

Tutorial of the Joinpoint Regression Program (version 4.8.01)

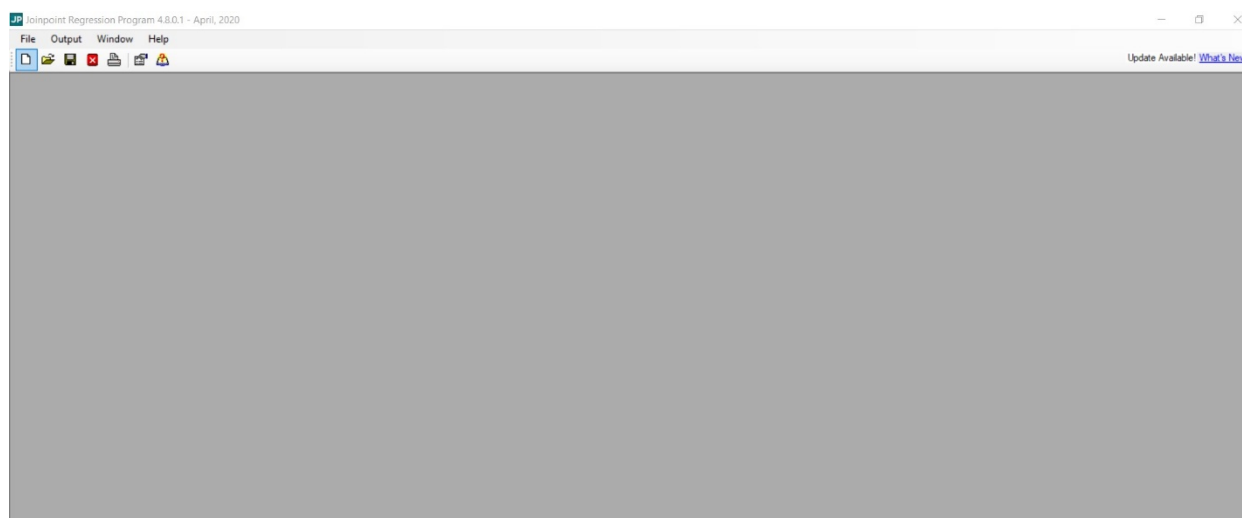
Taylor P. van Doren

Last updated: October 22, 2021

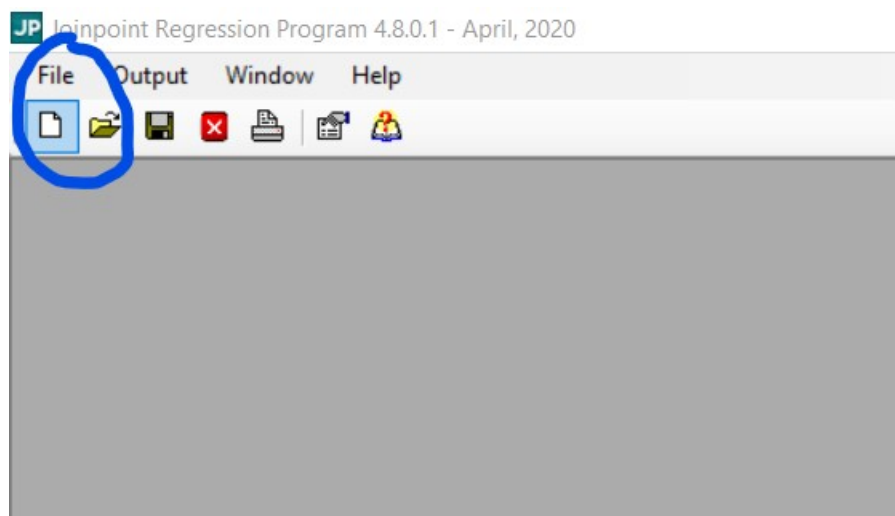
This document will provide a walkthrough of the [Joinpoint Regression Program](#), which is an open-source software program that you can request to download from the NIH through the above link. This program is simple and powerful, but the major downsides of running analyses in a software outside of R or other statistical software is that you must physically change programs in order to do your analyses. Additionally, the data visualizations provided automatically in the Joinpoint Regression software are not very appealing, and they only plot one fitted model at a time. One of the major benefits of this software, however, is unlike the **ljr** package in R, you do have the option to run a linear fit model ($y = mx$) rather than a logistic model ($\ln(y) = mx$) if you want. The Joinpoint Regression software also generates all the data you need to create attractive data visualizations of your data and all the model fits that you tested, and it provides a robust data selection procedure so you can easily see what the best fit model for your data is.

Given the above information, it's clear that there are some solid benefits to using this software as opposed to simply running the analyses in R. Therefore, I've dedicated this document (and a part of my [Github repository](#)) to a tutorial of this program. It's also great to have options, especially when they are good options!

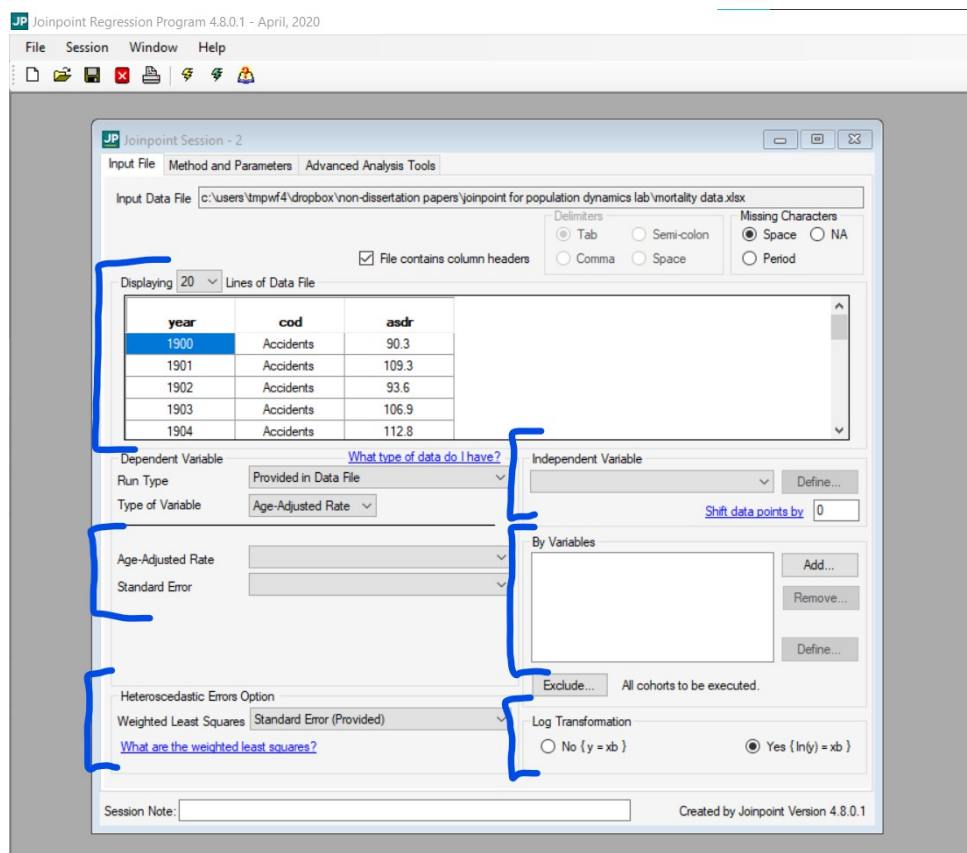
Once you've downloaded your program, open it up. This is the empty home screen of the Joinpoint Regression Program interface.



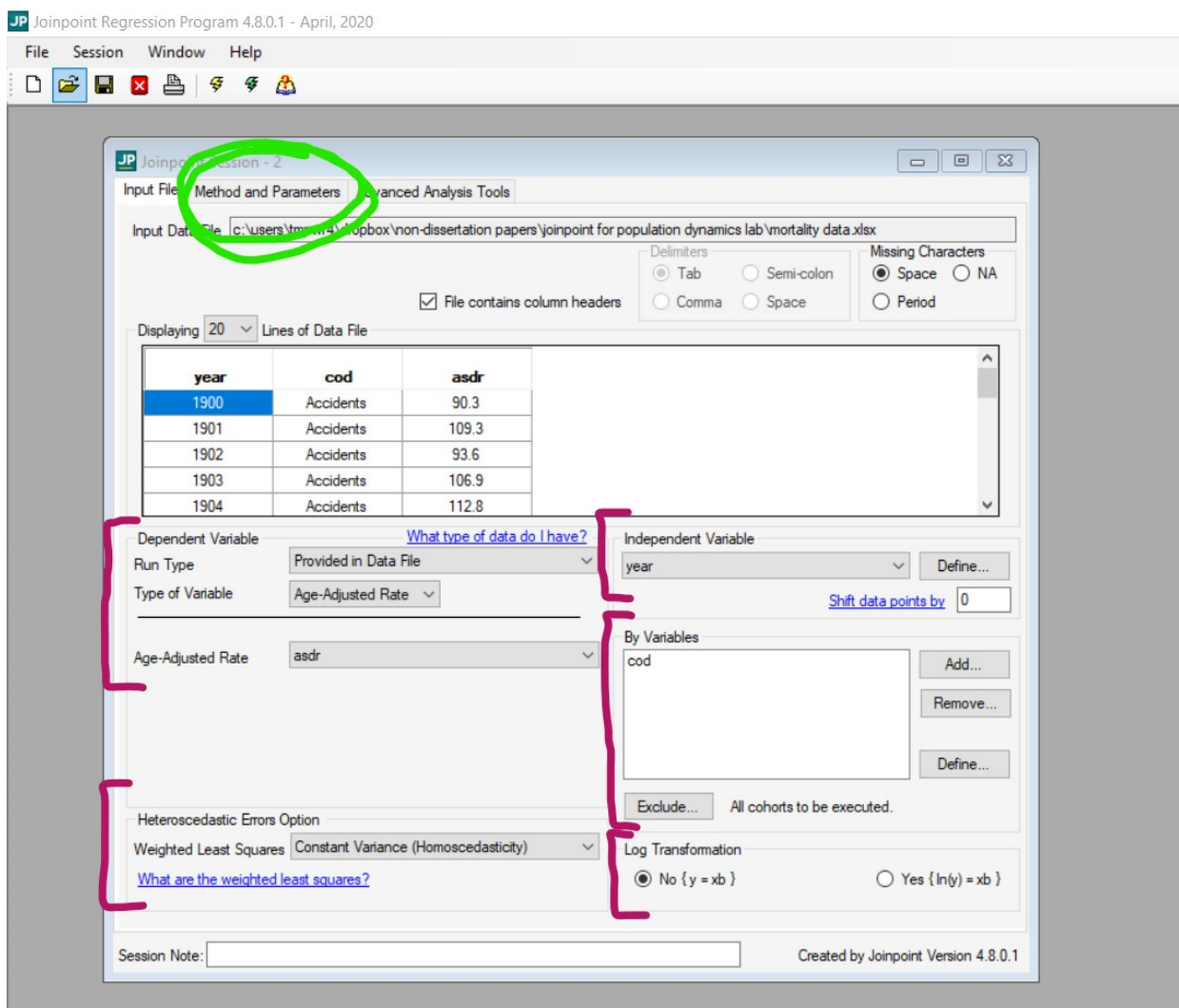
In order to do any analyses, you'll have to click on the black sheet to upload some data. In the Github repository, there is a file called '**mortality data.csv**'. That is the file you'll want to upload for this analysis



When you upload the raw data file, you should see this:



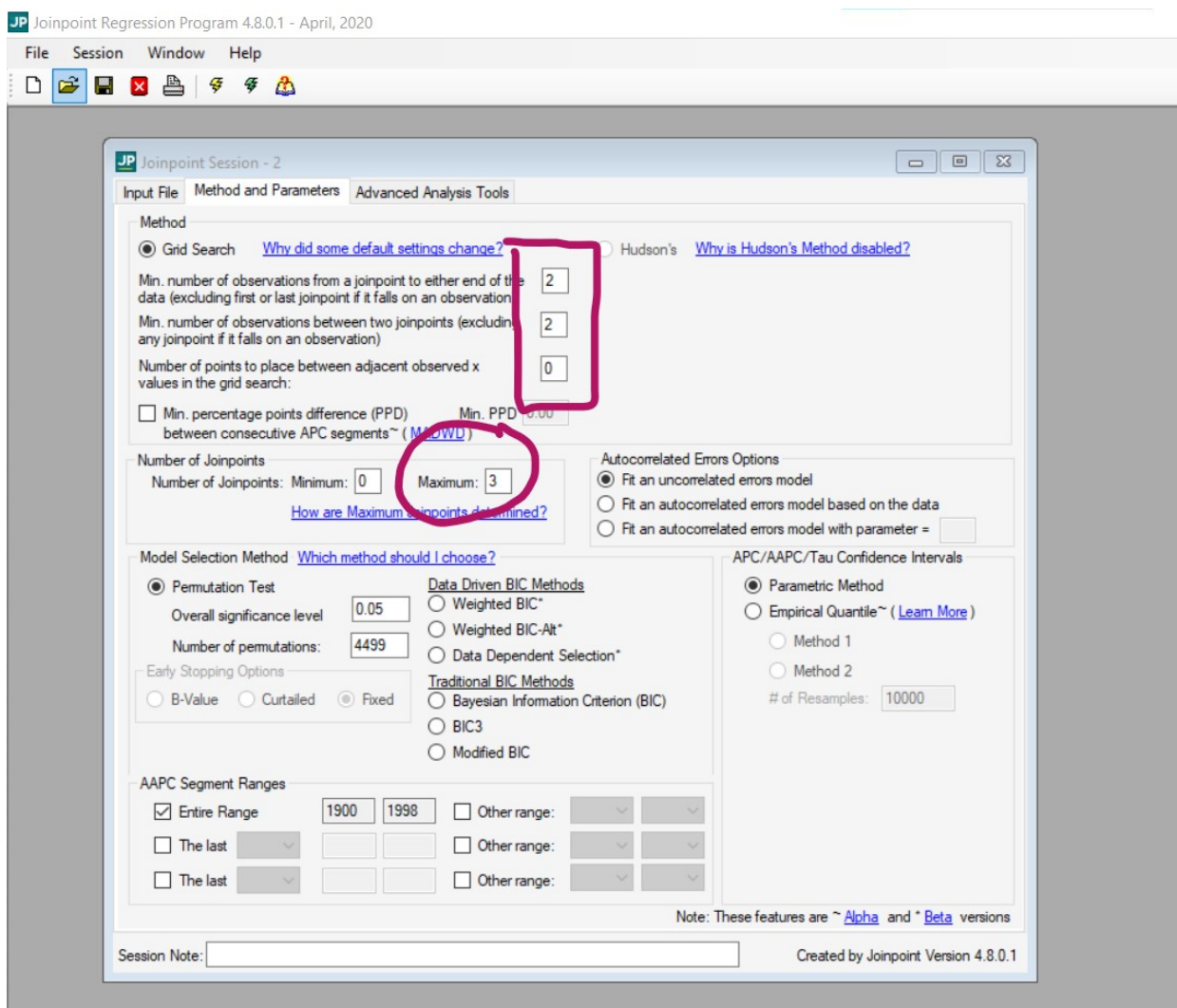
This interface has a lot of information in it. You will see the large box, which holds the information in the data file we've uploaded. Below that, you will have options to choose the column name that holds the age-adjusted rate and standard error. Below these boxes on the same side is a box to choose what kind of standard error we can input or expect. On the right hand side, there is a drop-down menu to choose your dependent variable, a box to choose your co-variants, and finally the option to analyze your data using the log transformation or not.



This is how we will fill out these boxes. Choose “Provided in Data File” for “Run Type”, and for “Type of Variable”, choose “Age-Adjusted Rates”. If you’re working with data that needs to be age-adjusted so that you are comparing apples to apples, make sure you do that either beforehand, or the Joinpoint software can do it for you in another procedure. Since our rates do not require age adjustments, we can select this. That variable in our data set is titled “asdr”, so this is what we select.

For the options for heteroscedastic errors, we choose “Constant Variance (Homoscedasticity)”. This obviously is not a perfect measurement, but it’s a good enough assumption to make. Our “Independent Variable” is “year”, and our “By Variables” is “cod”. Finally, in this procedure we will choose “NO” for “Log Transformation”. We have already analyzed these data using logarithmic transformations in our R code, but we can certainly go through the same procedure here.

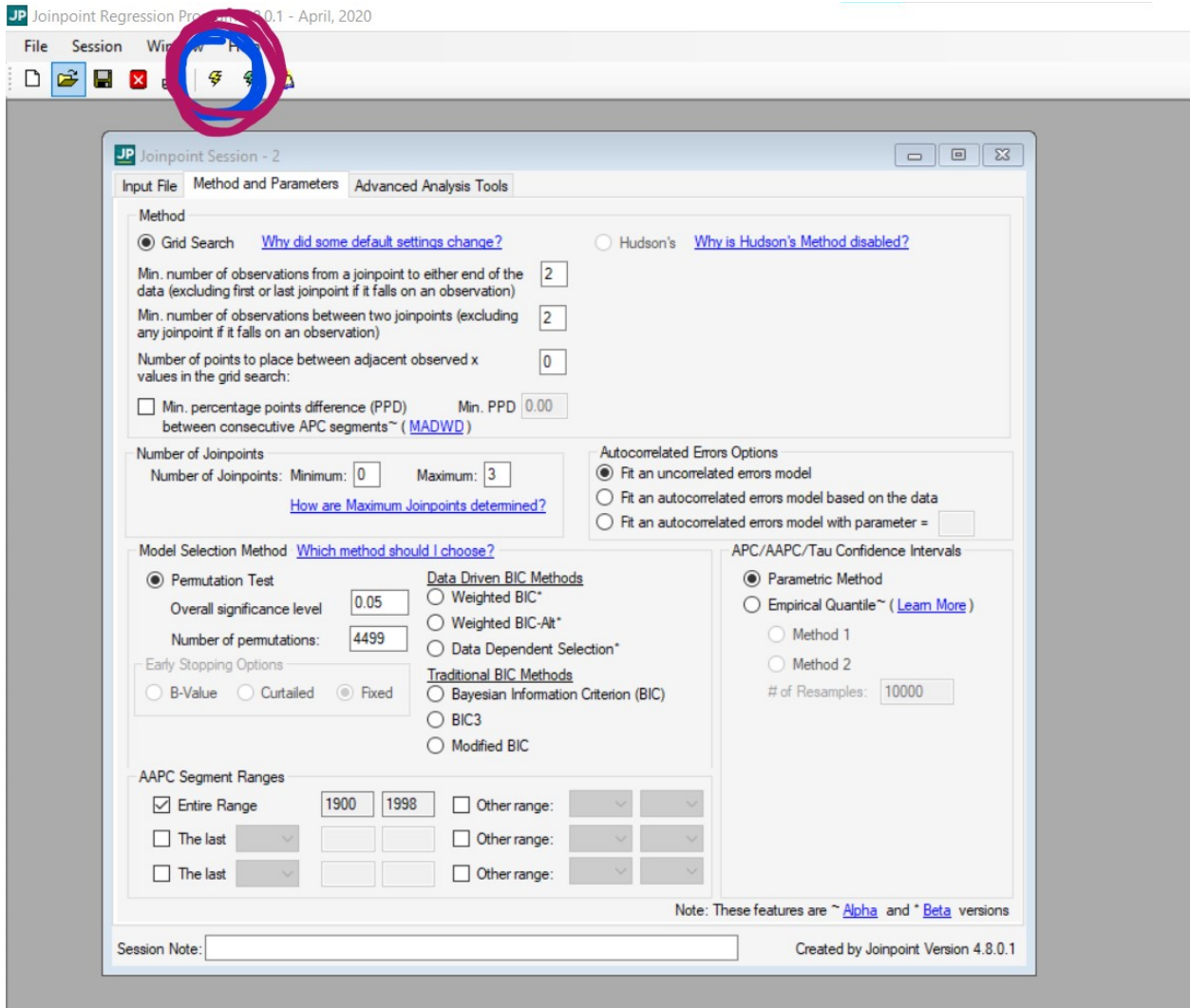
After we have all of these boxes filled in, head up to the tabs at the very top of the dialogue box and choose “Methods and Parameters” so we can tell the program how many joinpoints we want to fit.



The numbers that appear in the pink box above are: “Minimum number of observations from a joinpoint to either end of the data (excluding first or last joinpoint if it falls on an observation)” = 2; “Minimum 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. These are the automatic values provided by the program, and if you think about the assumptions behind each of these parameters, they are good ones. At the ends of the time series and between joinpoints, you want to make sure that there is actually a *trend* being established that you (or, the program) can statistically analyze.

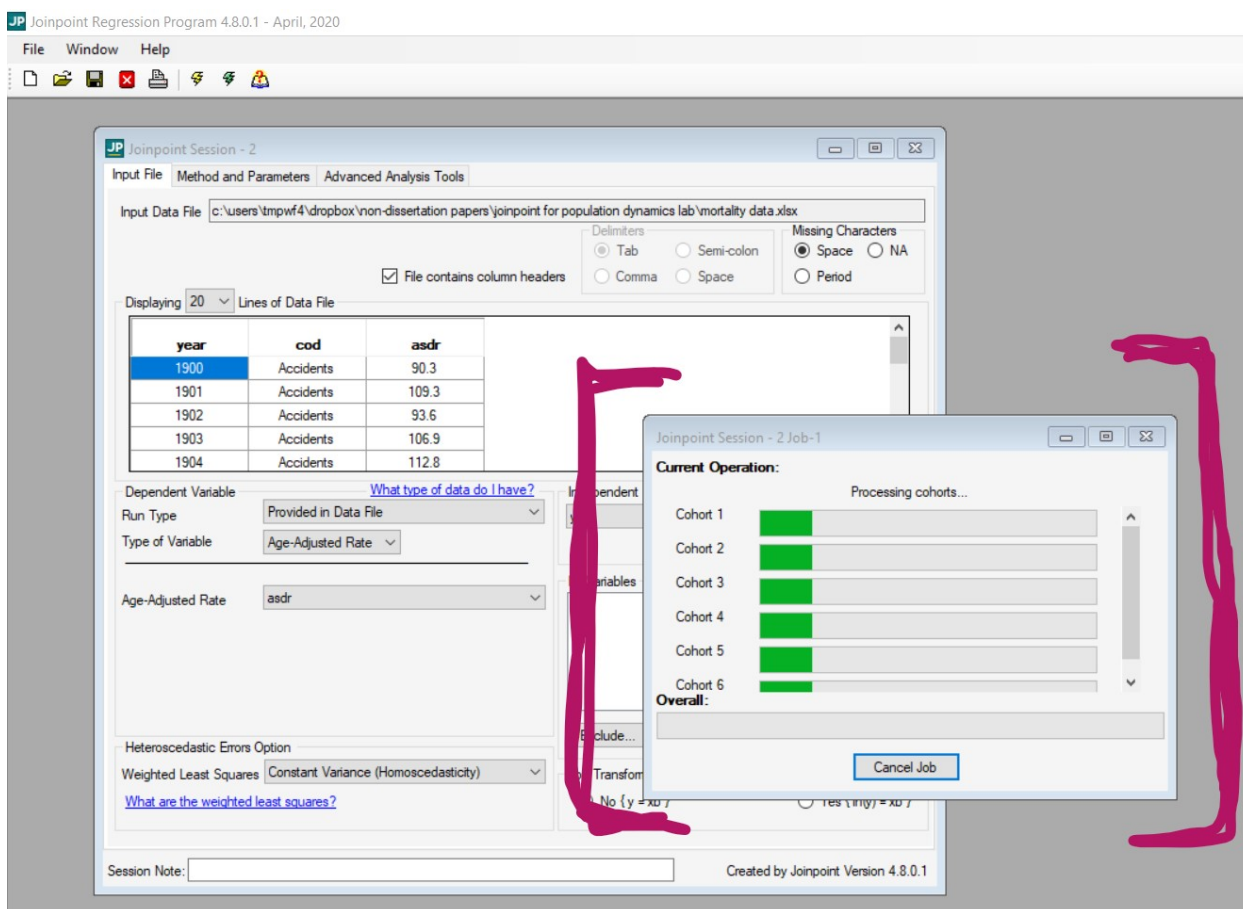
In the pink circle above, I have entered Maximum [number of joinpoints] = 3. This is probably as high as you will want to go with a relatively large data set with a large set of co-variants. The program will take a long time to run, and you have to think critically about what it would mean to ask the program to fit a large number of joinpoints to your data. While it would certainly do it, you run the risk of overfitting your data and getting results that are not meaningful anymore. You can always come back and change this parameter later.

All the rest of the parameters are automatically filled in, and we don’t need to change them.



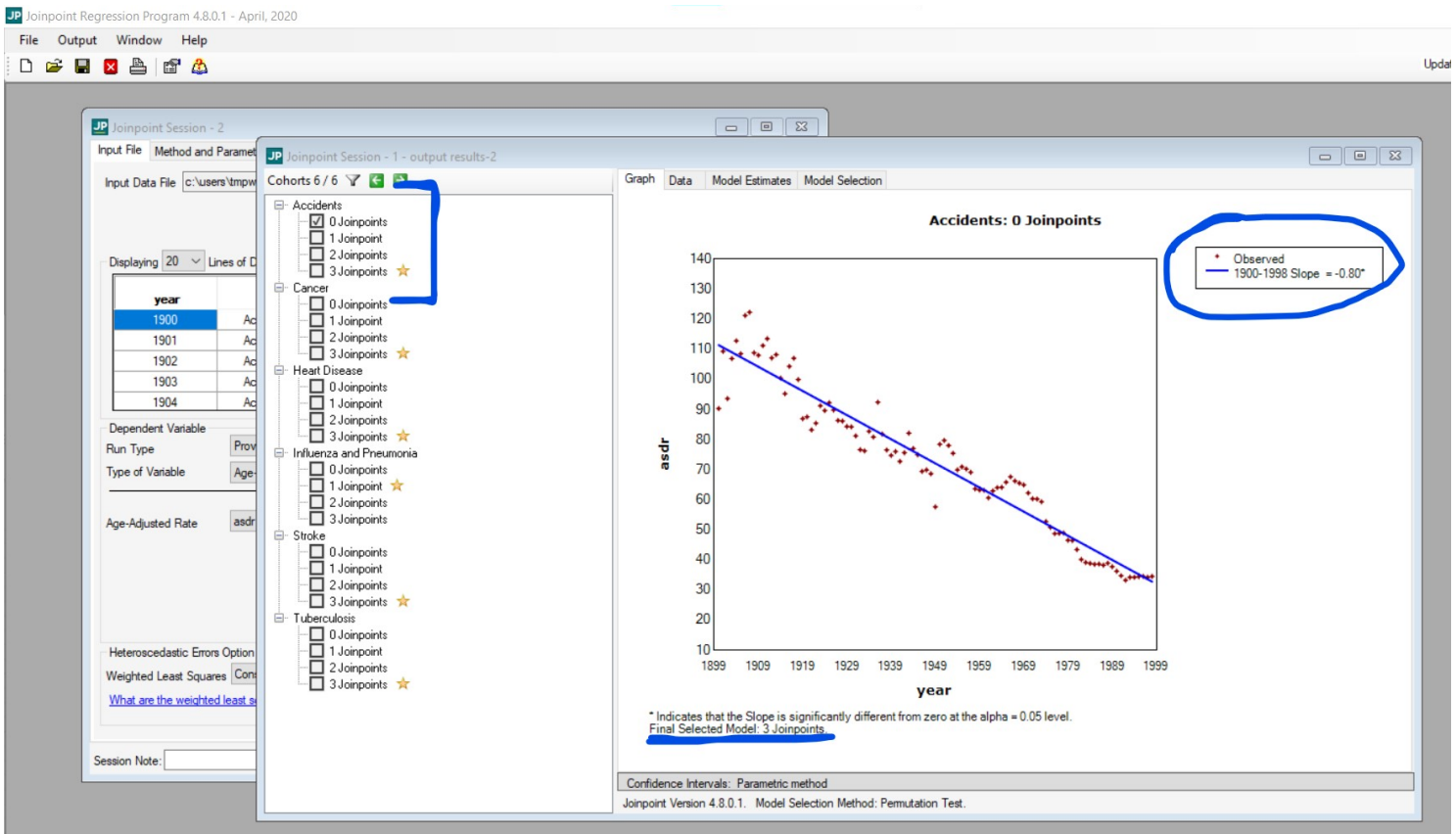
When everything is filled out to your liking, click the yellow lightning bolt, as shown above in the blue and pink circles. There is a small blue bolt right next to the yellow bolt, but that one will run the model and ignore any errors that pop up. We want to see if there are errors, so click the yellow lightning.

The next thing you will see is that your model is running!



In the new dialogue box that I have outlined with pink square brackets, you should see progress bars for each of the “cohorts”—those are your causes of death that we’re modeling. This could take a while, depending on the size of the data (the length of the time series) and/or the number of co-variables you’re modeling.

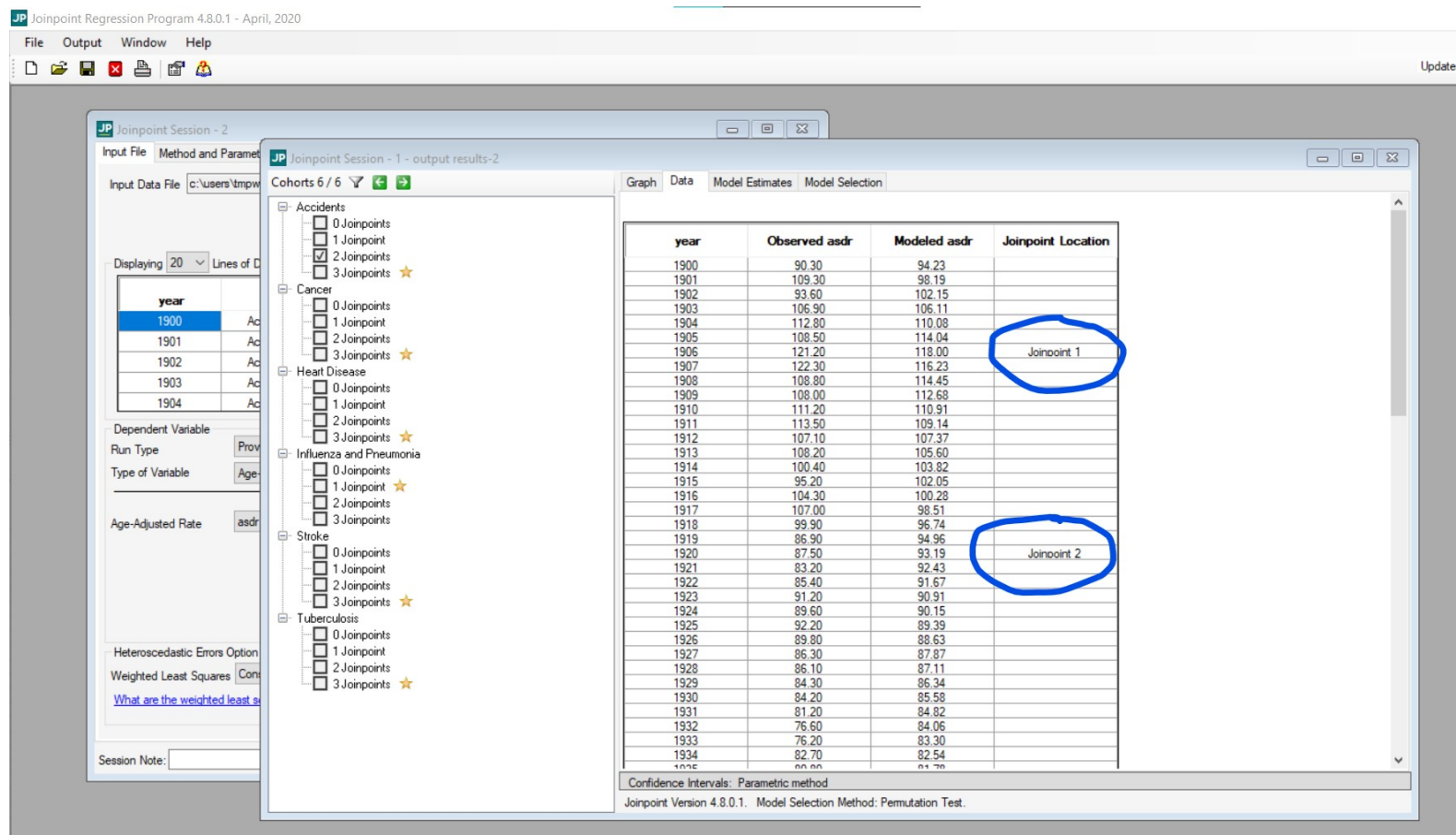
When it’s all finished running, what you should see is another brand new dialogue box with results!:



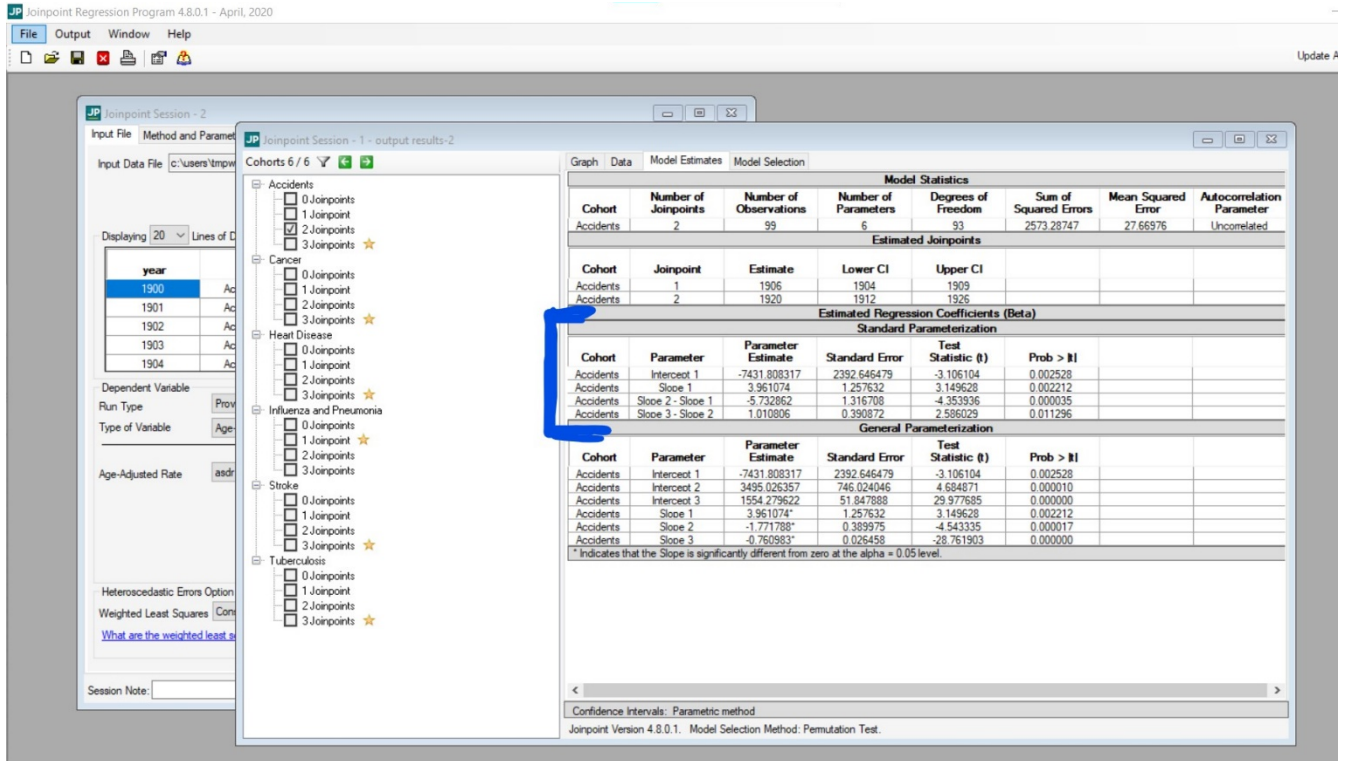
These are what your results will look like. On the left-hand side, there is a list of all the causes of death modeled and all of the models you've asked it to fit: 0 joinpoints through 3 joinpoints for each cause of death. You should also see a star next to one of the models for each cause of death—that is your best fit model!

The large window in this dialogue box is dedicated to the visualization of the data and the currently-activated model (back to the left-hand size real quick: those boxes are clickable, and you can go through each one to see the data and the model fit visualized in the large box). On the right hand side in what looks like the legend, it holds the segment of the time series for the modeled slope and if that model is significant (it is!).

At the bottom of the box, where I have underlined in blue, is the best fit model for this set of data. So for "Accidents", the best fit model has 3 joinpoints.



At the very top of the results dialogue box is a set of tabs. We were previously on the “Graph” tab, but now click the “Data” tab to see all the data. The program very conveniently shows us the year, observed asdr (what we input into the model), the modeled asdr, and the joinpoint location in the time series. So here, we are looking at the data for Accidents with 2 joinpoints (as you can see, we have checked that box on the left hand side). There is an estimated joinpoint in 1906, and a second estimated joinpoint in 1920 (circled in blue).



In the “Model Estimation” tab, this is where you will find all of the coefficients for the model that is activated in the list on the left-hand side. We are still looking at Accidents with 2 joinpoints estimated. In the Standard Parameterization, there is a coefficient estimate for the (1) intercept, (2) slope before any identified joinpoints, (3) slope between the first and second joinpoint, and (4) slope after second joinpoint to the end of the time series. It also has information on the standard error and the significance of those fitted points. Remember, the slopes are representative of the **Annual Percent Change** in rate.

JP Joinpoint Regression Program 4.8.0.1 - April, 2020

File Output Window Help

Update Available!

JP Joinpoint Session - 1 - output results-2

Cohorts 6 / 6

Accidents

Cancer

Heart Disease

Influenza and Pneumonia

Stroke

Tuberculosis

Model Selection Method

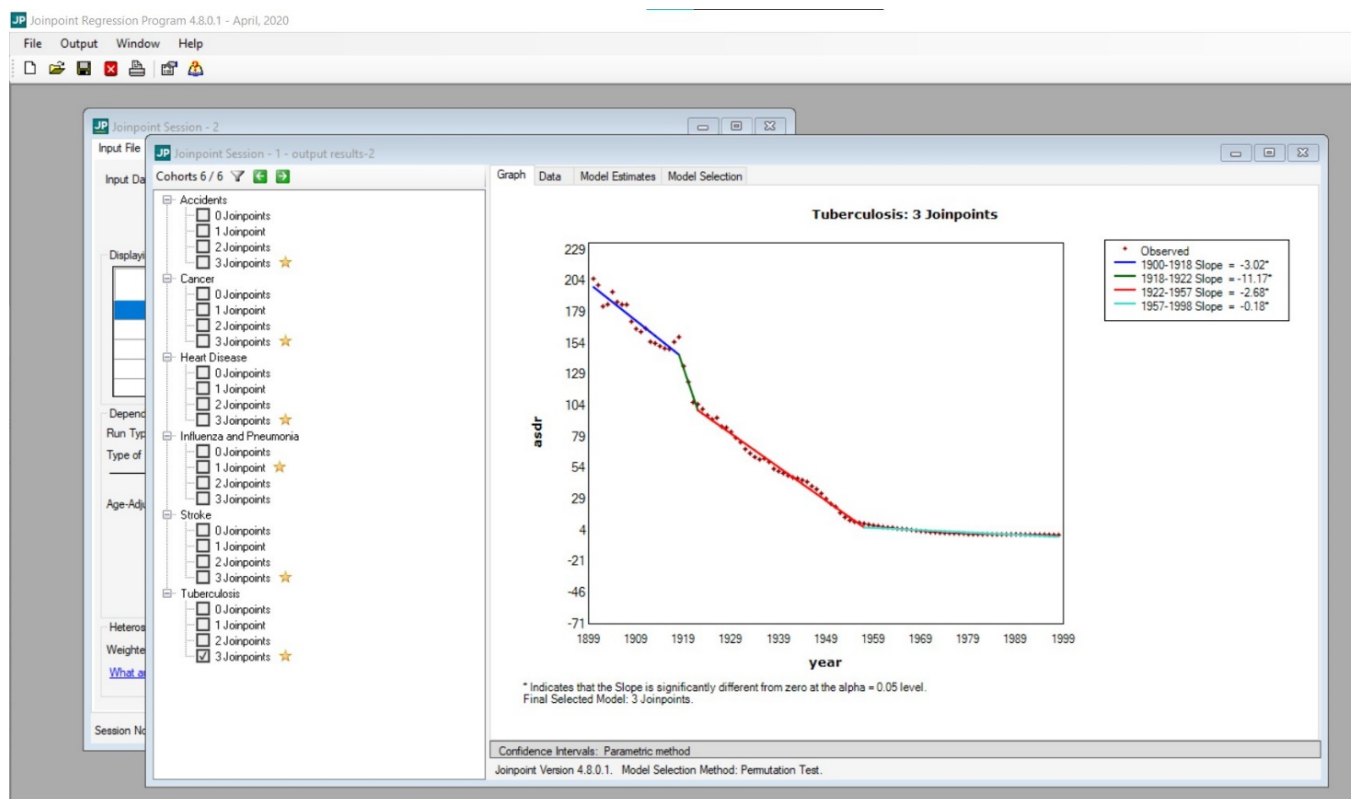
Cohort	Model Selection Method	Test For Number of Joinpoints	Test Number	Null Hypothesis	Alternate Hypothesis	Numerator Degrees of Freedom	Denominator Degrees of Freedom	Number of Permutations	P-Value	Significance Level
Accidents	Permutation Test		#1	0 Joinpoints	3 Joinpoint(s) ^	6	91	4500	0.0002222	0.0166667
Accidents			#2	1 Joinpoint	3 Joinpoint(s) ^	4	91	4500	0.0002222	0.0250000
Accidents			#3	2 Joinpoints	3 Joinpoint(s) ^	2	91	4500	0.0002222	0.0500000

^ Selected Model
 ~ Significance level for individual test (alpha = 0.05).
 Final Selected Model: Accidents - 3 Joinpoint(s)

Confidence Intervals: Parametric method

Joinpoint Version 4.8.0.1. Model Selection Method: Permutation Test.

In the “Model Selection” tab, this is where you can find the statistical significance test results that were automatically performed on all model fits for that specific cause of death to see what model is the best fit for the data. For Accidents, the three joinpoint model is clearly the best fit model, as seen by the comparisons I’ve highlighted with the blue circles.



You can go through and investigate all of the data and the models that we have fit (there are a total of 24!). Above, I've taken a screenshot of the best model fit for tuberculosis from 1900-1998. As you can see, the three joinpoint model is the best fit, and in the legend on the right-hand side, you can see the results of the model and the slopes of the lines between joinpoints. As a brief interpretation of this best fit model for tuberculosis:

- From 1900-1918, tuberculosis mortality decreased ~3.0% per year
- From 1918-1922, mortality decreased ~11.1% per year
- From 1922-1957, mortality decreased ~2.8% per year
- From 1957-1998, mortality decreased ~0.1% per year

What is important to point out here is that the model fit for tuberculosis actually starts to creep below zero the later we get into the 20th century. This is where you have to think critically about what the statistical results are actually telling you: we know tuberculosis mortality is decreasing slowly by this time, and the mortality rate is already very, very low. We know that mortality rates cannot be less than zero. So what this likely means is that the US has about as low of a tuberculosis mortality rate in the 1990s and later as it can possibly be.