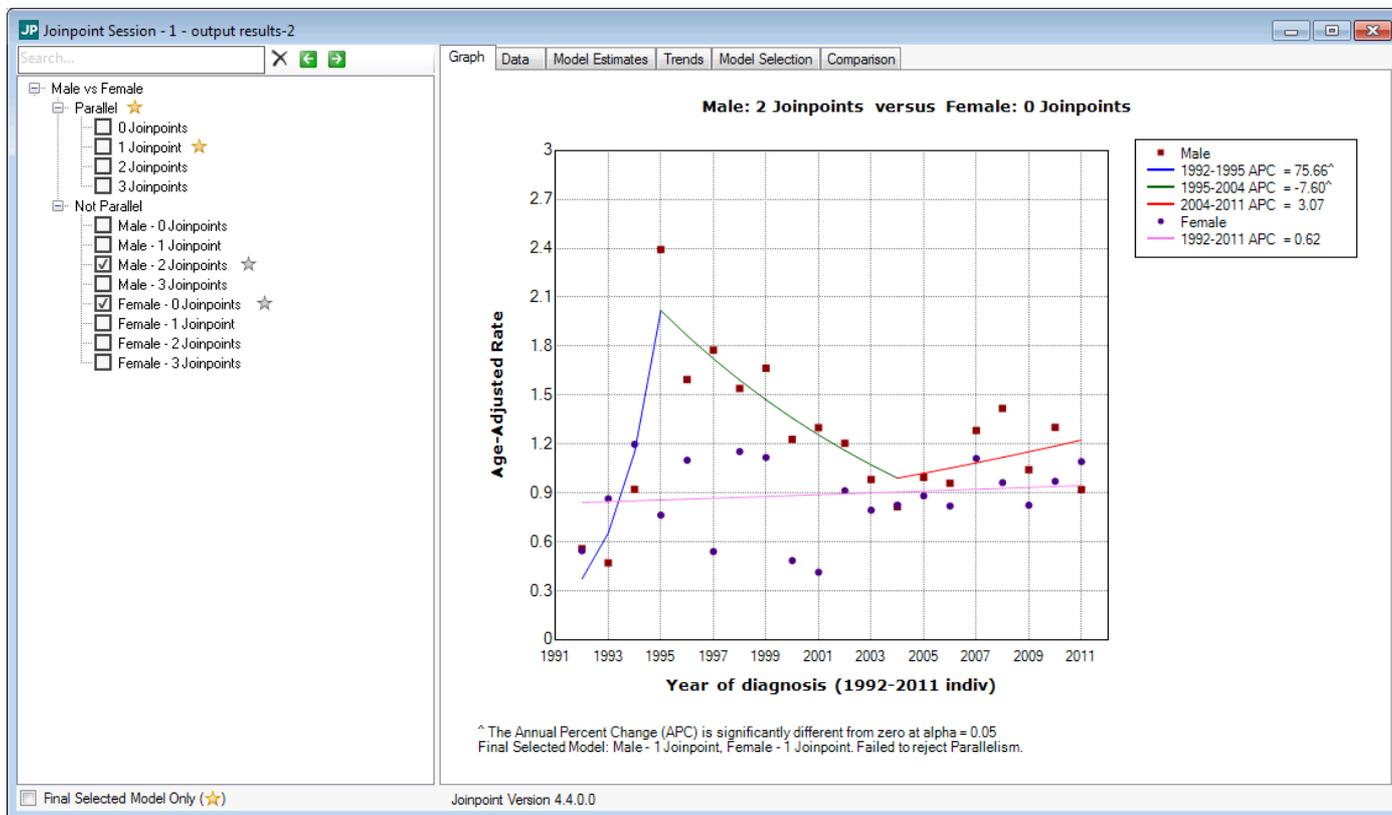
For the Pairwise comparison, you have the following three best-fit models:

- Male - Joinpoint selects 2 Joinpoints.
- Female - Joinpoint selects 0 Joinpoints.
- Combined Male and Female - Joinpoint selects 1 Joinpoint.

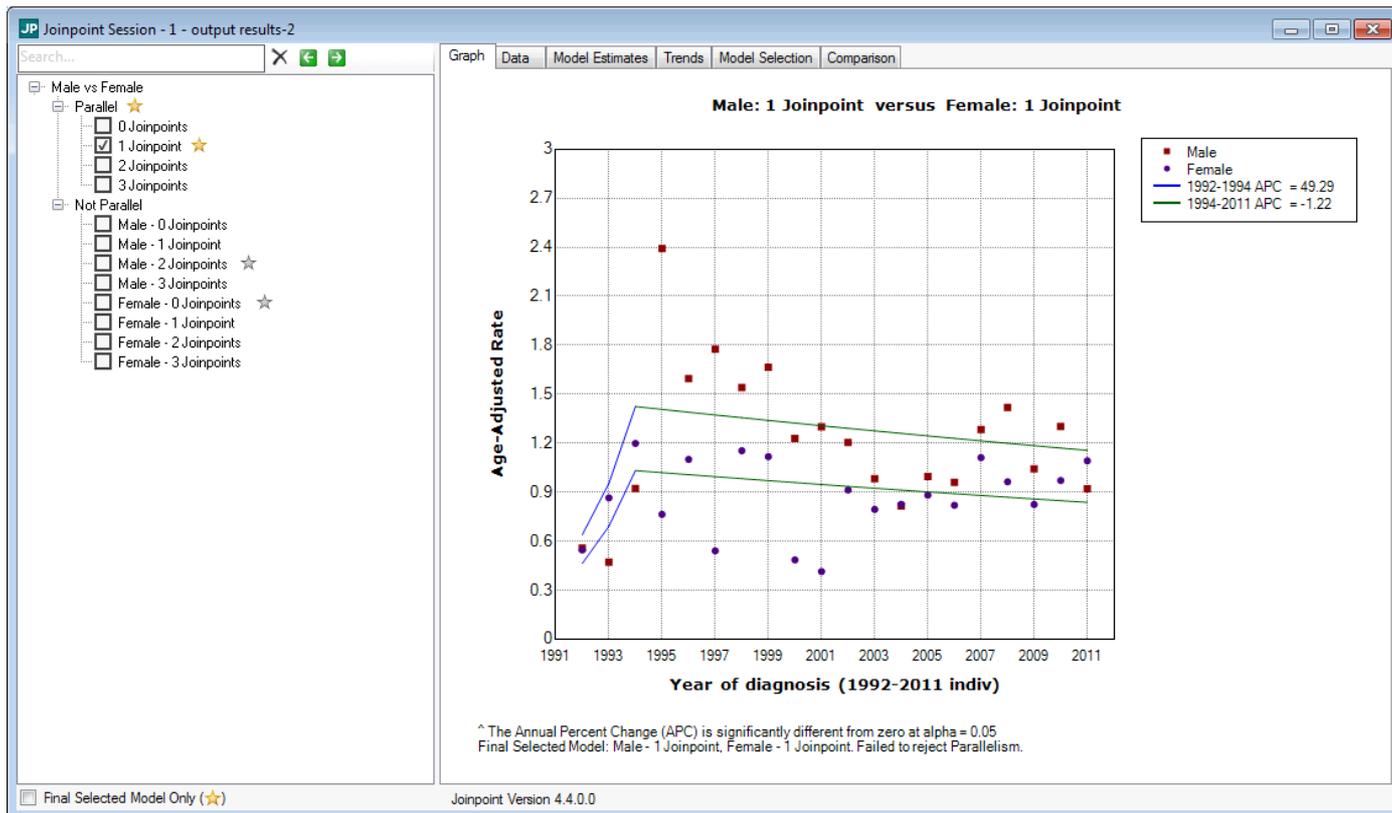
In the Pairwise Comparison help, there is a section labeled "How is the Kmax chosen?" That section describes the process of how Joinpoint determines the "2 Joinpoints" on the Comparison tab. This selection is based on the 3 models above. The number shown on the comparison tab ("2 Joinpoints") is the maximum number of Joinpoints in the above 3 models. So, it is the max of 2, 0, and 1.

The test for parallelism is performed AFTER the three models are fit. So, the number of joinpoints displayed on the comparison tab is not necessarily the number of joinpoints selected for the model fit for the combined data. Below are the screen shots of the three models and the test for parallelism.

Male - Joinpoint selects 2 Joinpoints, Female - Joinpoint selects 0 Joinpoints:



Combined Male and Female - Joinpoint selects 1 Joinpoint:



Test for Parallelism:

JP Joinpoint Session - 1 - output results-2

Search...

Graph | Data | Model Estimates | Trends | Model Selection | Comparison

Male vs Female

- Parallel ☆
 - 0 Joinpoints
 - 1 Joinpoint ☆
 - 2 Joinpoints
 - 3 Joinpoints
- Not Parallel
 - Male - 0 Joinpoints
 - Male - 1 Joinpoint
 - Male - 2 Joinpoints ☆
 - Male - 3 Joinpoints
 - Female - 0 Joinpoints ☆
 - Female - 1 Joinpoint
 - Female - 2 Joinpoints
 - Female - 3 Joinpoints

Final Selected Model Only (☆)

| Test For Parallelism | | | | | |
|----------------------|----------------------|------------------------|------------------------|-------------|---------------------------------|
| Kmax [^] | Numerator Degrees of | Denominator Degrees of | Number of Permutations | P-Value | Significance Level [~] |
| 2 Joinpoint(s) | 5 | 28 | 4500 | 0.127555556 | 0.050000000 |

Final Selected Model: Failed to reject Parallelism
[~] Significance level for individual test
[^] Assumed number of joinpoints for the comparison (parallel or coincident) test [Learn More](#)

Joinpoint Version 4.5.0.0

Estimating the Autocorrelation Parameter

Can the Joinpoint software estimate the autocorrelation parameter?

Answer:

Yes, starting with Version 3.5, the software can estimate the autocorrelation parameter.

If you select "Fit an autocorrelated errors model based on the data", the autocorrelation parameter will be estimated separately for each by-group using the method described in Section 2.3 of Kim et al. (2000). Under this option, the autocorrelation parameter is estimated for the model with the default maximum number of joinpoints or the maximum number of joinpoints set by a user.

Although the autocorrelation may be estimated from the data, correcting for autocorrelation with this estimate may seriously reduce the power to detect joinpoints (see Section 3 of Kim et al. (2000)). We found in our simulations in Table IV of that paper that adjusting for autocorrelation was helpful in maintaining proper size of the tests of joinpoints when there was large autocorrelation. We also found that if there was no autocorrelation then the adjustment seriously affected the power of the test to detect joinpoints. For example we see in Table IV (b) with $\phi = 0$, the power goes from 90% to 68%. This is because it is difficult to differentiate between autocorrelation and joinpoints in a model.

If you suspect that your data are positively autocorrelated, we suggest using the "Fit an autocorrelated errors model with parameter =" option to see how sensitive your results are to changes in autocorrelation. The option should be used as follows:

1. Fit the model with the uncorrelated errors option.
2. If the user suspects that there is positive autocorrelation in the data, then repeat the analysis trying several values of the autocorrelation parameter, say for example 0.1, 0.2, and 0.3. If the results are very similar with different values of the autocorrelation parameter, then the user knows their results will still hold if there is autocorrelation present. If the results change as the autocorrelation parameter changes, then the user may end up presenting the series of results, to show how the results depend on different assumptions about the autocorrelation.

If you suspect negative autocorrelation, the uncorrelated errors model will suffice (see Kim et al., 2000).

Related content

- [Method and Parameters Tab](#)
- [Autocorrelated Errors Option](#)

New default settings for Grid Search

Why did the default settings for Grid Search change?

Answer

In versions prior to 4.5.0.0, the default settings for Grid Search were:

- Minimum number of observations from a Joinpoint to either end of the data (including the first or last Joinpoint if it falls on an observation): 3
- Minimum number of observations between two joinpoints (including any joinpoint that falls on an observation): 4
- Number of points to place between adjacent observed x values in the grid search: 0

Starting in version 4.5.0.0, the default settings for Grid Search are now:

- Min. number of observations from a Joinpoint to either end of the data (excluding first or last joinpoint if it falls on an observation): 2
- Min. number of observations between two joinpoints (excluding any joinpoint if it falls on an observation): 2
- Number of points to place between adjacent observed x values in the grid search: 0

As you can see, the description for these parameters has changed slightly. These new default settings are identical in results to what you used to get with 3, 4, 0. We made these value and text changes to more accurately describe what was going on in the Grid Search algorithm. For example, if you used to enter 4, 5, 0, then the new values will be 3, 3, 0. If you load an old session with the previous values, they will automatically be updated to the corrected values.

We have also separated these settings from those of Hudson's Method to make it clear that they function differently for each method.

Related content

- [Grid Search Method - Details](#)

Hudson's Method Disabled

Why is the Hudson's Method disabled?

Answer

In version 4.5.0.0, Hudson's Algorithm (used to estimate the exact best fitting location of joinpoints in continuous time) was temporarily disabled. It was determined that the algorithms for setting how close joinpoints can be to each other and to the ends of the series, were not working properly when using the Hudson's Algorithm. When examining this issue in detail, it was determined that there were some complicated interactions between these settings and the use of the Hudson's Algorithm. We were not fully able to work out these issues in time for this release. Since version 4.5.0.0 had some important other updates, we decided to temporarily disable Hudson's Algorithm. We apologize for any inconvenience this may cause.

But what about sessions that were created in the past with Hudson's Algorithm? Were they wrong? The answer is no. The parameters supplied to Hudson's Algorithm were just not being used correctly. Hudson's was incorrectly bounding the location of potential Joinpoints more than should have been done with the given parameters. Hudson's should have considered a wider area for potential Joinpoint placement. The use of the same parameters for both Grid Search and Hudson's Method in older versions of Joinpoint led to Hudson's being forced to use parameters that weren't specific for its own use.

Related content

- [Hudson's Method - Details](#)

Discrepancies with SEER Cancer Statistics Review (CSR)

I used the program on SEER data and do not get the same answers as in the SEER Cancer Statistics Review. Why?

Answer:

The SEER Cancer Statistics Review (CSR) uses the same Joinpoint Regression Program. Their analysis uses the "Input Standard Error of Response" setting for the Heteroscedastic Errors Option using the standard error of the rate as calculated by SEER*Stat. Using these options and the current default number of permutations (4499), if the Joinpoint Regression Program chooses the model with 0 joinpoints, the annual percentage rate change will agree with the calculation of this value given in the CSR.

Related content

- [Heteroscedastic Errors Option](#)

Joinpoint APCs not matching SEER*Stat APCs

When I fit a Joinpoint model to find the rate trend, why doesn't the APC for one of the segments match the APC found by SEER*Stat for the corresponding segment?

Answer:

The Joinpoint model assumes that the trend is continuous at the joinpoint, whereas the APC for the corresponding segment calculated by SEER*Stat does not impose the continuity constraint with the consecutive segment. Hence, when there are one or more joinpoints, the APCs from Joinpoint and from SEER*Stat are different, but they should be close to each other. When there are zero joinpoints, the APCs from Joinpoint and SEER*Stat will be the same, except for differences due to the precision level of the data.

Related content

- [Annual Percent Change \(APC\) and Confidence Interval](#) A description of APC and Confidence Interval.
- [APC/AAPC/Tau Confidence Intervals](#)

Non-Significant Change in Slopes

Sometimes, the change in the slopes between two segments is not statistically significant (p -value >0.05) from the t -test, but there is a joinpoint between the two segments or vice versa. Why?

Answer:

The Joinpoint program shows the estimates of the regression coefficients, i.e., intercepts and slopes, and the changes of the slopes. The p -values of the slope changes are calculated from the t -test based on asymptotic normality. Based on a p -value greater than 0.05, one might say the two slopes are not statistically different and hence conclude that the two segments are the same. But using the p -value from a t -test is not as accurate as that from the permutation test, since the t -test is an asymptotic test and the variances are calculated using the information matrix conditional on the estimated joinpoints without imposing the continuity constraint and omitting the offending observations. The software (with the permutation test procedure) does not require the asymptotic normality and maintains the correct Type I error probability level and hence the number of joinpoints determined by the software is more reliable.

Related content

- [Joinpoint p-values](#)

Consecutive Non-Significant Segments

Sometimes, the APC for one segment is significantly different from zero, but when an extra joinpoint in the segment is determined by the Joinpoint software, neither APCs for the two consecutive segments are significant. Why?

Answer:

First, the test of APC is based on asymptotic *t*-test and the number of joinpoints is determined by permutation test, which is more reliable. When a segment is divided into two segments by a joinpoint, there are less data points in each segment, which usually decreases the power of each test. An omnibus test, which combines both segments, is a more powerful test to detect the overall trend for both segments.

Related content

- [Joinpoint p-values](#)
- [Annual Percent Change \(APC\) and Confidence Interval](#) A description of APC and Confidence Interval.

Non-Equally Spaced Time Intervals or Non-Integer Time Points

Can I do Joinpoint regressions where the time intervals are not equally spaced or the time points are not integers?

Answer:

Yes. The Joinpoint software since Version 3.0 allows unequally spaced time intervals and non-integer time points.

Related content

- [Method and Parameters Tab](#)
- [Grid Search Method - Details](#)

Modeling Proportions & Percents

Can Joinpoint be used to model proportions or percents?

Answer:

Joinpoint can be used to model proportions or percents, and for consistency, assume that percents are transformed to proportions before analysis.

For confidence intervals of the model parameters and related p -values, Joinpoint assumes that the data being analyzed arises from a normal distribution or the number of observations is large enough to use the asymptotic normality of the estimated model parameters.

If n_i is the sample size for the i th observation and is large enough that both $n_i p_i$ and $n_i(1-p_i)$ are larger than or equal to 10, then the sample proportions can be considered to asymptotically follow a normal distribution with standard error equal to $\sqrt{p_i(1-p_i)/n_i}$.

If the proportions arise from a complex survey, then the standard error from a complex survey analysis statistical package (e.g. SUDAAN) can be used. If n_i is not large enough for a normal approximation (that is, either $n_i p_i$ or $n_i(1-p_i)$ or both are smaller than 10), then the distribution may be skewed, and Joinpoint results based on asymptotic normality may not be accurate unless the number of observations is large.

To correct heteroscedasticity, one would want to incorporate the standard errors of the proportions and the standard error of $\sqrt{p_i(1-p_i)/n_i}$ can be used. These are the standard errors that are automatically computed and incorporated when proportions are calculated within Joinpoint. If n_i is not available, but all the proportions are in approximately the same range, and the sample sizes are known to be approximately the same, then it may not be necessary to enter the standard errors (i.e. in this case all the standard errors are approximately the same, which is the implicit assumption in Joinpoint if no standard errors are entered).

Related content

- [Dependent Variable](#)
- [Heteroscedastic Errors Option](#)

Zeros in the Dependent Variable

How does the Joinpoint software deal with zeros in the dependent variable?

Answer:

It depends on the settings that are selected:

- If all records for a cohort have zero counts, then the cohort is NOT processed.
- If some of the records for a cohort have zero counts, then Joinpoint will add 0.5 to zero counts in the following scenarios:

| Type of Variable | Run Type | Heteroscedastic Errors Option | Log Transformation |
|------------------|---------------------------|-------------------------------|--------------------|
| Count | Provided in Data File | Poisson Variance | YES or NO |
| Count | Provided in Data File | Constant Variance | YES |
| Crude Rate | Calculated from Data File | Poisson Variance | YES or NO |
| Crude Rate | Calculated from Data File | Constant Variance | YES |

There must be one or more records in the cohort with a count greater than zero in order for Joinpoint to adjust zero counts within the cohort. When a session is executed, Joinpoint will scan the input file and notify users of any cohort that will have a count adjusted. Users will have the option to cancel the job. If the session is executed, the output will display a message for each cohort that had one or more counts adjusted.

- For all other situations, the cohort is NOT processed. In this case, a message will be displayed by Joinpoint and the executed job cancelled.

Related content

- [Dependent Variable](#)

Project or Extrapolate Into the Future

Can I use Joinpoint to project or extrapolate into the future?

Answer:

While the Joinpoint program itself does not do projections/extrapolations, it is not unreasonable to use Joinpoint to project a few years into the future. To do so, one would need to extrapolate the final segment, and compute the standard errors of the projection, using the betas displayed in the output. The Joinpoint Team has had some success using Joinpoint for projections (validated against actual values when the data is rolled back several years so the projected values can be validated against actual values). However, a projection using Joinpoint may or may not be more accurate than other methods (e.g. time series methods), depending on the situation. Long term projections using Joinpoint are not advised. Also, AAPCs, since they are summary measures over an interval, are not as appropriate to use for projections as is the APC of the final segment.

References which used Joinpoint for projections are:

Chen HS, Portier K, Ghosh K, Naishadham D, Kim HJ, Zhu L, Pickle LW, Krapcho M, Scoppa S, Jemal A, Feuer EJ. Predicting US- and state-level cancer counts for the current calendar year: Part I: evaluation of temporal projection methods for mortality. *Cancer* 2012 Feb 15;118(4):1091-9. [[Full Text](#)]

Zhu L, Pickle LW, Ghosh K, Naishadham D, Portier K, Chen HS, Kim HJ, Zou Z, Cucinelli J, Kohler B, Edwards BK, King J, Feuer EJ, Jemal A. Predicting US- and state-level cancer counts for the current calendar year: Part II: evaluation of spatiotemporal projection methods for incidence. *Cancer* 2012 Feb 15;118(4):1100-9. [[Full Text](#)]

Related content

- [AAPC Segment Ranges](#)
- [APC/AAPC/Tau Confidence Intervals](#)
- [Annual Percent Change \(APC\) and Confidence Interval](#) A description of APC and Confidence Interval.

Importing Graphs into Graphics Software

How do I import Joinpoint's graphs into other software such as Word, Excel, or PowerPoint?

Answer:

The graphs themselves, or the data to produce the graphs, can be exported from Joinpoint into a format which can be used by other software packages. You have the following options:

I. Save the graph as a BMP or JPEG file

- **Option 1:** Right click on the graph and a menu will appear. Select the “Export Graph...” menu item. A window will appear where you can specify the file name, the type to save it as (BMP or JPEG image), and the directory to save the file in. Once these are specified, click the Save button.
- **Option 2:** Use the **Output > Export...** menu item. Once the menu item is selected, the Export dialog window will appear. Click the “Displayed Graph” option and then use the associated “Browse” button to select a file name and location to save the file. When browsing, you will be able to set the file type (BMP or JPEG) at the bottom of the window. Once the file name and location have been specified and you have returned to the main Export dialog window, click on the Ok button and the graph will be exported and saved to file.

Then, the saved file can be imported to other software like any other image or picture would be.

II. Save the data coordinates (X, Y-observed, Y-fitted) used for creating the graph

- **Option 1:** Copy and paste the coordinates from Joinpoint into your software (Excel, PowerPoint, etc.). To accomplish this, please perform the following steps:
 1. Go to the data tab.
 2. Copy the contents of the data grid to the Windows clip board. To do this, follow these steps:
 - Right-click on the data grid and a menu will appear.
 - Select the “Select Current Grid” option. Once you do that, the entire grid will turn blue (indicating the grid is selected).
 - Right-click on the grid again, and select the “Copy” menu item. The data grid contents have now been copied to the Windows clip board.
 3. Paste the copied data tab coordinates into your software (Excel, etc.). For Excel: Open Excel, go to the spreadsheet where you want the information copied, right-click on a cell in the spreadsheet and select Paste (or use Ctrl-V).
- **Option 2:** Use the **Output> Export...** menu item. Select to export the contents of the data tab to a file. The file can be imported into other graphics software packages. For a more detailed explanation of how to use this option, please see the following FAQ: [Exporting Results](#).

- [Exporting High Resolution Graphs](#) How do I export high resolution graphs?
- [Exporting Results](#) How do I export Joinpoint results to a text file or to Excel?

Printing Results

How do I print Joinpoint results?

Answer:

Joinpoint can print both session parameters and output results. In order to print either, please follow these steps:

1. Make sure the output or session window you want to print is the active window.
2. Click the print icon () shown in the main Joinpoint toolbar or select the **File > Print...** menu item
3. If you are printing a session:
 - A Report Viewer window will appear showing you the 'Print Preview' of the session parameters.
 - Select the print icon () shown in the upper toolbar to print the report.
4. If you are printing output results:
 - A window will appear providing you with Print Options.
 - Select the Cohort, Model, and Output information you wish to print.
 - Click the Print button.
 - A report viewer window will appear containing your selected output.
 - The number of pages to be printed will appear in the upper left-hand corner of the Report Viewer window. Please review this number to verify that your specified Print Options do not print an unexpectedly large number of pages.
 - Select the print icon () shown in the upper toolbar to print the report.

Related content

- [Print Options](#)
- [Saving Results to PDF or Word](#) How do I save Joinpoint results to PDF or Word?

Saving Results to PDF or Word

How do I save Joinpoint results to PDF or Word?

Answer:

Joinpoint results from all of the tabs in the Output Window can be exported to PDF or Word via the Report Viewer. You must follow the same steps as you would to print results, but instead of sending the results to a printer, you can direct Joinpoint to save the results to PDF or Word. Please follow the steps below:

- Make sure the output or session window you want to print has focus.
- Click on the print icon () shown in the main Joinpoint toolbar or select the **File > Print...** menu item.
- A window will appear providing you with Print Options.
- Select the Cohort, Model, and Output information you wish to print.
- Click the Print button.
- A report viewer window will appear containing a 'Print Preview' of the information you have selected to print.
- The results displayed in the report viewer can now be exported to PDF or Word. At the top of the report viewer window click on the Export icon (). If you place the mouse over the icon, a 'hint' will be displayed saying 'Export'.
- Once the icon is pressed, a menu will appear under it. From this menu, select the output type you want: PDF or Word.
- Once an output type is selected, you will be able to specify a file name and location to save the resultant file.

Related content

- [Printing Results](#) How do I print Joinpoint results?
- [Exporting Results to Text](#)

Exporting Results

How do I export Joinpoint results to a text file or to Excel?

Answer:

Joinpoint results can be saved to a text file in order to be used as input into other software or saved as an Excel spreadsheet. To export the Joinpoint results, please follow these steps:

To export to text files:

- Make sure the output window containing the results you want to export has focus.
- Select the **Output > Export > Text...** menu item.
- A window will appear providing you with various export options.
- Click the check boxes associated with the information that you wish to export, specify the file name associated with each selected item, and define the various other output options.
- Click the Ok button.

To export to an Excel spreadsheet:

- Make sure the output window containing the results you want to export has focus.
- Select the **Output > Export > Excel...** menu item.
- A window will appear providing you with various export options.
- Click the check boxes associated with the information that you wish to export. Each export selection will be placed in its own tab within the spreadsheet.
- Click the Ok button.

Related content

- [Exporting Results to Excel](#)
- [Exporting Results to Text](#)

Exporting High Resolution Graphs

How do I export high resolution graphs?

Answer:

There are two ways to export high resolution graphs from Joinpoint:

- Option 1: If you want to export one graph at a time, you can right-click on the graph and select "Export High Resolution Graph".
- Option 2: If you want to export multiple graphs, use the "Output" -> "Export" -> "Text" menu option. On the Export Window, select the "Graphs" checkbox at the top of the window. Next, select the "Use High Resolution Graph" option and set the pixel size of the graph image. See [Exporting Results](#) in the help for more information.

Related content

- [Graph Tab](#)
- [Exporting Results](#) How do I export Joinpoint results to a text file or to Excel?

Using Excel as a Data Source

When I use Excel as my data source, why do I get "#####" in the data?

Answer:

When using Excel as your data source, Joinpoint will import the values from the Excel spreadsheet exactly as they look. It does not take the actual value in the cell, it takes the formatted display value.

For example, in your Excel spreadsheet you have a cell with the value 5.1234. You have formatted the column it's in to display just 1 decimal place. So in the spreadsheet you see 5.1. When you import this data, Joinpoint will import the value as 5.1. The data displayed in the grid on the Input File tab of your Joinpoint session will reflect this.

The most common reason to get "#####" in a field is that the column is too small. You just need to resize your column large enough to display all of the data in the column. You then need to save your Excel spreadsheet, and then reimport it into Joinpoint.

Related content

- [Input Data File](#)

Empirical Quantile CI's can use Method 1 or Method 2

Which Method is better to use for Empirical Quantile CI calculations?

Answer:

In version 4.6.0.0, Joinpoint allows the use of Method 1 or Method 2 for Empirical Quantile confidence interval calculations. Studies suggest using Method 2 when the number of observations is as small as 10 and Method 1 when the number of observations is greater than or equal to 20.

See [Empirical Quantile](#) for a complete description.

Related content

- [Empirical Quantile Confidence Interval](#) A description of Empirical Quantile Confidence Interval.

Empirical Quantile CI is an Alpha Version

Why is the Empirical Quantile Confidence Interval option labeled as "Alpha"?

Answer:

In version 4.6.0.0, the Empirical Quantile Confidence Interval (CI) calculation was changed and can now use Method 1 or Method 2. It also now applies to the APC and Tau CI's as well as the AAPC CI's. Due to the method addition and the fact that APC and Tau CI's are also computed using the Empirical Quantile method, this feature has set as "alpha" for this release. Being marked as alpha indicates that this feature is still being heavily tested and that it may change in the future.

See [Empirical Quantile](#) for a complete description.

Related content

- [Empirical Quantile Confidence Interval](#) A description of Empirical Quantile Confidence Interval.
- [APC/AAPC/Tau Confidence Intervals](#)
- [Empirical Quantile CI's can use Method 1 or Method 2](#) Which Method is better to use for Empirical Quantile CI calculations?
- [Alpha and Beta Versions](#) Features in Joinpoint which are labeled Alpha or Beta versions

Data Dependent Selection is a new Model Selection Method

Why would I choose Data Dependent Selection as a Model Selection Method?

Answer:

In version 4.6.0.0, a new Model Selection Method has been added called "Data Dependent Selection". This method will compute both BIC and BIC3 for each model, and then choose the best for the final result. Being marked as "beta" indicates that this feature is still being heavily tested and that it may change in the future.

See [Data Dependent Selection](#) for a complete description.

Related content

- [Data Dependent Selection \(DDS\)](#)
- [Model Selection Method](#)

Clustering controls are no longer available

What happened to the Clustering controls on the Advanced Analysis Tools Tab?

Answer:

In version 4.6.0.0, we have removed the controls used for Clustering, which were visible but disabled in previous versions of Joinpoint. Some users were confused about these controls which would never become enabled for use. It was decided to remove these controls until the Clustering feature is fully implemented and usable.

Related content

- [Advanced Analysis Tools Tab](#)

SEER*Stat Data Import

What type of SEER*Stat sessions are compatible?

Answer:

Results from SEER*Stat rate, frequency, and trend sessions can be imported into Joinpoint. You can learn more about how to use SEER*Stat from this [webpage](#).

Related content

- [Input File Tab](#)

APC Definition

What is an APC?

Answer:

Annual Percent Change (APC) is one way to characterize trends in cancer rates over time. With this approach, the cancer rates are assumed to change at a constant percentage of the rate of the previous year. For example, if the APC is 1%, and the rate is 50 per 100,000 in 1990, the rate is $50 \times 1.01 = 50.5$ in 1991 and $50.5 \times 1.01 = 51.005$ in 1992. Rates that change at a constant percentage every year change linearly on a log scale.

Related content

- [Annual Percent Change \(APC\) and Confidence Interval](#) A description of APC and Confidence Interval.

AAPC Definition

What is an AAPC?

Answer:

Average Annual Percent Change (AAPC) is a summary measure of the trend over a pre-specified fixed interval. It allows us to use a single number to describe the average APCs over a period of multiple years. It is valid even if the joinpoint model indicates that there were changes in trends during those years. It is computed as a weighted average of the APCs from the joinpoint model, with the weights equal to the length of the APC interval.

Related content

- [Average Annual Percent Change \(AAPC\) and Confidence Interval](#)

Tau Definition

What is Tau?

Answer:

The tau parameters are the unknown joinpoints in the joinpoint regression model described by Kim et al. (2000). For fitted models with one or more joinpoints, the Joinpoint software displays the estimates of the tau parameters and estimates of other joinpoint model parameters.

Related content

- [Parameterizations](#)
- [Estimated Joinpoint \(Estimated Tau\) and Confidence Interval](#)

The Best Fit number of joinpoints can decrease

Why does the number of joinpoints for the best fit sometimes decrease when the maximum number of joinpoints is increased?

Answer:

Users of joinpoint need to specify the minimum (denoted MIN) and maximum (denoted MAX) number of joinpoints to be considered. To find the number of joinpoints for the best fit, a sequence of tests is performed, first testing $H_0: K_0 = \text{MIN}$ vs. $H_a: K_1 = \text{MAX}$ number of joinpoints. If we reject the null hypothesis, we next test $H_0: K_0 = \text{MIN} + 1$ vs. $H_a: K_1 = \text{MAX}$. If we fail to reject the null, we test $H_0: K_0 = \text{MIN}$ vs $H_a: K_1 = \text{MAX} - 1$. Testing continues in this manner until we determine the number of joinpoints by continuing to step the lower limit up by one if we reject the null hypothesis or by stepping the upper limit down by one if we fail to reject the null hypothesis. To maintain an overall significance level α for this sequence of tests, each test is performed at the $\alpha / (\text{MAX} - K_0)$ level. Therefore, even when the overall significance level is kept constant, a change in the MAX or MIN number of joinpoints, which would change the significance levels of the individual tests, may result in a change to the number of joinpoints in the best fitting model.

Related content

- [Number of Joinpoints](#)
- [Selecting the Final Model](#) Why doesn't the joinpoint program give me the best possible fit? I can see other models with more joinpoints that would fit better. Exactly how does the program decide which tests to perform and which joinpoint model is the final model?
- [Permutation Test Details](#) Describe the permutation test used here.
- [Permutation Test Details](#)

Continuous Data

How do you handle an independent variable that is not yearly data, but is continuous data and is presented in discrete categories, or is ordinal data?

Answer:

In general, the joinpoint software can handle independent variables that are not yearly data. For example, if the data is reported in months instead of years, the model can be run exactly the same as using “year” as independent variable. The interpretations of annual percentage change (APC) and average annual percentage change (AAPC) should change accordingly. APC now is “monthly percentage change”, and AAPC “average monthly percentage change”. However, if the data the monthly data is coded in years (i.e. 1/12th, 2/12ths,..._ instead of 1,2, ...) then the APC would be interpretable as an “annual percent change”.

If the independent variable represents chronological age instead of calendar years, then the APC would represent “percent change per year of age”, and if it represents dollars, then the APC would represent “percent change per dollar”.

If the independent variable represents intervals, e.g. age intervals of 10-19, 20-29, 30-39, etc, then it is recommended using the mid-point of the interval (e.g. 15, 25, 35) or the mean or median of the observations in each interval. For open-ended groups (e.g., age 80+), one should consider eliminating it or one may use the median (or mean) of all the observations that are above 80 years old for this group.

If the independent variable is an ordinal variable such as stage of disease (e.g. Stage I, II, III, IV), while the analysis can still be run, the APC cannot be interpreted because the underlying data is not continuous.

Related content

- [Independent Variable](#)
- [Input Data File](#)
- [APC/AAPC/Tau Confidence Intervals](#)

Joinpoint Revision History

What are the most recent changes to the Joinpoint program?

Version 4.8.0.1 (released April 22, 2020)

Bug Fixes

- Data files with very small standard errors caused the Permutation Test to miscalculate the P-Values and potentially affected the Final Selected Model. This miscalculation is only present in version 4.8.0.0. It is recommended users re-run any analyses conducted with Joinpoint 4.8.0.0 that used the Permutation Test.

Version 4.8.0.0 (released April 15, 2020)

New Features

- Minimum APC Difference Worth Detecting (MADWD) - In some situations Joinpoint may determine that there are one or more pairs of segments where the difference in the Annual Percent Change (APC) between some consecutive segments is quite small. Users can enter a minimum APC difference worth detecting as a minimum percentage point difference (Min. PPD). Joinpoint will start with the final selected model that was chosen using the specified analysis criteria and lower the number of Joinpoints until all consecutive segments pass the Min. PPD criteria.
- Copy Graph - The Joinpoint graph can now be copied to the user's clipboard from the right click menu.
- Cohort Tree Filter - Certain filter options can now be set as default options for each new Joinpoint Output.

Bug Fixes

- By-variables with equal signs in their text would not display properly in the Joinpoint Output.
- By-variables with leading or trailing spaces in their values prevented sessions from executing.
- Sessions with the Pairwise Comparison option that had varying number of observations were allowed to execute when they should not.
- The Control Box was added to the Nag warning for Default System Font allowing users to close the warning when the "Ok" button is not visible.
- The Output Properties was displaying the Jump Location for the Comparability Ratio.
- Removal of duplicate Independent Variable values in the Define Variable Window
- If a session had missing Dependent Variable values and the Type of Variable was set to "Count", Joinpoint would throw an exception.
- Joinpoint would crash if a session had varying number of observations and too few observations to fit any Joinpoints.
- Quotes around the Adjustment Variable in the data file caused an exception.

Version 4.7.0.0 (released February 26, 2019)

New Features

- Two new model selection methods were added: Weighted BIC and Weighted BIC-Alternative.
- The statistically significant flag was changed from ^ to *.
- The restriction requiring Standard Populations to be the same across cohorts when calculating age-adjusted rates was removed.
- When using a data file that has 32 or more observations, a message is displayed to users to recommend setting the maximum number of Joinpoints to 6.
- A "Session Note" field was created. This allows users to add a brief note on the Joinpoint session that is then displayed in the output.
- A new export file was created that contains the number of Joinpoints in the Final Selected Model for each cohort.
- A second Execute button was added that prevents Joinpoint from displaying any warning messages before running the session.
- A custom display labels option was added to the independent variable define.

Bug Fixes

- Toggling the "File contains column headers" option could result in a double header being displayed in the data preview window.
- SEER*Stat export with variables as "Quoted Labels" caused Joinpoint to crash.
- The Include/Exclude cohorts feature did not exclude Delay & Non-Delay cohorts correctly.
- Opened JPO now displays the file name in the header.
- Double clicking a JPS file whose data file had changed caused that file to open multiple times.
- Fixed the Nag Warning display for zoomed displays.
- Added correct .dll files required to use the "Print" feature.
- Fixed a bug when Joinpoint encountered blank lines in the data file.
- Fixed a bug involving missing character=space and data files containing missing data.
- Updated the help links to the correct base page on the Plone site.
- Update the Parameters report to correctly reflect the current UI.

Version 4.6.0.0 (released April 16, 2018)

New Features

- Added "Data Dependent Selection" as a new Model Selection Method option. This is an Alpha feature.
- Added "BIC3" as a new Model Selection Method option.
- Added the ability to select Parametric or Empirical Quantile methods for APC and Tau Confidence Intervals.
- Two separate methods of the Empirical Quantile confidence interval calculations are now provided: Method 1 and Method 2. This is an Alpha feature.
- Added the option to set the seed value used for generating random numbers during resampling under the Empirical Quantile method as well as an option to keep the seed constant or have it change with each cohort.
- A new Joinpoint File Type was created. The Joinpoint Template (.jpt) allows users to save their Session settings, including variable locations, and then create a new session with a new data file. This allows users with multiple similar data files to easily recreate sessions.
- Added a more advanced filter control to the Cohort Tree of the Output Window.
- Added slope significance indicator on the Output Window graph when Log Transformation is No.
- Added a new grid to the Model Selection tab of the Output Window which displays the

Model Selection Method chosen.

- Added a footer to the Output Window to display the Confidence Interval Method used.
- Added examples on how to call the console version from a SAS or R program.
- Right-click “Save As” for graphs now supports PNG and TIFF formats (for normal and high resolution graphs).

Bug Fixes

- Changed the data file preview grid (on the Input Tab of the Joinpoint session) to allow keyboard navigation.
- Removed Multi-Group Clustering group box from the Advanced Analysis Tools Tab of the Joinpoint session.
- Improved the loading and processing of input data files.
- Improved the display of formulas in the online help.
- Fixed the High Resolution Graph to look more like the normal graph.

Version 4.5.0.1 (released June 12, 2017)

New Features

- The value 100 has been added to the list of possible “Rates per” (population at risk) when computing Crude or Age-Adjusted Rates.

Bug Fixes

- As of version 4.5.0.0, users can set the Alpha levels for the Joinpoint Locations, APCs, AAPCs, and Jump Value and Comparability Ratio. Upon installation of that version, the default values for each were set incorrectly to 0.5. The default values should have been 0.05. **Installing the new version of Joinpoint (4.5.0.1) will change any 0.5 Alpha setting to 0.05.**
- The significance level footnote on the Model Selection tab referenced the wrong alpha level. This has been corrected.

Version 4.5.0.0 (released May 17, 2017)

New Features

- Two tabs on the Session Window have been renamed. “Advanced” has become “Method and Parameters”. “Comparison” is now “Advanced Analysis Tools”. Several controls on these tabs have been moved. More details are below.
- The Grid Search now has specific parameters for controlling Joinpoint placement. The default values for Grid Search’s “Minimum number of observations from a joinpoint to either end of the data” and “Minimum number of observations between two joinpoints” have been changed. For more details on this change, please go to the [Help section](#).
- Hudson’s Method has been disabled.
- The Pairwise Comparison has two new parameters: Significance Level and Max Number of Randomly Permuted Data Sets.
- Alpha Values for APCs, AAPCs, and Jump Value and Comparability Ratio have been added to the Session Preferences. Session Preferences can now be defaulted to different values.
- For Jump Model and Comparability Ratio Model analyses, the “CR Estimate” is now displayed in the Graph Legend on the Output Window. Added CR Estimate precision setting to the Display Options window.
- An APC/Slope precision setting has been added to the Display Options window.
- Jump Model and Comparability Ratio display and precision options have been added to the

Display Options window.

- The layout of the Model Selection and Comparison tab settings in the Display Options window has been updated.
- Added the Alpha Level to the footer of the Model Selection tab on the Output Window.

Bug Fixes

- If the Jump Model option was run, the Comparability Ratio table (located on the Model Estimates output tab) contained an incorrect statistic. For the Jump Model, the column labeled "Variance of CR (a/b)" contained the standard error of the comparability ratio value produced from the Jump Model, not the variance. The variance column has been replaced by a column containing the standard error of the comparability ratio.
- The Comparability Ratio table, located on the Model Estimates tab when the Jump Model or Comparability Ratio Model are run, has been corrected with the following additional adjustments:
 - The Comparability Ratio Model now produces a Lower and Upper CI, Test Statistic and *p*-value. These are computed using a normal (z) distribution.
 - Several footnotes have been added to the table to better describe the statistics displayed.
- If all values of a by-variable weren't present in the first grouping for the by-variable, Joinpoint wouldn't execute due to encountering an unknown value for the by-variable.
- The fields for the Comparability Ratio model export were in the wrong position.
- 0.5 is added to zero counts only under the following conditions:
 - The Run Type is "Provided in Data File", the Type of Variable is "Count", and the Heteroscedastic Errors Option is "Poisson Variance".
 - The Run Type is "Provided in Data File", the Type of Variable is "Count", the Heteroscedastic Errors Option is "Constant Variance", and the Log Transformation is "YES".
 - The Run Type is "Calculated from Data File", the Type of Variable is "Crude Rate", and the Heteroscedastic Errors Option is "Poisson Variance".
 - The Run Type is "Calculated from Data File", the Type of Variable is "Crude Rate", the Heteroscedastic Errors Option is "Constant Variance", and the Log Transformation is "YES".
- A message is displayed in the Output Window for each cohort that had 0.5 added to one or more zero counts.
- A warning message is now displayed when the number of permutations for the Permutation Test is set below 4499. The minimum number of permutations allowed is now 1000.
- Fixed a problem where APC and AAPC exports were adding an extra field delimiter at the end of each row.
- Corrected the display of the Dependent Value column headers on the Data Tab of the Output Window.
- Added restriction that you can only use Poisson Variance if using Provided Count or Calculating Crude Rate.
- The Maximum number of joinpoints was not being validated on the correct Grid Search settings.
- Fixed a problem where grids on the Output Window were not always left aligned.
- Fixed a crash in the Output Window when the Non-Delay cohort of a Delay \ Non-Delay session was a bad cohort.
- The Comparability Ratio grid on the Output Window will display an explanation footer when an incomputable value is displayed.
- For Jump and Comparability Ratio runs:
 - The Last Observation of Old Coding must be at least 4 or more data points from either end of the data.
 - If you use Constant Variance, the Variance of Comparability Ratio will be set to N/A.

- You cannot set Log Transformation to No.

Version 4.4.0.0 (released January 4, 2017)

New Features:

- Session Options:
 - Jump Model and Comparability Ratio calculation options have been added, which allows analysts to estimate trends even when there is a coding change in the data series. For further information on these new options please refer to the [help manual](#).
 - The ability to analyze only select cohorts from the input data.
 - Joinpoint can now analyze data files containing cohorts with an unequal number of observations.
- Output Viewer Options:
 - The output viewer window now contains a tree view for selecting cohorts.
 - Multiple cohort models can be viewed on a single graph. All other output tab information is also updated to contain all of the selected cohort models.
 - AAPC ranges can be specified using the output menu.
- Joinpoint can now export high resolution graphs.
- By Variables now allow text values.
- Joinpoint now has an on-line tech support request form.

Bug Fixes:

- Logarithmic scale display on the Output Window graph sometimes displayed incorrect Y axis values.
- In rare occurrences, the graph axis positions in the Output Window did not show all observed data values.
- JPO files over 1 megabyte would sometimes cause loading errors.
- Dictionaries with a variable format of "labels" would show as blank in the Format Editor.
- Some session properties were not being displayed in the Output Properties window.
- Joinpoint will not let users add or remove variables when a SEER*Stat Dictionary file is used.

Version 4.3.1.0 (released April 19, 2016)

Bug Fixes:

- Several library files were missing from the 4.3.0.0 installation. Those library files prohibited some users from reading or exporting to Excel and generating reports.
- A bug in the By-Variable format generator was fixed.

Version 4.3.0.0 (released April 15, 2016)

New Features:

- The Joinpoint session now has an input data tab. The new tab combines the Data Import Wizard and the session Specifications Tab.
- AAPC ranges can now be set on the Advanced Tab.
- Joinpoint now links to an on-line help system. If you do not have internet connectivity, a PDF version of the help system is installed with Joinpoint.
- Matrix export was expanded to export to Excel.
- Joinpoint now has 3 sample sessions. These sessions can be loaded from the menu File-

>Sample Joinpoint Sessions. These new sample sessions have replaced the “Blank Session” option.

Bug Fixes:

- Various bugs were fixed in the matrix export.

Version 4.2.0.2 (released June 23, 2015)

- The data import wizard was updated to allow non-integer values.
- When computing age-adjusted rates, Joinpoint will now skip cohorts when the total population for the cohort is zero.

Version 4.2.0.1 (released May 6, 2015)

- The data import wizard was updated to correctly handle files containing the information needed for Joinpoint to compute age-adjusted rates.

Version 4.2.0.0 (released April 29, 2015)

New Features:

- Joinpoint now provides a new method for computing AAPC Confidence Intervals. The new method is called the Empirical Quantile method. The previous AAPC CI method is still available and is called the Parametric method.
- When performing the AAPC comparison and using the parametric method to calculate the AAPC Confidence Intervals, a new method has been implemented to calculate the confidence interval for the difference between the two true AAPCs. The AAPC comparison cannot be performed when using the Empirical Quantile method to calculate the AAPC Confidence Intervals.
- Joinpoint can now process Delay and Non-Delay Adjusted statistics output by SEER*Stat. For SEER*Stat exports containing both delay and non-delay statistics, Joinpoint can graph both together for each cohort.
- When computing crude or age-adjusted rates, Joinpoint now provides an option for users to display their rates per a certain population (e.g. 100,000 population at risk).

Bug Fixes:

- There was a problem with the handling of a specific data file via the import wizard that was identified and resolved.
- A bug that prevented users from defining their independent variable values was resolved.

Version 4.1.1.5 (released February 26, 2015)

- A bug preventing the use of CSV files was fixed

Version 4.1.1.4 (released February 19, 2015)

- The p -value for a two-sided test that the true APC is zero is calculated based on a t distribution.

Version 4.1.1.3 (released December 29, 2014)

- A bug was fixed involving the text file import wizard.

Version 4.1.1.2 (released December 15, 2014)

- A small bug was fixed to prevent Joinpoint from terminating its calculations under certain circumstances.

Version 4.1.1.1 (released October 7, 2014)

- A small bug was fixed to prevent Joinpoint from terminating its calculations under certain circumstances.

Version 4.1.1 (released August 5, 2014)

Bug fixes:

- In pairwise comparison runs, all values in the AAPC comparison table were being flagged as statistically significant, whether or not that was actually the case. This problem has been resolved and the program now flags only those that are significant.
- The X and Y axis labels in the output graphs were at times not being displayed. This problem has been resolved.

Version 4.1.0 (released April 25, 2014)

Enhancements:

- Joinpoint now has the ability to process multiple cohorts simultaneously (threading). Joinpoint will inspect the PC it is running on and determine the number of cohorts that can be processed concurrently. Many times this is equal to the number of CPU cores on the PC. Processing multiple cohorts will decrease the time it takes Joinpoint to run a session. Please note that this feature will not work when executing Joinpoint under Windows XP. Windows XP users will only be able to process one cohort at a time (similar to previous versions of Joinpoint).
- Users can adjust the maximum number of threads (cohorts to process concurrently) that Joinpoint will attempt to use when executing a session. Joinpoint will automatically set the number of threads to use to the maximum possible for the PC it is running on. Users can then lower the number of threads if they want. This option is provided via the File->Preferences menu.
- A new session execution progress window has been created. The window will now show the progress of each cohort that it is currently processing. When multiple cohorts are being processed concurrently, a progress bar for each will appear in the progress window.
- The APC and AAPC tables on the Trends tab in the output have two additional statistics: Test Statistics and *P*-Value.
- The slope can now be displayed in the data table when the Log Transformation is set to NO. The report printing was also updated to reflect this change.

Bug Fixes:

- The AAPC Confidence intervals will only be computed if the variance of each associated APC is available.
- The graphing of X-axis values where the range between the minimum and maximum value was less than one was updated to display 10 axis points.
- The correlation and covariance tables on the Model Estimates tab in the Joinpoint output were computed only if unconstrained estimates with offending data points deleted is available.

Version 4.0.4 (released May 6, 2013)

- The program was recompiled so that it would be compatible with Windows XP.

Version 4.0.3 (released April 29, 2013)

Enhancements:

- If standard errors are used in the calculations, then they will be displayed on the Data tab (whether they are provided or computed). Users have the ability to change the display precision via the output options.
- The APC footnote character was changed from * to ^.
- The By Variable “Delete...” button was renamed to “Remove...”
- When using a SEER*Stat dictionary file (*.DIC), any independent variable value with a dash (“-”) located in its format label will not be used in the analysis. The label will be updated to contain the following text: (not used in calculations). This functionality was added so that users would not have to reproduce their SEER*Stat analyses in order to remove totals or sub-totals in their independent variable.
- Graph Display Option enhancements:
 - The legend can be moved to the Right, Left, Top, or Bottom of the graph.
 - X and Y axis label precision can be changed.
 - The locations of each Joinpoint can be displayed on the graph (e.g. “Joinpoint 1: 1981”)
 - The legend can be adjusted to display the APC ranges.
 - The modeled line for each cohort can be set to a single color (e.g. Blue). In comparison runs, the second cohort modeled line will be Green.
- Added a Print button to the main Joinpoint toolbar.
- File associations have been enabled for Joinpoint. Joinpoint will automatically open when either a Joinpoint session (*.JPS) or output matrix (*.JPO) file name is double-clicked.
- Users can now drag and drop files onto the Joinpoint program and it is a file type Joinpoint recognizes it will open it up. Session (*.JPS) and Output (*.JPO) files will automatically open. If a .TXT file is dropped onto Joinpoint, this will trigger the Data Import Wizard. If a SEER*Stat DIC file is dropped, this will trigger the session open process.
- In the output window, depending on which Model Selection Method was specified by the user, a tab with one of the following three labels was shown: “Permutation Tests”, “BIC”, or “Modified BIC”. Those three labels have been replaced by the following label: “Model Selection”. This modification was made to standardize the labeling in the Joinpoint application.
- A standard way of labeling the best, selected model was implanted into the Joinpoint application. The label “Final Selected Model” will now appear instead of labels such as ‘Best Fit” and “Best Model”. This modification was made to standardize the labeling in the Joinpoint application.

Bug Fixes:

- The Trends tab on the output form will not be displayed when the Log Transformation option is set to No. This is because no APCs or AAPCs are produced when this option is selected. Export and Output Options have also been similarly updated.
- When having Joinpoint compute Crude Rates while using the Poisson Variance Heteroscedastic Errors Option, there was the potential to have Joinpoint read in the wrong data column for the counts. This has been resolved.
- The covariance and correlation matrices for pairwise comparisons were not always refreshing themselves correctly when turning them on and off.

Version 4.0.1 (released January 9, 2013)

- Fixed a bug dealing with the decimal symbol for international users.

Version 4.0 (released December 17, 2012)

- Joinpoint can now calculate crude rates, age-adjusted rates, proportions, and percentages if their components are provided in the input data file. Associated standard errors can also be computed. Please see the Dependent Variable section for more details.
- The following modifications were made to the Joinpoint Session:
 - The new Comparison tab was created. The tab contains the parameters related to Pairwise comparisons and Clustering. Clustering is a future enhancement.
 - The “Test for pairwise differences between by-groups” control (originally on the Specifications tab) was moved to the Comparison tab and renamed “Pairwise Comparison”. In order to produce Comparison analyses, users must also use the “Comparison Type” control on the Comparison tab.
 - The Specifications tab was reorganized. Some items have been moved to other parts of the screen and have more descriptive labels (i.e. “Model” was moved from the upper left to the lower right and is labeled “Log Transformation”).
 - The Dependent Variable section (on the Specifications tab) was expanded to accommodate Joinpoint’s new ability to compute various statistics (as previously mentioned). Users can now specify whether the Dependent variable will be calculated or provided and what specific type of variable it is (Count, Crude Rate, Age-Adjusted Rate, Proportion, Percentage, or Other).
 - The Heteroscedastic Errors Option was reorganized. The two Poisson options in the previous version of Joinpoint are now combined into one. “Poisson Variance using Crude Rate” is now specified by selecting Dependent Variable=Calculated, Type=Crude Rate, and Heteroscedastic Errors Option=Poisson Variance. “Poisson Variance using Count” is now specified by selecting Dependent Variable=Provided, Type=Count, and Heteroscedastic Errors Option=Poisson Variance.
- A new report viewer has been implemented. Reports can be printed or exported to PDF, Word, or Excel.
- The Autocorrelation Parameter is now displayed on the Model Estimates tab of the Output window when “Fit an automated errors model based on the data” is selected for the Autocorrelated Errors Options.
- Errors found in the input data file are now reported by cohort. When a session is executed, if any data errors are found a special window is displayed with the errors found for each cohort.
- Joinpoint is now compiled under Microsoft’s Visual Studio. The installation of Joinpoint includes Microsoft’s .NET framework 4.0.
- All user-specific options are now saved in a parameter file instead of the Window’s

registry.

Alpha and Beta Versions

Features in Joinpoint which are labeled Alpha or Beta versions

There are features in Joinpoint which are labeled either "Alpha Version" or "Beta Version". These are features that are in various states of development.

Alpha Version

Features that are labeled "Alpha" are brand new features. Being marked as alpha indicates that this feature is still being heavily tested and that it may change or be removed in the future. We strongly recommend that you use these features carefully and only for experimentation.

Current "Alpha" features:

- Empirical Quantile
- MADWD

Beta Version

Features that are labeled "Beta" are features that have passed through the Alpha stage already. They have been tested and are safe to use. The usefulness of the feature is still being evaluated and it may be removed in a future version. You are welcome to use these features, but we discourage relying on them for any published works.

Current "Beta" features:

- Weighted BIC
- Weighted BIC-Alt
- Data Dependent Selection

References

Basic Method

Kim, H-J, Fay, M.P., Feuer, E.J., and Midthune, D.N. (2000) "Permutation Tests for Joinpoint Regression with Applications to Cancer Rates", *Statistics in Medicine* 19, 335-351. (correction: 2001;20:655). Correction to Table 1(a) of Kim, et al. is provided as a PDF at <http://surveillance.cancer.gov/documents/joinpoint/table1.pdf>.

This is a paper where joinpoint regression is applied to describe cancer rates and the permutation test is proposed to determine the number of significant joinpoints. The grid search of Lerman (1980) was used to fit the segmented regression function and the p-value of each permutation test is estimated using Monte Carlo methods, and the overall asymptotic significance level is maintained through a Bonferroni adjustment.

Background Papers for Basic Method

Feder, P.I. (1975) "On Asymptotic Distribution Theory in Segmented Regression Problems: Identified Case", *Annals of Statistics* 3, 49-83.

Feder studied asymptotic properties of the least squares estimators in multi-segment regression and proved that under some technical conditions on the independent variable, the least squares estimators are consistent and asymptotically normal.

Lerman, P.M. (1980) "Fitting Segmented Regression Models by Grid Search", *Applied Statistics* 29: 77-84.

Lerman proposed the grid search method to fit segmented line regression where the joinpoint estimates occur at discrete grid points, and studied asymptotic inference using asymptotic normality proved by Feder (1975).

Hudson, D. (1966) "Fitting segmented curves whose join points have to be estimated", *Journal of the American Statistical Association* 61, 1097-1129.

Hudson proposed a procedure to fit a segmented regression curve whose joinpoints can be estimated anywhere in the data range. In the first stage, the model is fit to every feasible partition of the data without imposing continuity across the change-points. We then examine the locations of the intersection points of the least square lines. When the intersection point of the two adjacent least squares lines is not between the two data points which separate the segments, we make an adjustment in the fit by mathematically imposing the continuity constraints at one of the two boundary data points. The final estimates are obtained by searching for the global minimum of residual sum of squares.

Fitting Joinpoints in Continuous Time

Yu, B., Barrett, M., Kim, H-J, and Feuer, E.J. (2007) "Estimating Joinpoints in Continuous Time

Scale for Multiple Change-Point Models", Computational Statistics and Data Analysis 51, 2420-2427.

In this paper, we extend the Hudson's continuous fitting method to a multiple joinpoint model and discuss some practical issues in the implementation. We also compare computational efficiencies of the Lerman's grid search and the Hudson's continuous fitting.

Early Stopping Rules for the Permutation Test

Fay, M.P., Kim, H-J, and Hachey, M. (2007) "On using Truncated Sequential Probability Ratio Test Boundaries for Monte Carlo Implementation of Hypothesis Tests", Journal of Computational and Graphical Statistics 16 (4), 946-967.

Joinpoint selects a final model conducting a series of permutation tests and we can save computation time by using sequential stopping boundaries. This paper proposes a truncated sequential probability ratio test boundary to stop resampling when the early replications indicate a large enough or small enough p-value, and studies its properties.

Comparing Two Joinpoint Regression Lines

Kim, H-J, Fay, M.P., Yu, Binbing, Barrett, M.J., and Feuer, E.J. (2004) "Comparability of segmented line regression models", Biometrics 60, 1005-1014.

We propose a procedure to compare two segmented line regression functions, especially to test (i) whether two segmented line regression functions are identical or (ii) whether the two mean functions are parallel allowing different intercepts. A general form of the test statistic is described and then the permutation procedure is proposed to estimate the p-value of the comparability test.

AAPC

Clegg, L.X., Hankey, B.F., Tiwari, R., Feuer, E.J., Edwards, B.K. (2009) "Estimating average annual percent change in trend analysis", Statistics in Medicine 28(29): 3670-8.

AAPC CI

Kim H-J, Luo J., Chen H.S., Green D., Buckman D., Byrne J., Feuer E.J. (2017) "Improved Confidence Interval for Average Annual Percent Change in Trend Analysis", Statistics in Medicine 36(19): 3059-74.

Studies on the performance of Joinpoint

Kim, H-J, Yu, B., and Feuer, E.J. (2008) "Inference in segmented line regression: A simulation study", *Journal of Statistical Computation and Simulation* 78:11, 1087-1103.

Via simulations, this paper empirically examines small sample behavior of asymptotic confidence intervals and tests, based on Feder (1975)'s asymptotic normality of least squares estimators in joinpoint regression, studies how the two fitting methods, the grid search and the Hudson's continuous fitting algorithm, affect these inferential procedures and also assesses the robustness of the asymptotic inferential procedures.

Kim, H-J, Yu, B., and Feuer, E.J. (2009) "Selecting the number of change-points in segmented line regression", *Statistica Sinica* 19(2):597-609.

In this paper, we show that under some conditions, the number of joinpoints selected by the permutation procedure of Joinpoint is consistent. Via simulations, the permutation procedure is compared with some information-based criteria such as Bayesian Information Criterion (BIC).

Model Selection

J. Kim and H.-J. Kim (2016), "Consistent Model Selection in Segmented Line Regression", *Journal of Statistical Planning and Inference* **170**, 106-116.

N.R. Zhang and D. O. Siegmund (2007), "A Modified Bayes Information Criterion with Applications to the Analysis of Comparative Genomic Hybridization Data", *Biometrics* **63**, 22-32.