Chapter 16: Analysis of covariance

TXCL7565/PHSC7565

What This Lecture Covers

- Overview of ANCOVA
- ANCOVA in GraphPad Prism
- Advantages of ANCOVA over a simple two-sample t-test

OVERVIEW OF ANCOVA

Analysis of Covariance (ANCOVA)

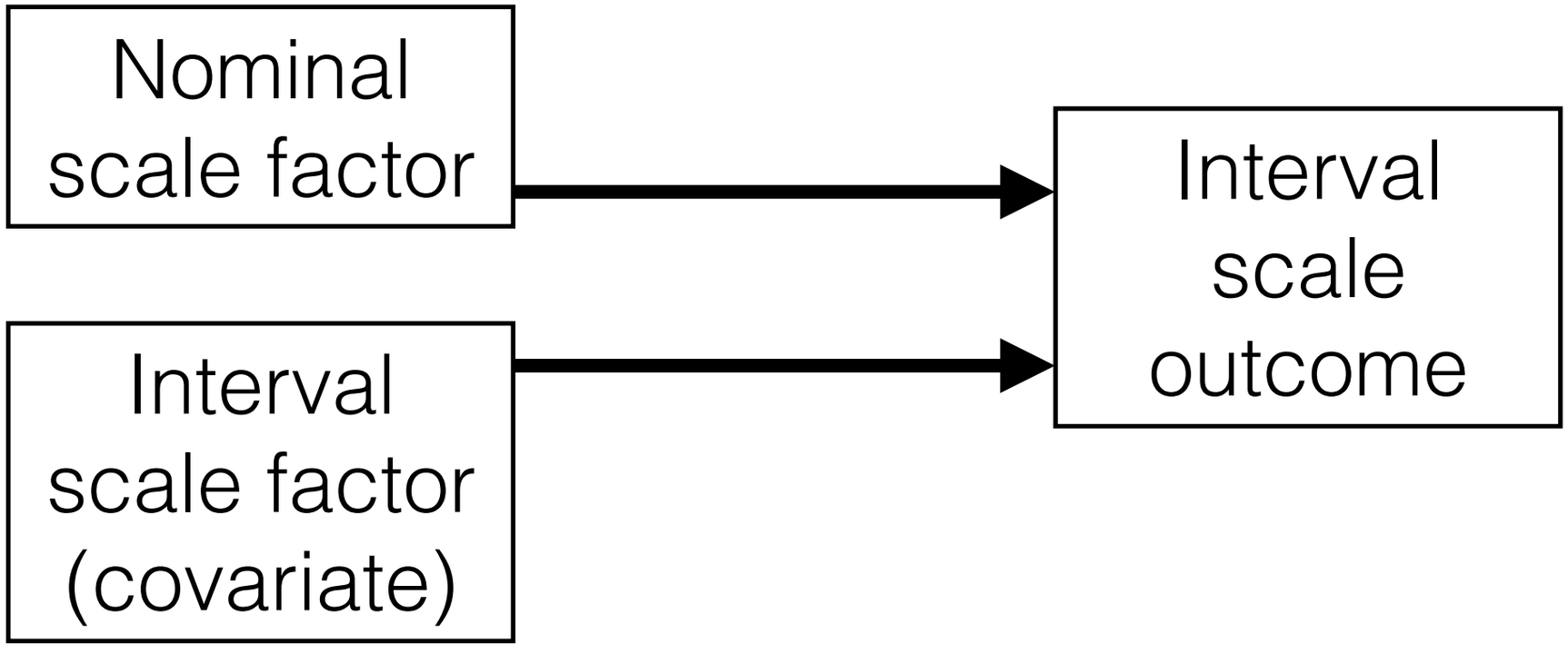


Figure 16.1 Structure of a study to be analyzed by analysis of covariance (ANCOVA)

Covariates

- “Covariate is any factor (independent variable) which would be recorded on an interval scale” - Essential Statistics for Pharmaceutical Sciences

- “Continuous variables such as these, that are not part of the main experimental manipulation but have an influence on the dependent variable, are known as covariates” - Discovering Statistics

ANCOVA and ANOVA

- In both models, we have a main effect for each factor (nominal and covariate) and an interaction effect.

- In ANCOVA, the interaction effect is used to estimate the difference in slopes (i.e., the difference in the quantitative relationship between the dependent variable and the continuous covariate).

ANCOVA and multiple regression

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{interaction} X_1 \times X_2$$

X_1 is the nominal factor
X_2 is the continuous covariate

General interpretation of ANCOVA results

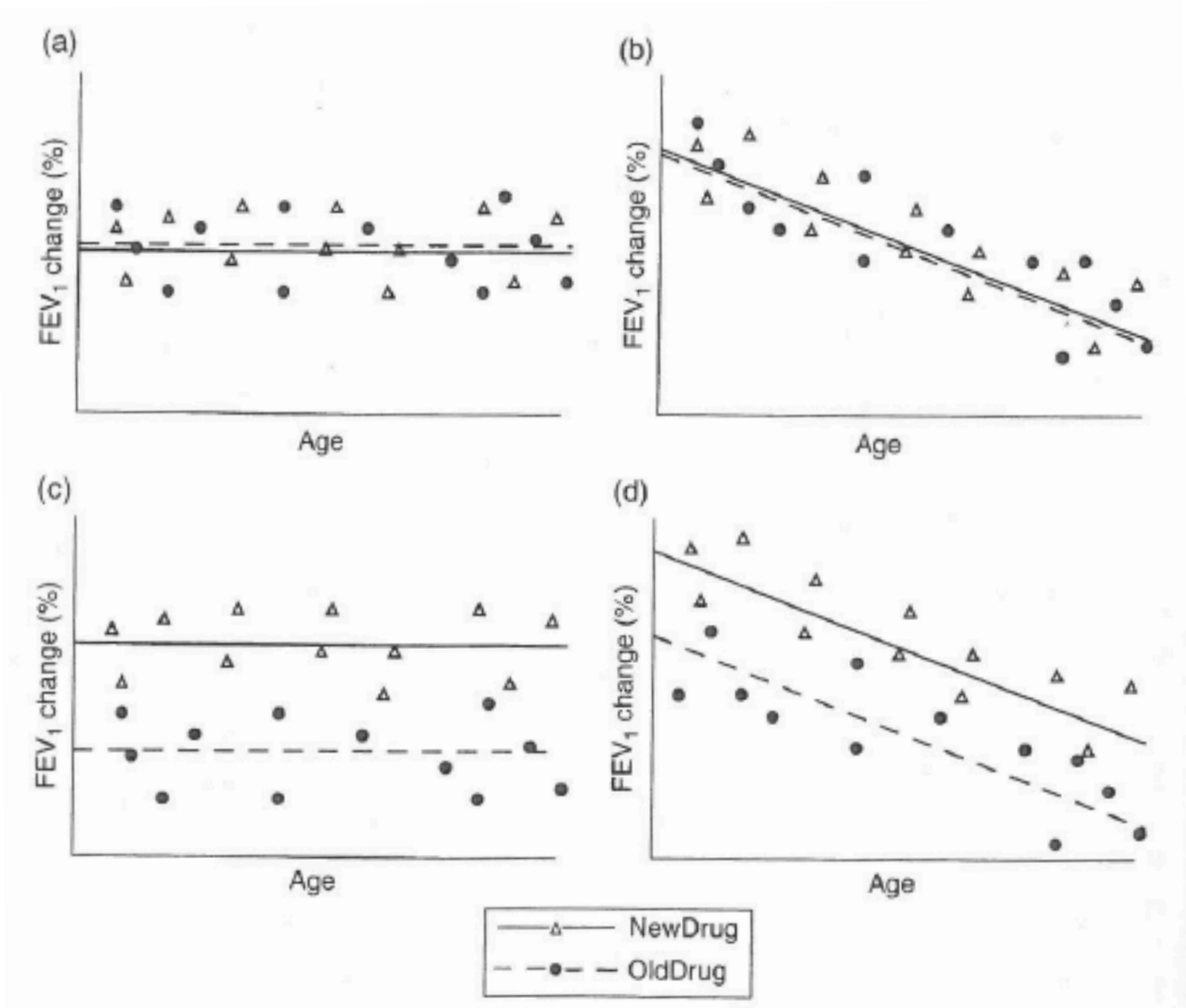
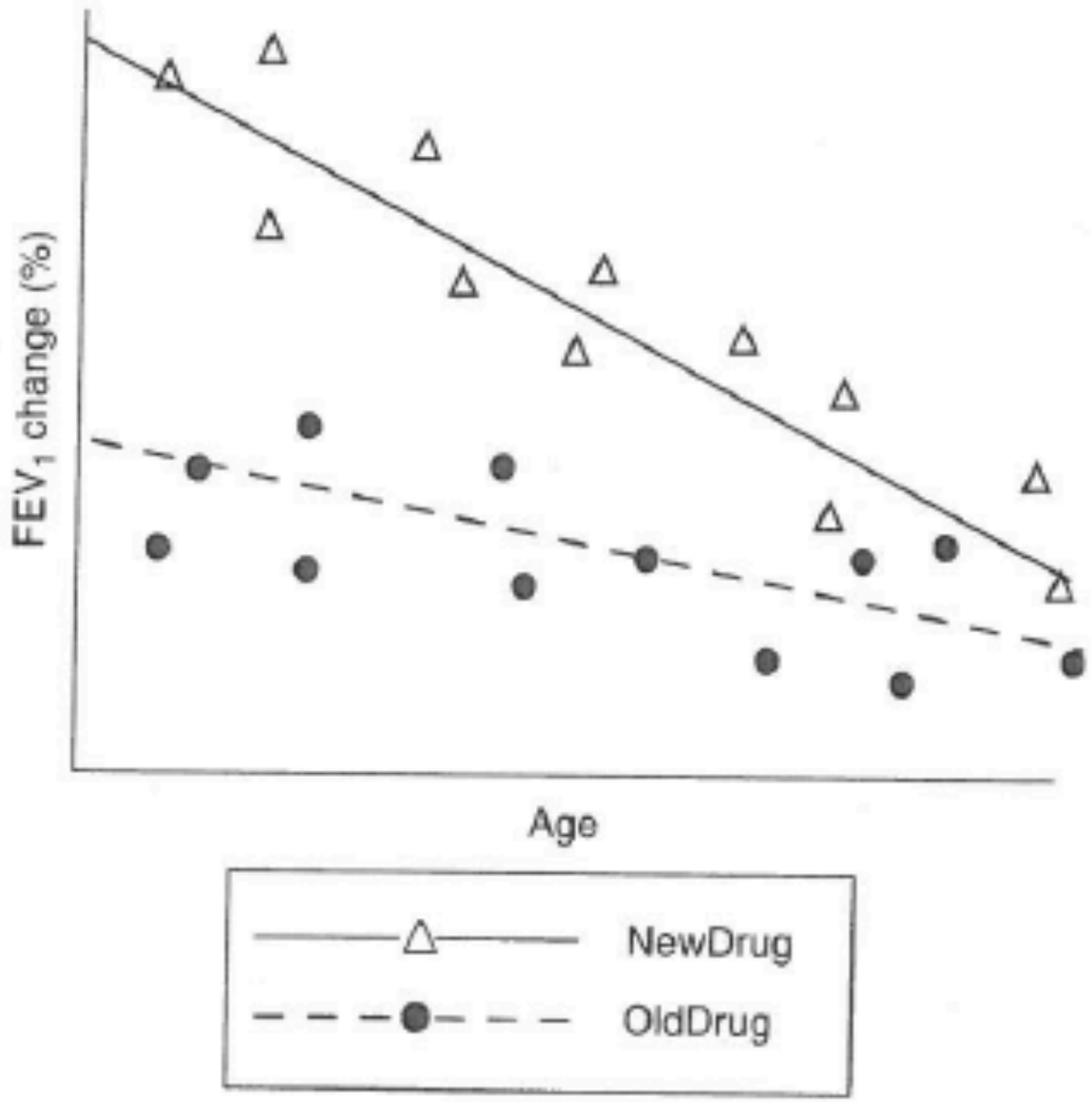


Figure 16.2 Four hypothetical outcomes for the COPD treatment trial (a, b, c, d)

Example of an interaction



Typical stages to ANCOVA

- Initial stage - Fit separate lines (dependent vs. continuous covariate) for each level of the nominal factor

    - If the interaction effect is significant than the gradient (or slope) of the lines are different and need to be modeled as such.

    - If the interaction effect is non-significant, then the gradient (or slope) of the lines are not statistically different and a ‘common slopes model’ should be used to determine difference between levels of the nominal factor

- Common slopes model - Model the data using parallel lines for each level of the nominal factor, i.e., estimate a common gradient (slope) for all levels of the factor.

COPD Trial Example

- Goal - compare two treatments for chronic obstructive pulmonary disease (COPD)

- Patients were randomly chosen from 40-80 year old males with COPD treated at National Jewish in 2015

- Patients were randomly assigned to two treatment groups

- Outcome variable - percent change in FEV_1

- Nominal factor - treatment (new vs. old drug)

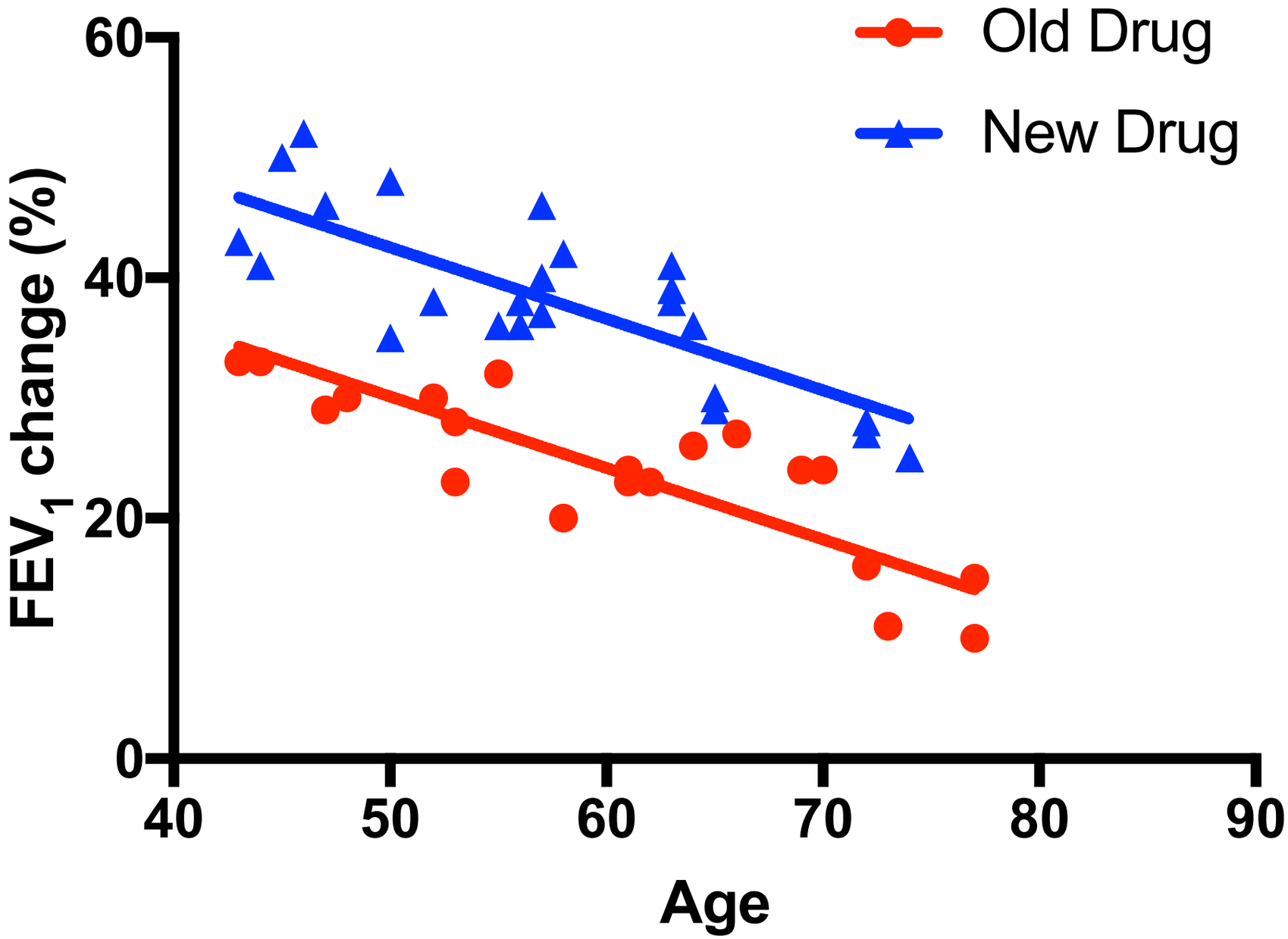
- Continuous covariate - age (not of primary interest, only included to improve analysis)

Type of study?

Can you make cause and effect conclusions?

What is the scope of inference for this study?

COPD Trial



Initial stage - COPD trial

Table 16.2 Generic example of part of the output of an initial ANCOVA for the COPD trial data. An interaction term was included to allow for differing slopes in the lines fitting the data for the two treatments

Factor	Coefficient	P value
Treatment	19.295	0.014
Age	-0.536	<0.001
Treatment*Age interaction	-0.117	0.356

The effect of treatment on FEV1 is not dependent on age (p-value = 0.36).

Common slopes model - COPD trial

Table 16.3 Generic example of part of the results of a common slopes ANCOVA for the COPD trial data. Interaction term has been excluded.

Factor	Coefficient	P value
Treatment	12.405	<0.001
Age	-0.594	<0.001

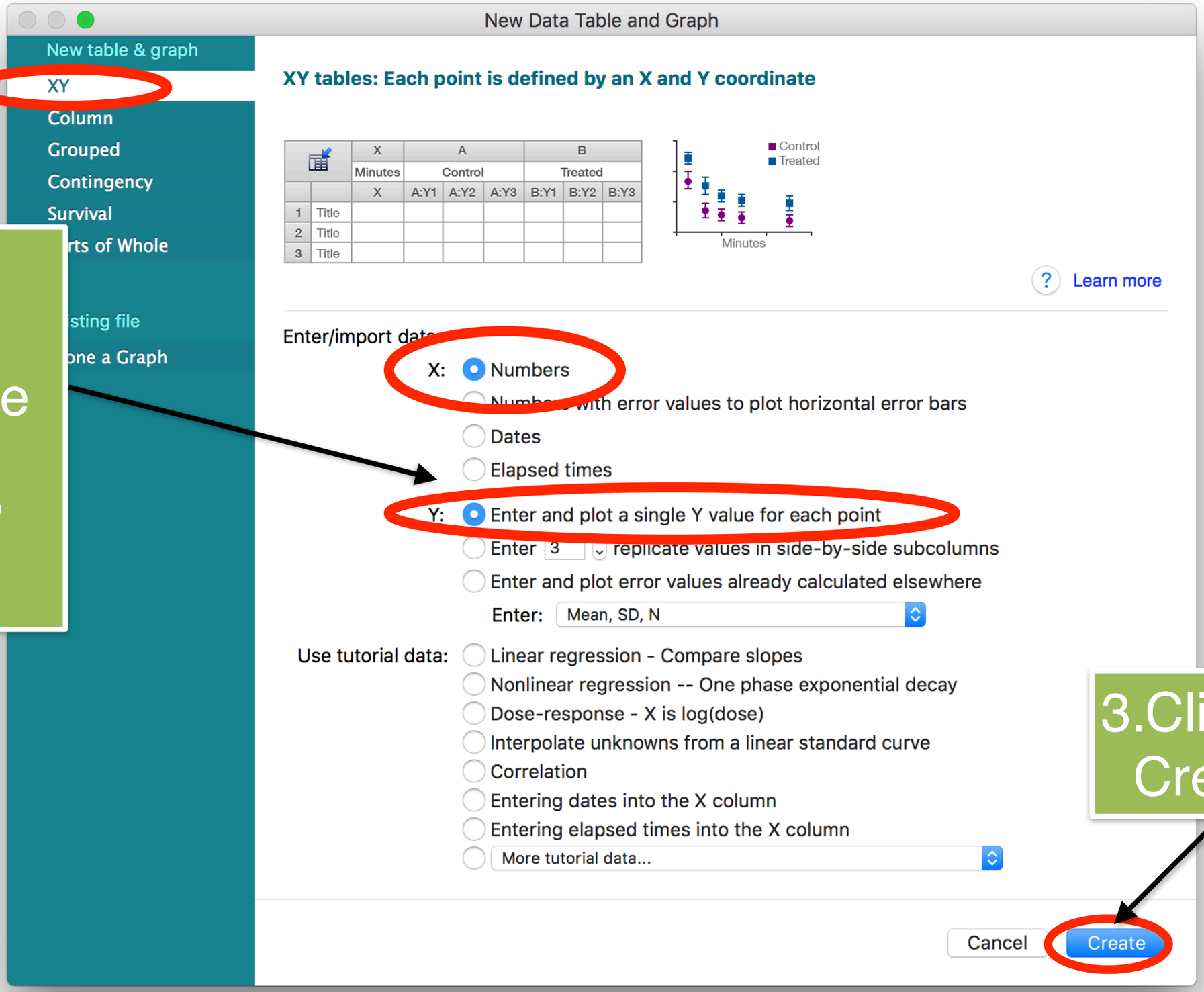
The new drug is associated with a significantly higher FEV_1 (estimated mean difference = 12.41, p-value < 0.001) than the FEV_1 in patients who received the old drug. A year increase in age is associated with a 0.59% decrease in FEV_1 (p-value <0.001).

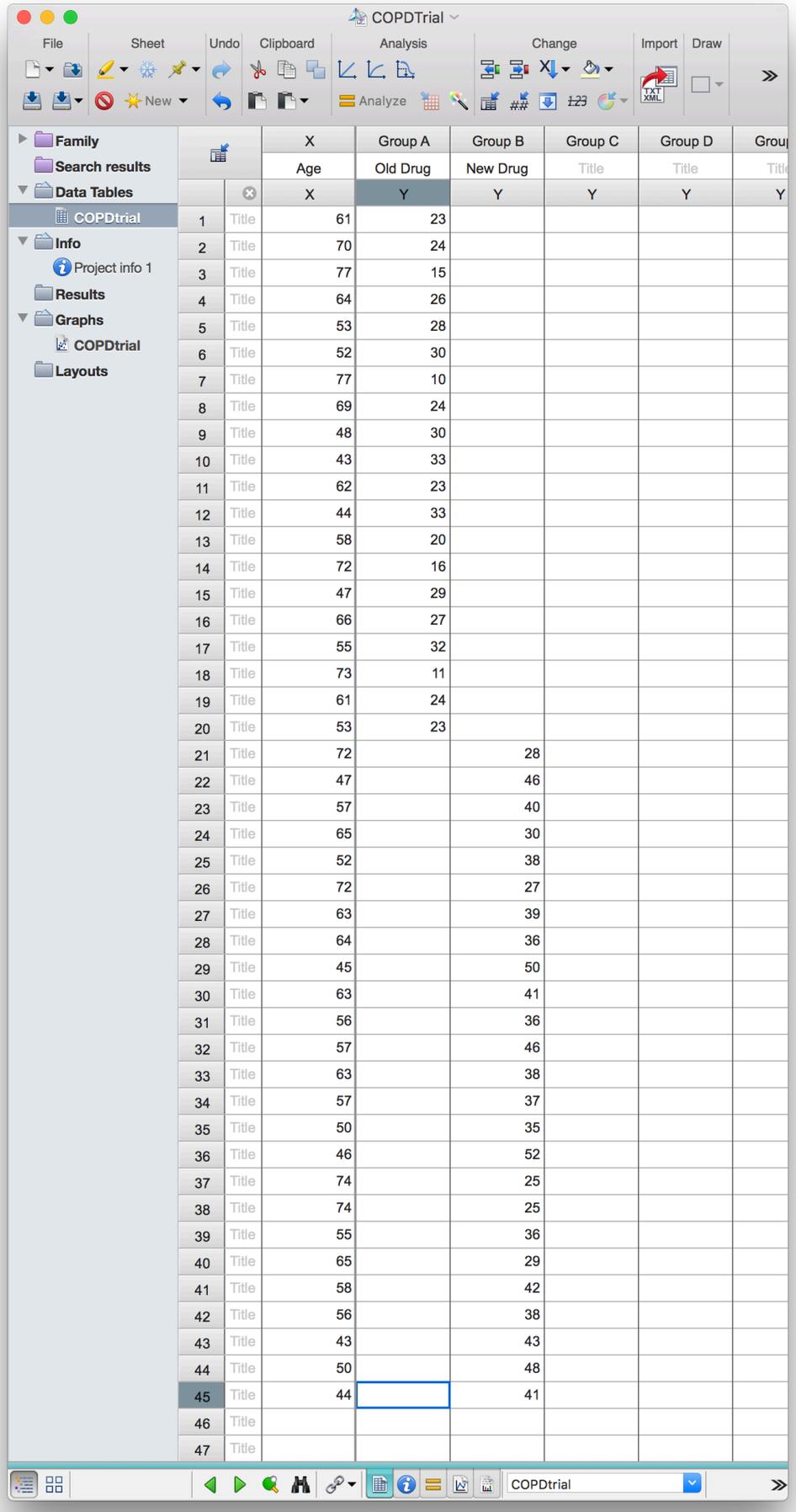
ANCOVA IN GRAPHPAD PRISM

1. Select ‘XY’ from New table & graph

2. Select ‘Enter and plot a single Y value for each point’ for Y:

3. Click Create





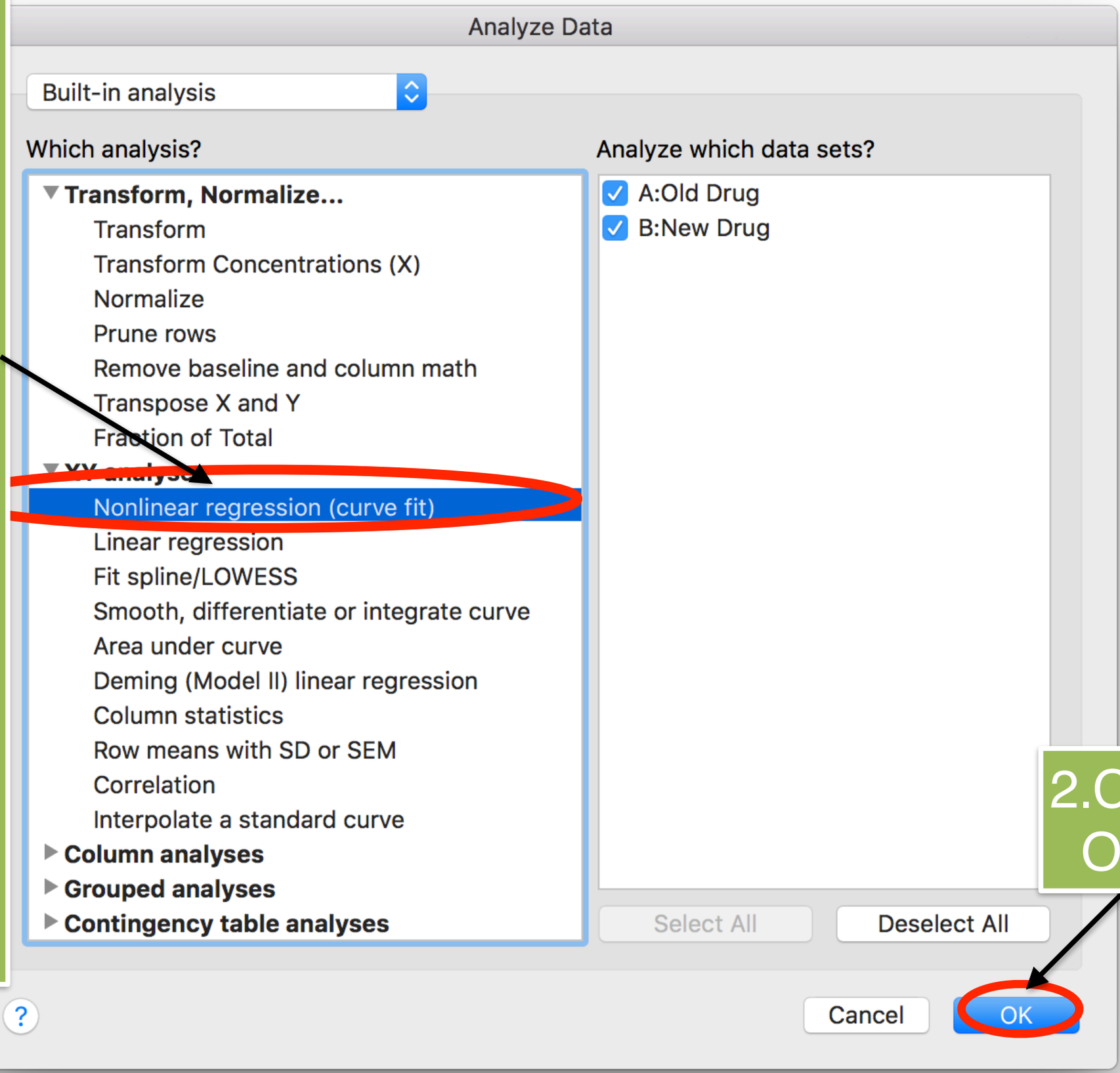
Enter the continuous covariate for each observation in the X column.

For observation from the first level of the nominal factor, enter the outcome variable in the first Y column.

For observations from the second level of the nominal factor, enter the outcome variable value in the second Y column.

And so on…

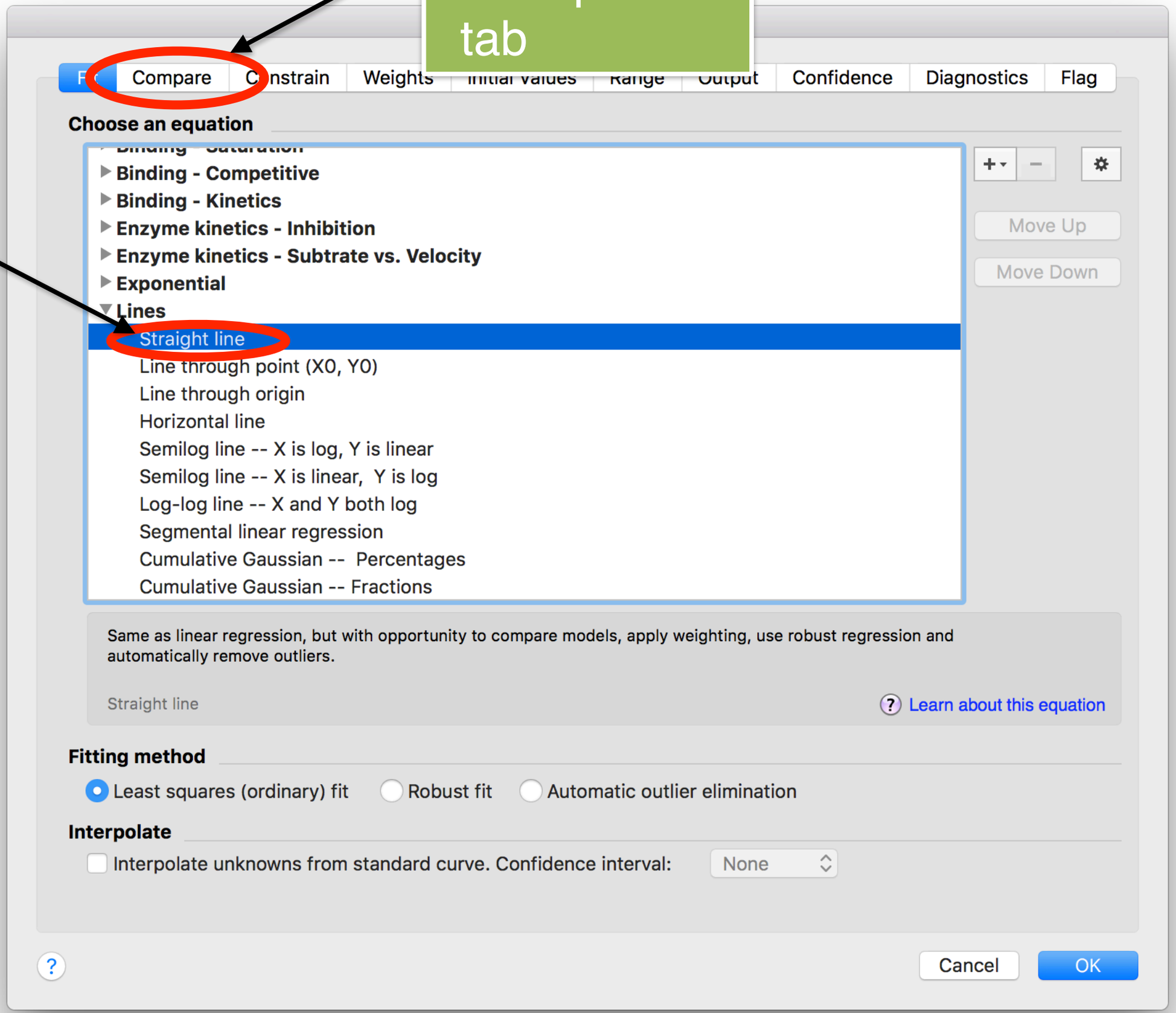
1.Select ‘Nonlinear regression’ from XY analyses. The nonlinear regression options allows for linear regression with more options for control of the model than the ‘Linear regression’ choice.



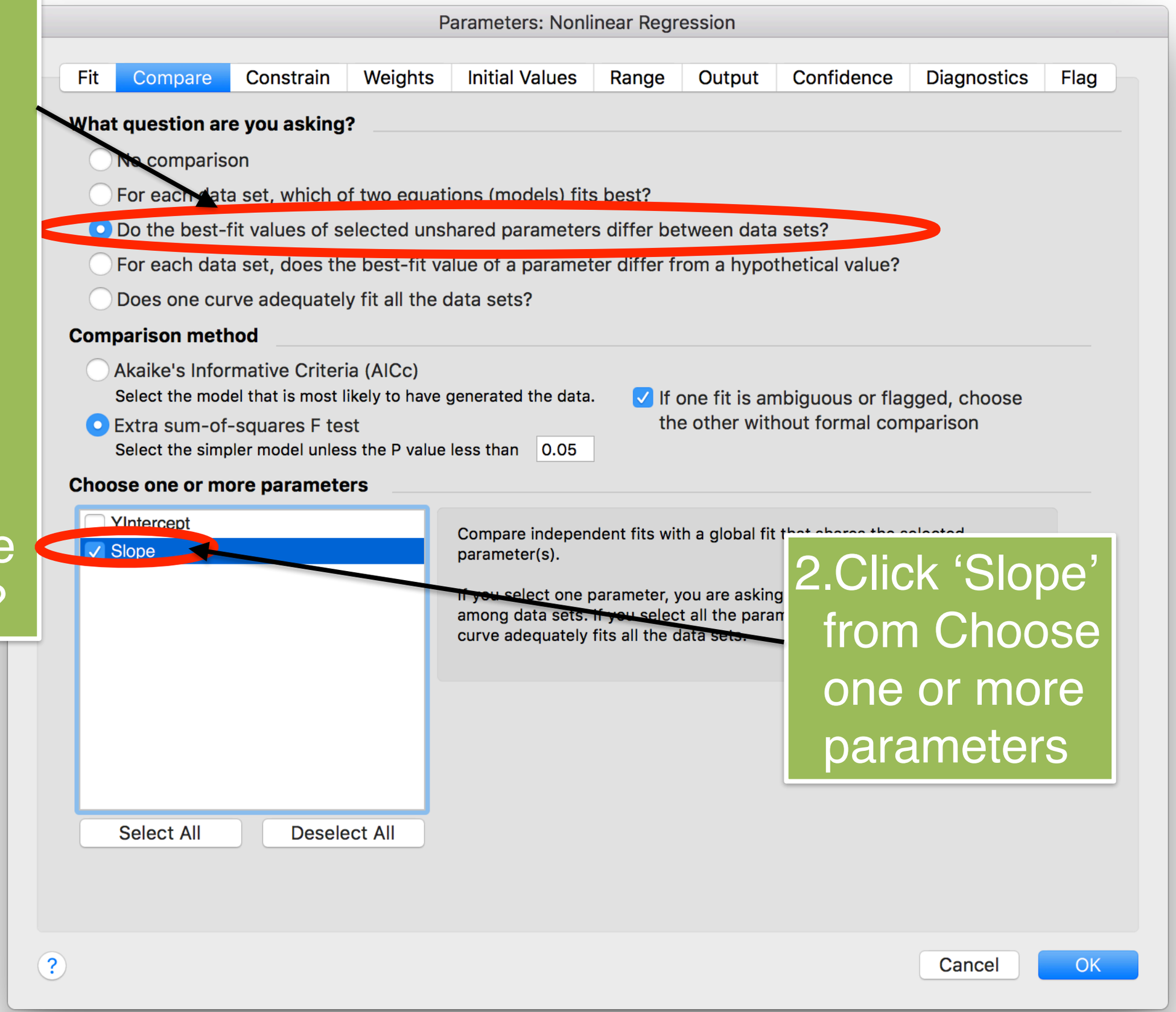
2.Click OK

1. Select ‘Straight line’ from Lines

2. Select ‘Compare’ tab



1.Select ‘Do the best-fit values of selected unshared parameters differ between data sets?’ from What question are you asking?



2.Click ‘Slope’ from Choose one or more parameters

COPDTrial — Edited

File Sheet Undo Clipboard Analysis Interpret Change Draw Write Text Export

New Analyze 10 Arial

- Family
- Search results
- Data Tables
  - COPDtrial
- Info
  - Project info 1
- Results
  - Nonlin fit of COP
- Graphs
  - COPDtrial
- Layouts

	Nonlin fit	A	B	C	D	E
		Old Drug	New Drug	Global (shared)	Title	Title
		Y	Y	Y	Y	Y
1	Comparison of Fits					
2	Null hypothesis			Slope same for all data sets		
3	Alternative hypothesis			Slope different for each data set		
4	P value			0.3564		
5	Conclusion (alpha = 0.05)			Do not reject null hypothesis		
6	Preferred model			Slope same for all data sets		
7	F (DFn, DFd)			0.87 (1, 41)		
8						
9	Slope different for each data set					
10	Best-fit values					
11	YIntercept	56.37	75.67			
12	Slope	-0.5364	-0.6531			
13	Std. Error					
14	YIntercept	4.924	5.528			
15	Slope	0.08054	0.09424			
16	95% CI (profile likelihood)					
17	YIntercept	46.03 to 66.72	64.23 to 87.1			
18	Slope	-0.7056 to -0.3673	-0.848 to -0.4581			
19	Goodness of Fit					
20	Degrees of Freedom	18	23			
21	R square	0.7114	0.6762			
22	Absolute Sum of Squares	254.2	439.6			
23	Sy.x	3.758	4.372			
24						
25	Slope same for all data sets					
26	Best-fit values					
27	YIntercept	59.86	72.27			
28	Slope	-0.5944	-0.5944	-0.5944		
29	Std. Error					
30	YIntercept	3.871	3.708			
31	Slope	0.06242	0.06242	0.06242		
32	95% CI (profile likelihood)					
33	YIntercept	52.05 to 67.68	64.79 to 79.75			
34	Slope	-0.7204 to -0.4684	-0.7204 to -0.4684	-0.7204 to -0.4684		
35	Goodness of Fit					
36	Degrees of Freedom			42		
37	R square	0.7031	0.6707	0.8372		
38	Absolute Sum of Squares	261.6	447	708.5		
39	Sy.x			4.107		
40	Constraints					
41	Slope	Slope is shared	Slope is shared			
42						
43	Number of points					
44	# of X values	20	45			
45	# Y values analyzed	20	25			
46						
47						

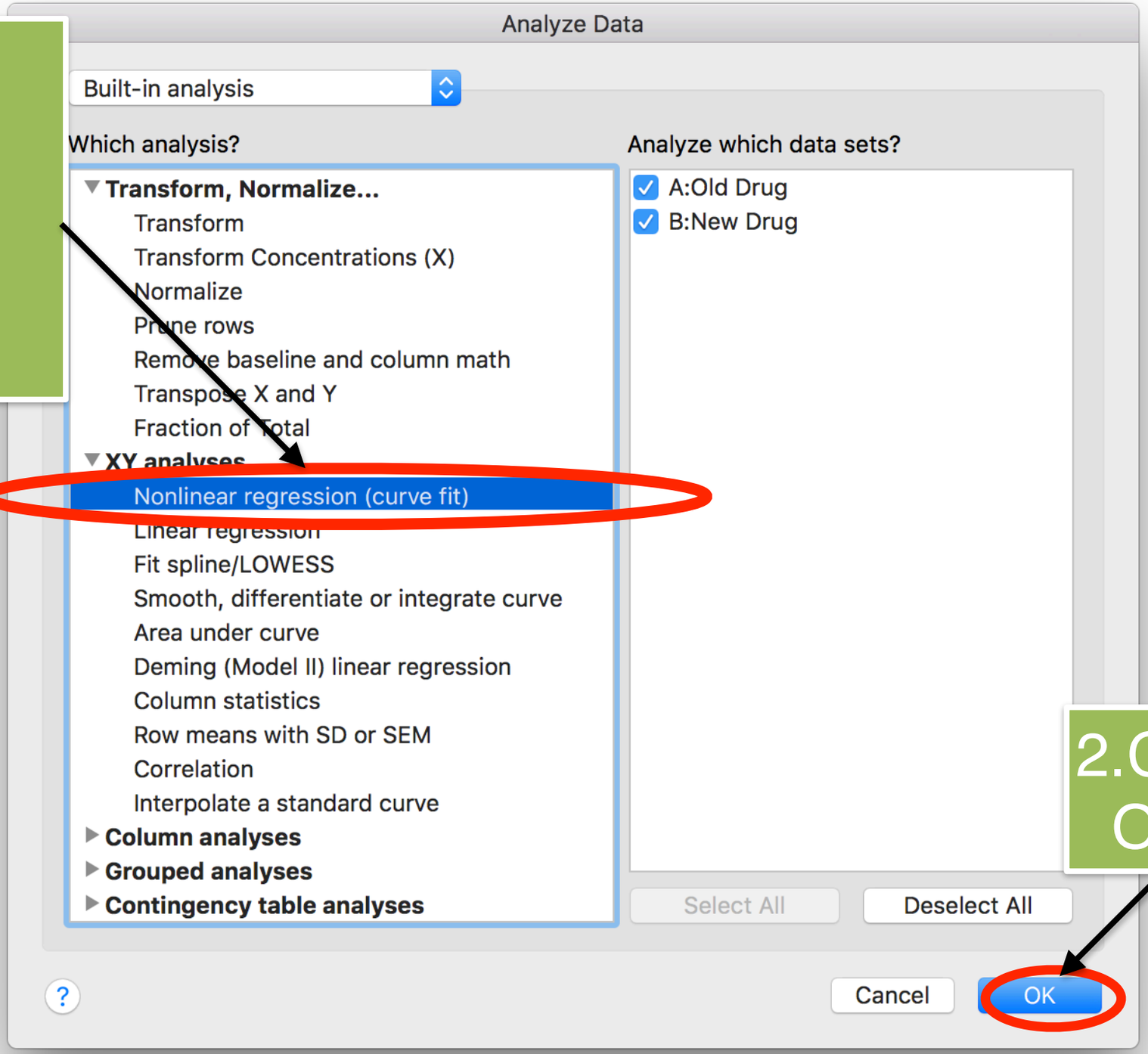
Straight line

Equation help

Nonlin fit of COPDtrial | Table of results

Common slopes model

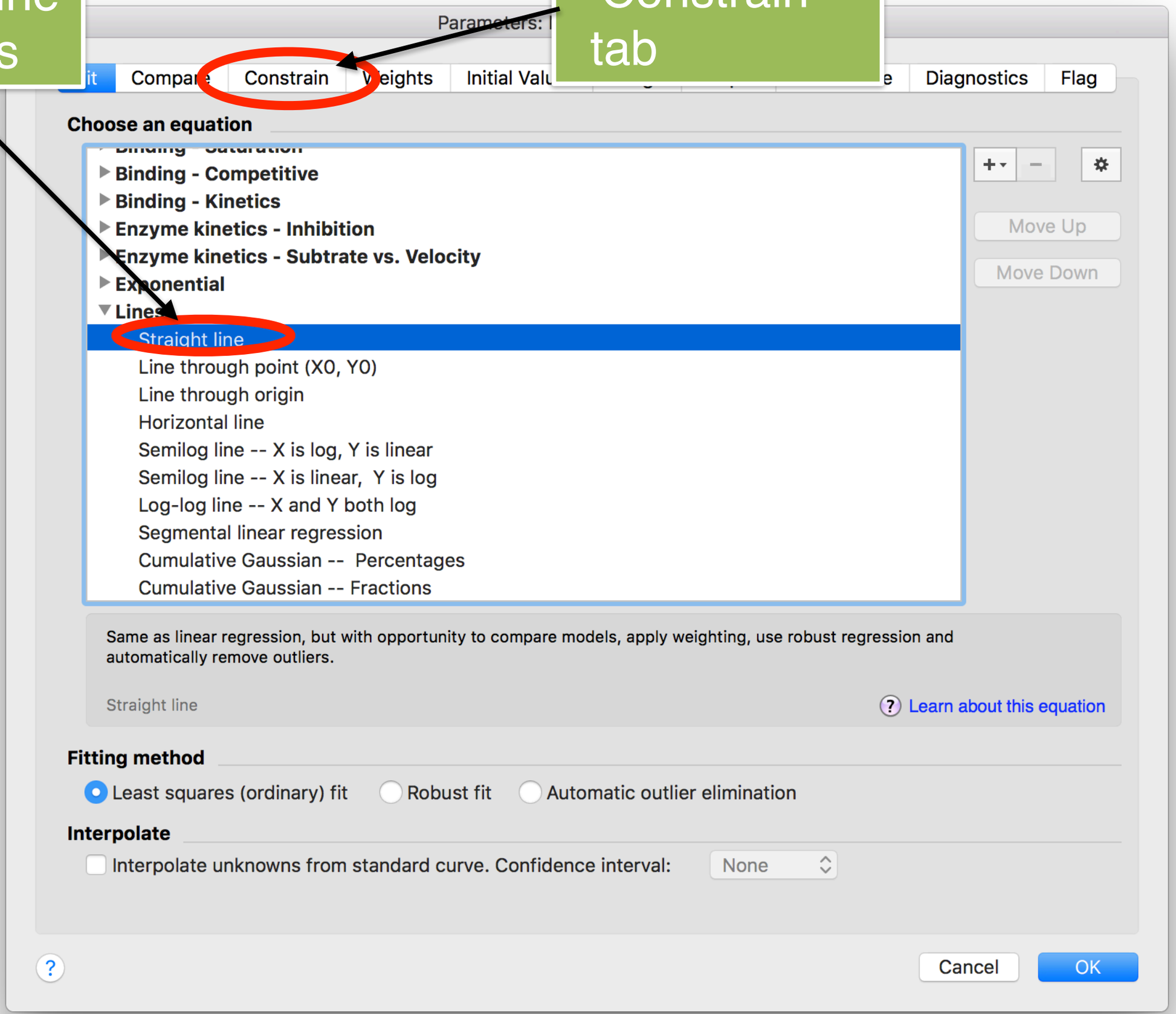
1. Select ‘Nonlinear regression’ from XY analyses.



2. Click OK

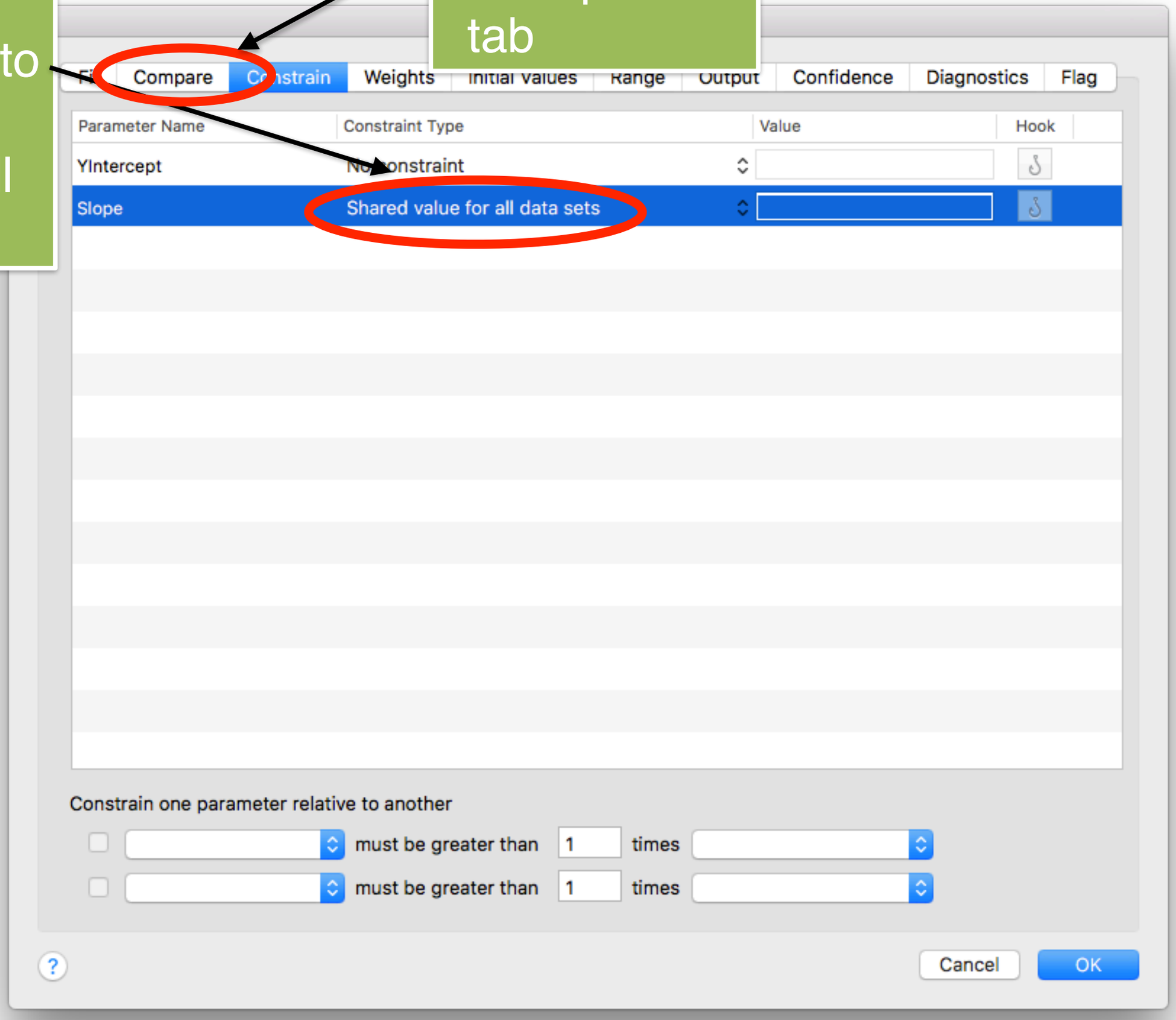
1.Select ‘Straight line’ from Lines

2.Select ‘Constrain’ tab

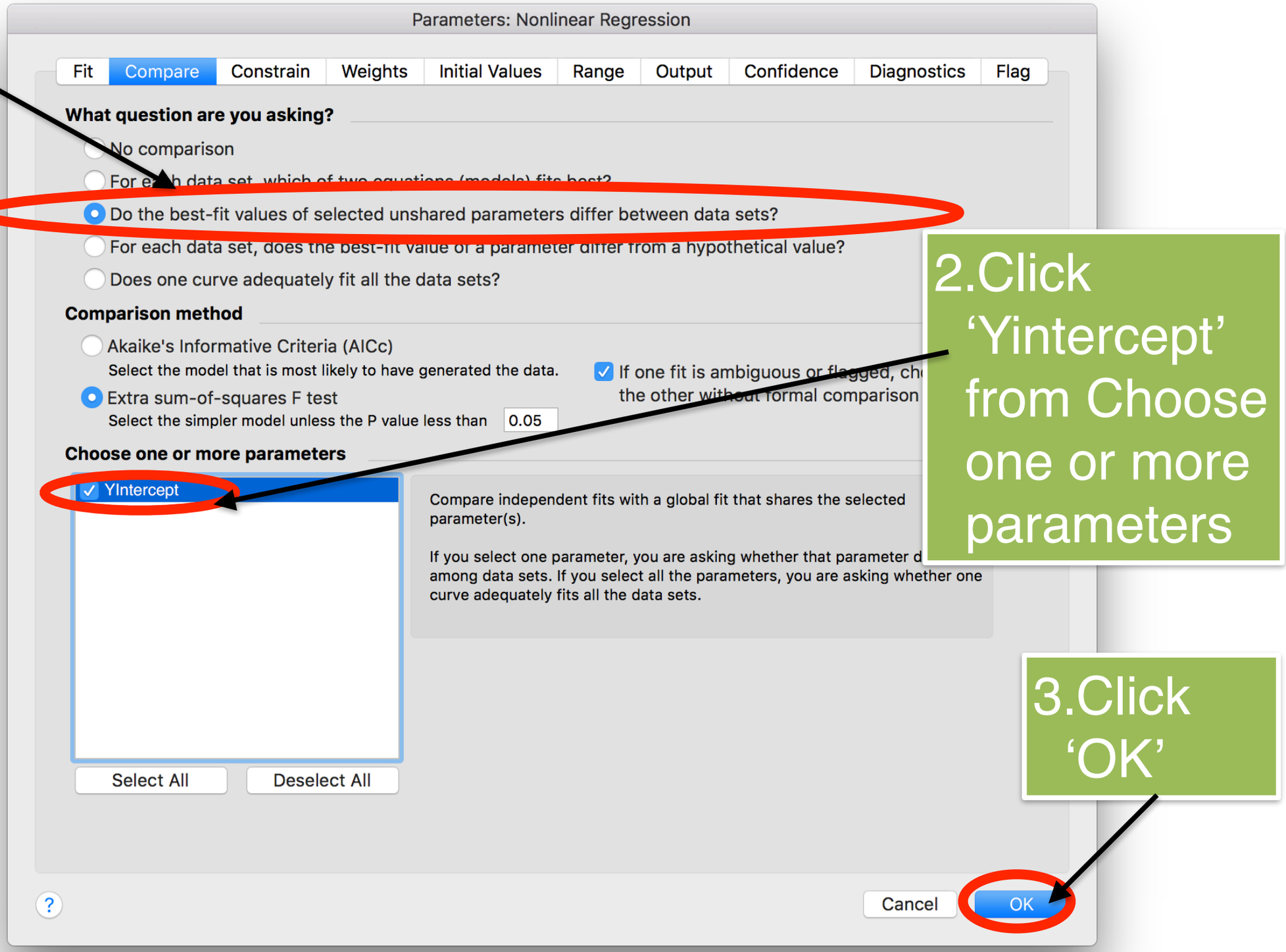


1. Constrain the slopes to ‘Shared value for all data sets’

2. Select ‘Compare’ tab



1.Select ‘Do the best-fit values of selected unshared parameters differ between data sets?’ from What question are you asking?



2.Click ‘Yintercept’ from Choose one or more parameters

3.Click ‘OK’

COPDTrial — Edited

File Sheet Undo Clipboard Analysis Interpret Change Draw Write Text Export

New Analyze

10 Arial

- Family
- Search results
- Data Tables
  - COPDtrial
- Info
  - Project info 1
- Results
  - Nonlin fit of COP
  - Nonlin fit of COP
- Graphs
  - COPDtrial
- Layouts

Nonlin fit		A	B	C	D
		Old Drug	New Drug	Global (shared)	Title
		Y	Y	Y	Y
1	Comparison of Fits				
2	Null hypothesis			YIntercept same for all data sets	
3	Alternative hypothesis			YIntercept different for each data set	
4	P value			<0.0001	
5	Conclusion (alpha = 0.05)			Reject null hypothesis	
6	Preferred model			YIntercept different for each data set	
7	F (DFn, DFd)			103.8 (1, 43)	
8					
9	YIntercept different for each data set				
10	Best-fit values				
11	YIntercept	59.86	72.27		
12	Slope	= -0.5944	= -0.5944		
13	Std. Error				
14	YIntercept	0.8296	0.8631		
15	95% CI (profile likelihood)				
16	YIntercept	58.13 to 61.6	70.49 to 74.05		
17	Goodness of Fit				
18	Degrees of Freedom	19	24		
19	R square	0.7031	0.6707		
20	Absolute Sum of Squares	261.6	447		
21	Sy.x	3.71	4.316		
22	Constraints				
23	Slope	Slope = -0.5944	Slope = -0.5944		
24					
25	YIntercept same for all data sets				
26	Best-fit values				
27	YIntercept	66.75	66.75	66.75	
28	Slope	= -0.5944	= -0.5944		
29	Std. Error				
30	YIntercept	1.105	1.105	1.105	
31	95% CI (profile likelihood)				
32	YIntercept	64.53 to 68.98	64.53 to 68.98	64.53 to 68.98	
33	Goodness of Fit				
34	Degrees of Freedom			44	
35	R square	-0.3752	0.1108	0.4442	
36	Absolute Sum of Squares	1211	1207	2418	
37	Sy.x			7.414	
38	Constraints				
39	YIntercept	YIntercept is shared	YIntercept is shared		
40	Slope	Slope = -0.5944	Slope = -0.5944		
41					
42	Number of points				
43	# of X values	20	45		
44	# Y values analyzed	20	25		
45					
46					
47					

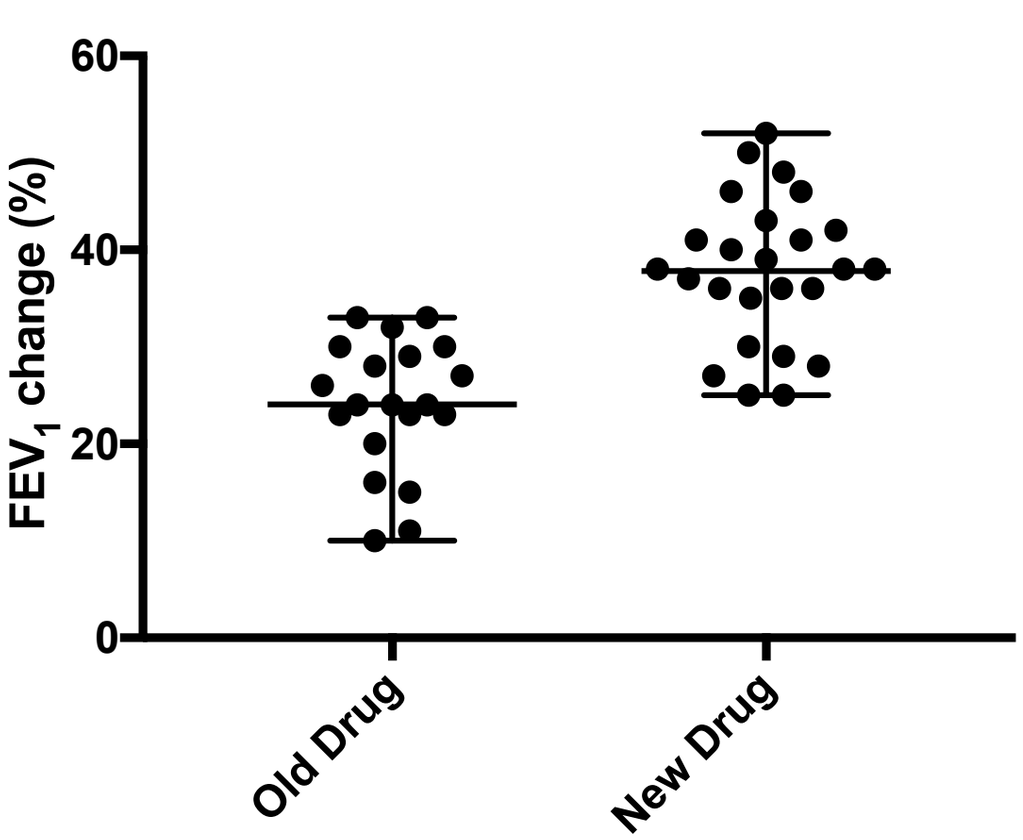
Straight line

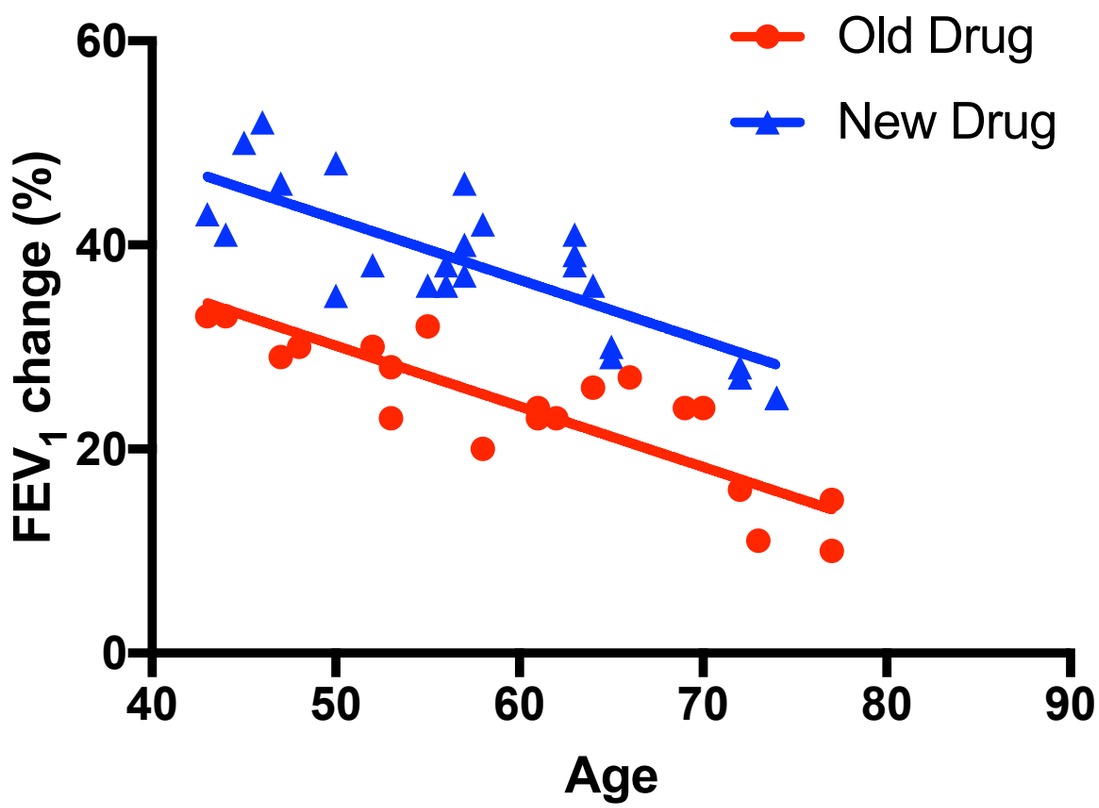
Nonlin fit of COPDtrial | Table of results

ADVANTAGES OF ANCOVA

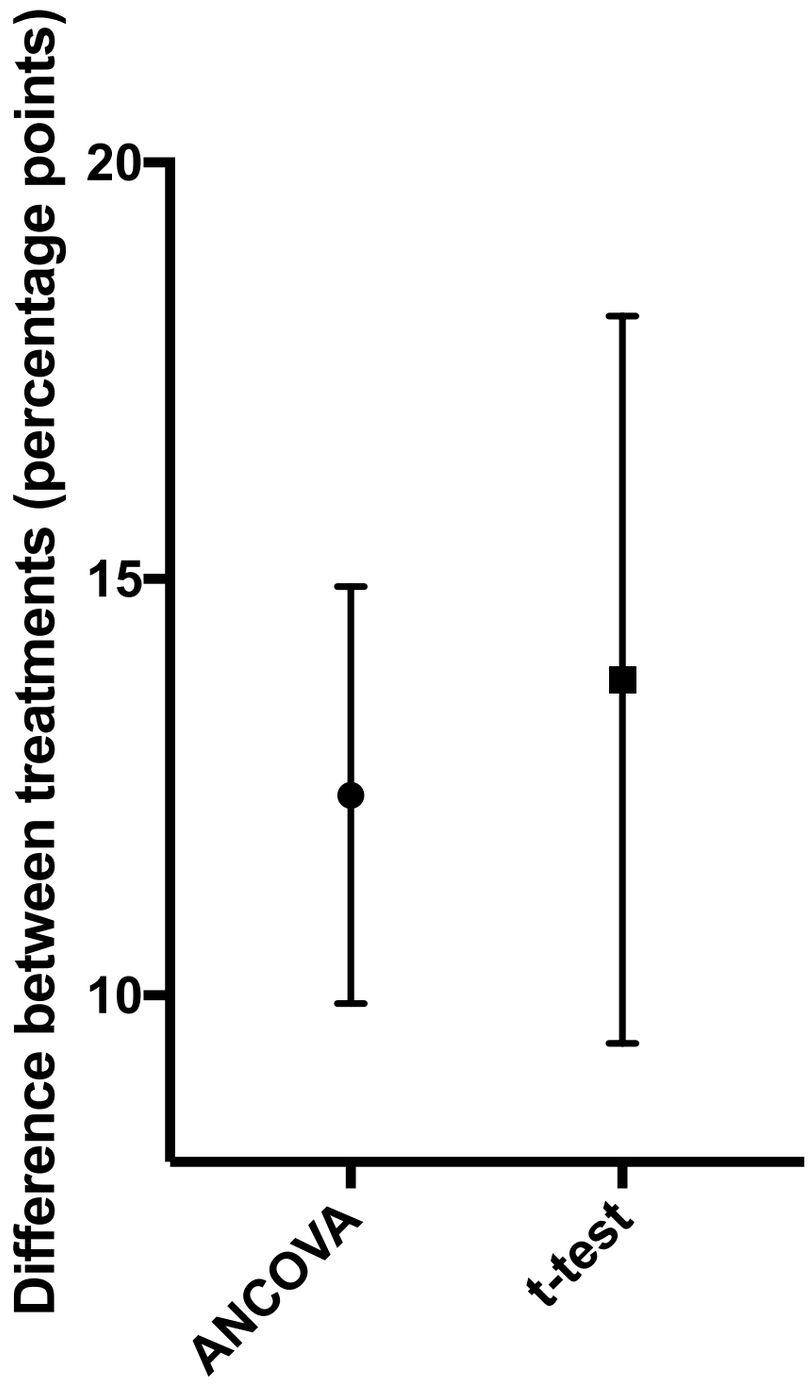
Greater statistical power and narrower confidence intervals

- More unexplained variability there is, the less likely we are to achieve statistical significance





Narrower confidence intervals



Correction for bias due to baseline imbalances

- The patients were randomly allocated to the two treatment groups, so we would expect the two groups to be broadly similar in all regards except drug received.

- Often there are small discrepancies, i.e., imbalances, that may bias the results if the imbalance is in a characteristic that is associated with the outcome.

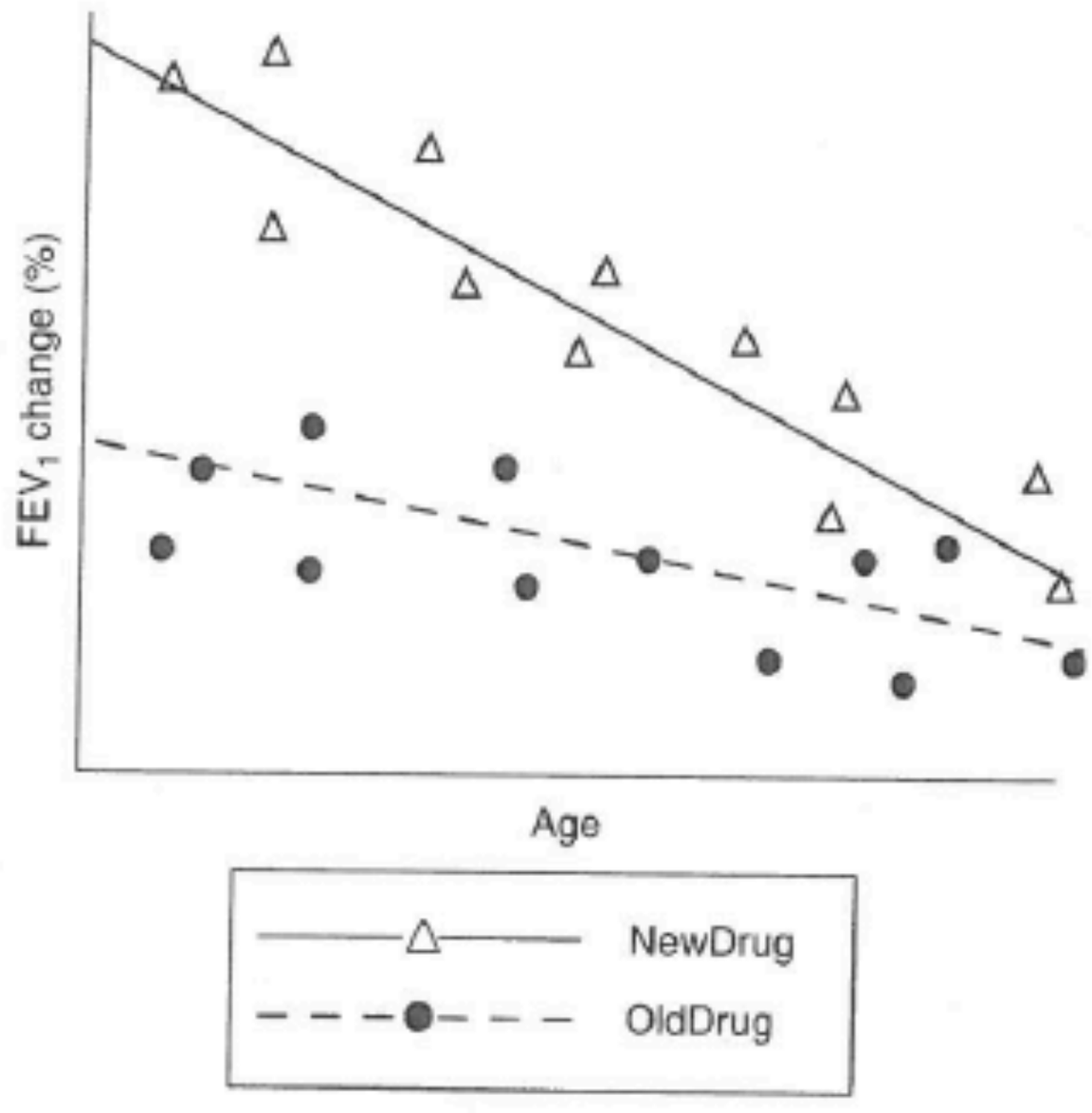
- ANCOVA can correct for this imbalance, but a t-test cannot.

- WARNING - It is unreasonable to attempt to use ANCOVA to correct for baseline imbalance where there is a *gross* difference between groups to be compared.

Identification of significant prognostic factors and possible interactions

- ANCOVA allows for the identification of factors that may influence the effectiveness of a treatment.

- **Prognostic factor** - factor that influences the effectiveness of a drug



What did we learn?

- Analysis of covariance (ANCOVA) is employed when an interval scale endpoint may be influenced by both a nominal and an interval scale factor.

- It is very power to inspect the data graphically first.

- The first stage of the analysis includes an interaction term to test whether there is clear evidence of non-parallelism. If the test for the interaction is non-significant then we proceed to a ‘Common Slopes Model’ where we drop the interaction term and force the two lines to adopt a common slope.

- ANCOVA is has several advantages over a two-sample t-test including the reduction of residual error, the ability to correct for baseline imbalances, and the interpretability in the context of prognostic factors.