Part 3
Lecture 2 Interaction

1

Who we are…

Pascal Tyrrell, PhD *Associate Professor*

Department of Medical Imaging , Faculty of Medicine
Department of Statistical Sciences , Faculty of Arts and Science

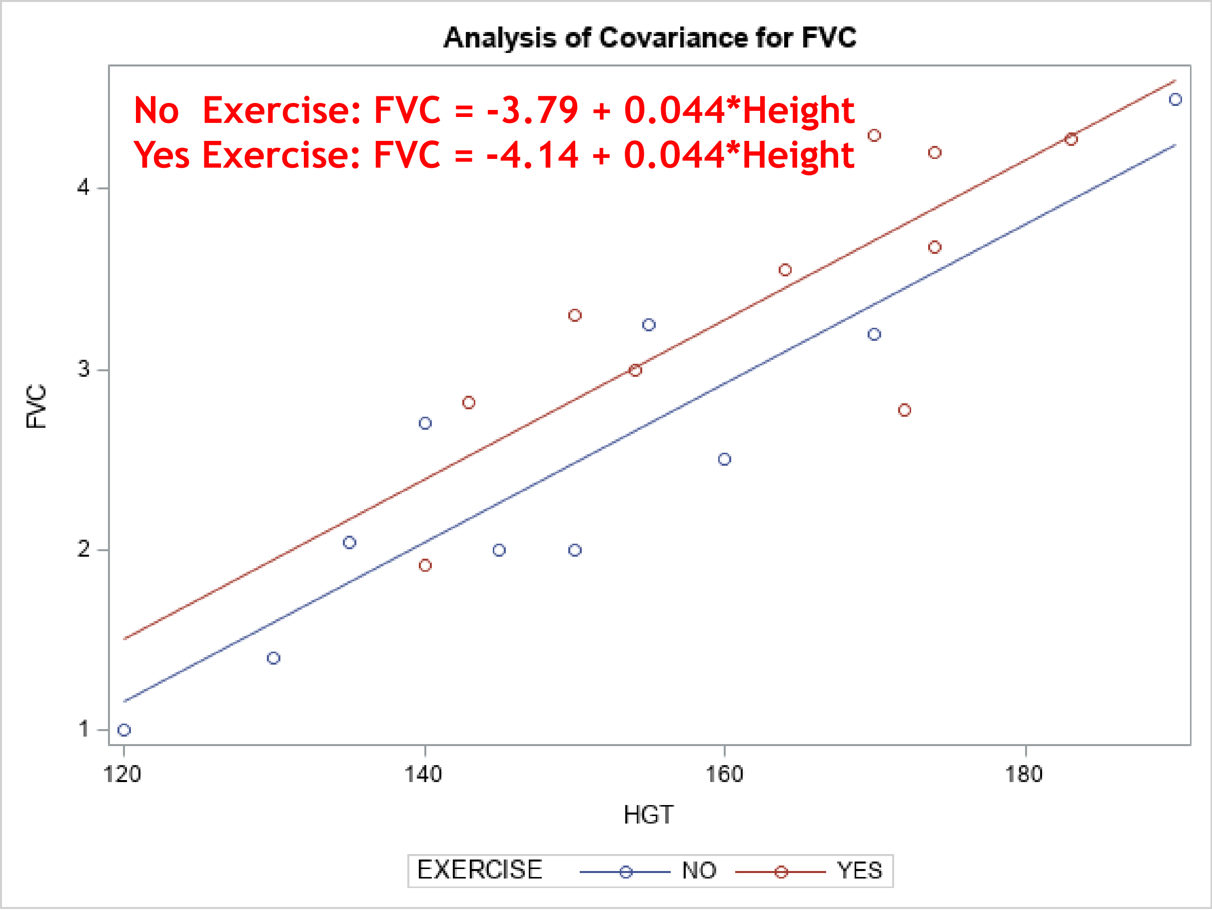
Paul Corey, PhD *Professor Emeritus*

Biostatistics Program, Dalla Lana Faculty of Public Health
Institute of Medical Science, Faculty of Medicine
Department of Statistical Sciences, Faculty of Arts and Science

Last lecture we discovered the sneaky ways of confounding variables…

… But what if we found that the lines in the figure to the right were actually NOT parallel?

This is what we call interation!



Let's recap: COMPARISON OF MEAN FVC

Among twenty persons who recently joined an exercise gymnasium, ten had no experience carrying out any exercises whereas the other ten had some experience doing exercises at home.

The forced vital capacity (FVC) was measured on each of the twenty individuals. The mean FVC was compared between the two groups.

Example 1 COMPARING MEAN FVC IN TWO EXERCISE GROUPS DATASET CONSISTS OF 20 DIFFERENT PATIENTS

```
DATA UNPAIRED_I ; INPUT ID $  EXER  HGT  FVC  @@ ;
EXERCISE = "YES" ;
IF EXER=0 THEN EXERCISE=" NO"; DATALINES ;
 1 0 120 1.00   2 0 130 1.40   3 0 135 2.04
 4 0 145 2.00   5 0 140 2.70   6 0 150 2.00
 7 0 155 3.25   8 0 160 2.50   9 0 170 3.20
10 0 190 4.45
11 1 140 2.12  12 1 150 3.10  13 1 154 3.10
14 1 143 2.22  15 1 164 3.65  16 1 170 4.40
17 1 174 4.01  18 1 172 3.98  19 1 174 4.80
20 1 183 5.28
RUN ;
```

Notice the _I which indicates the data has changed

TWO PROCEDURES FOR COMPARING UNPAIRED MEANS

```
PROC TTEST  DATA = UNPAIRED_I  CL = NONE ;
CLASS EXERCISE ;
VAR FVC  ;
RUN ;

PROC  GLM  DATA = UNPAIRED_I ;
CLASS EXERCISE ;
MODEL FVC  =  EXERCISE  /  SOLUTION SS3 ;
LSMEANS EXERCISE / TDIFF PDIFF STDERR CL ;
RUN ;
```

The TTEST Procedure

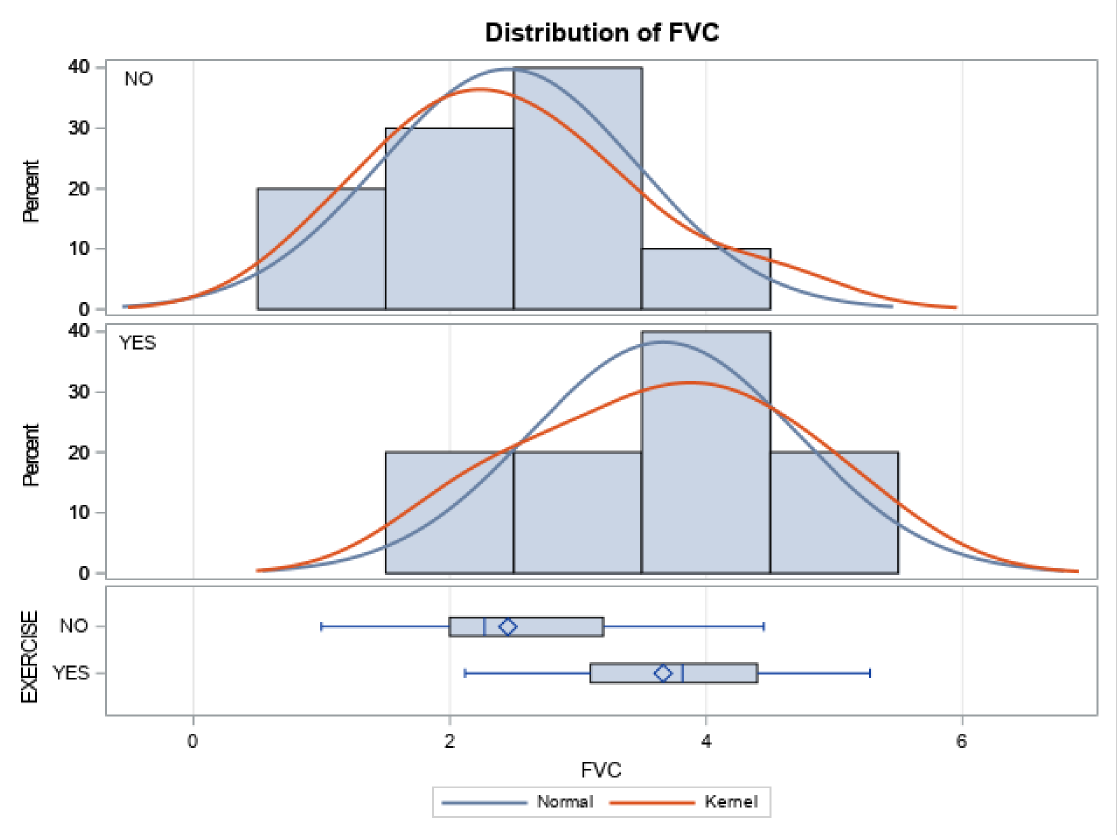
Variable: FVC

EXERCISE	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
NO		10	2.4540	1.0025	0.3170	1.0000	4.4500
YES		10	3.6660	1.0418	0.3295	2.1200	5.2800
Diff (1-2)	Pooled		-1.2120	1.0224	0.4572		
Diff (1-2)	Satterthwaite		-1.2120		0.4572		

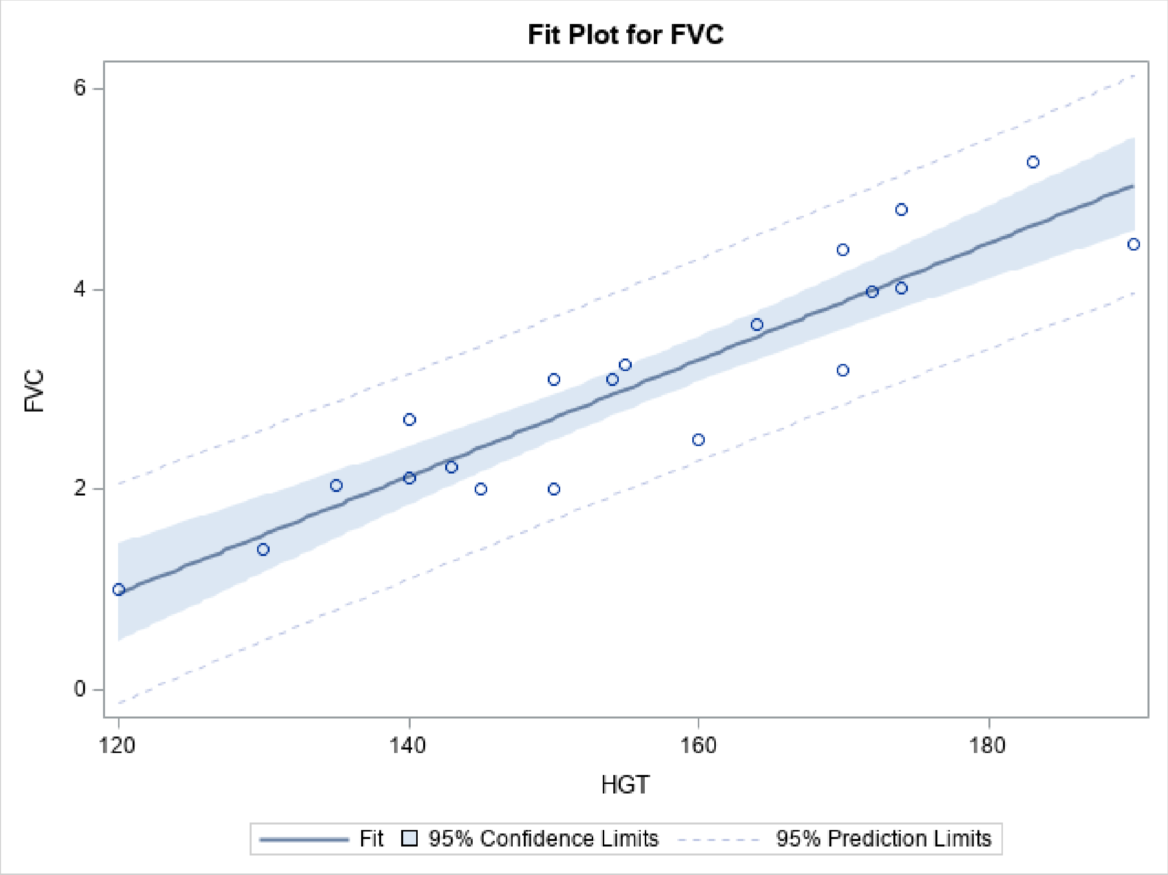
EXERCISE	Method	Mean	95% CL Mean	
NO		2.4540	1.7368	3.1712
YES		3.6660	2.9207	4.4113
Diff (1-2)	Pooled	-1.2120	-2.1726	-0.2514
Diff (1-2)	Satterthwaite	-1.2120	-2.1727	-0.2513

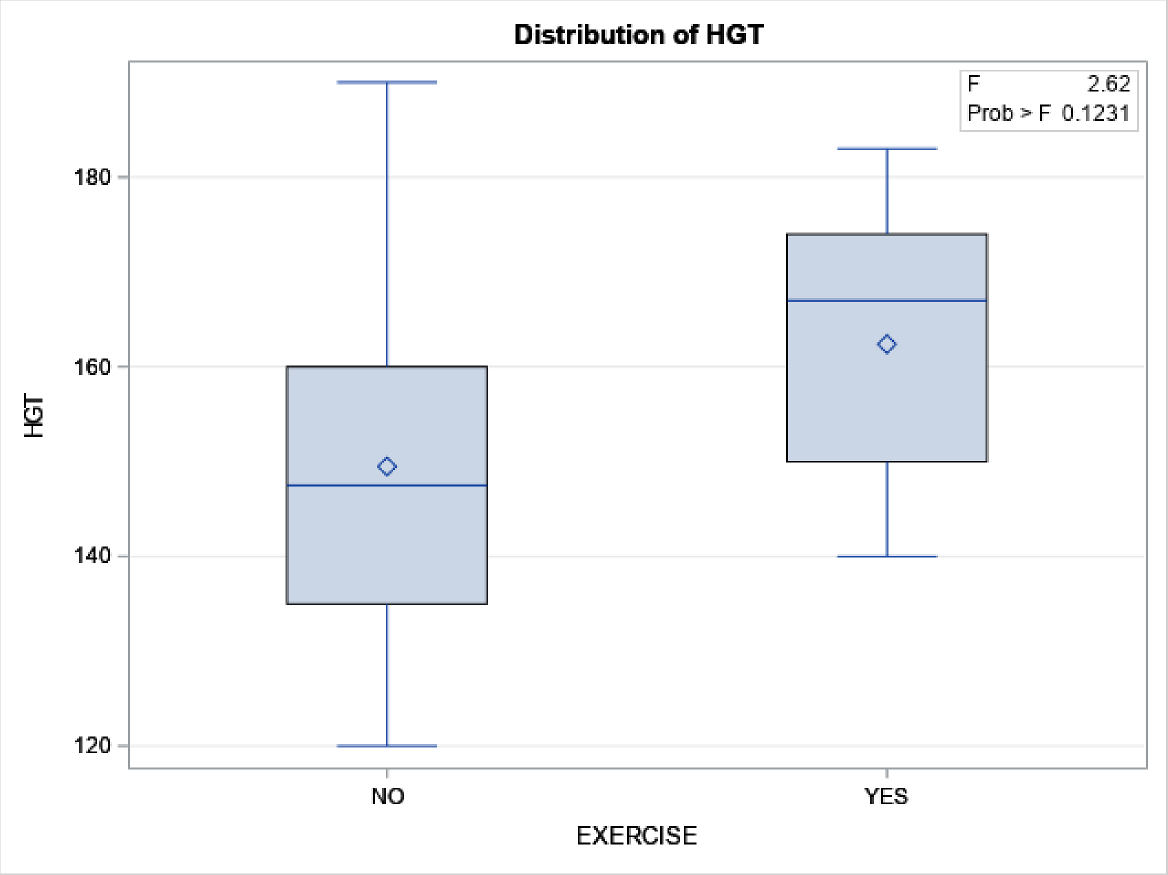
Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	18	-2.65	$\boxed{0.0163}$
Satterthwaite	Unequal	17.973	-2.65	0.0163

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	9	9	1.08	0.9107



But height is still positively correlated with FVC and there still exists a small but not statistically significant difference in height between groups





Interaction term is significant!

The GLM Procedure

Dependent Variable: FVC

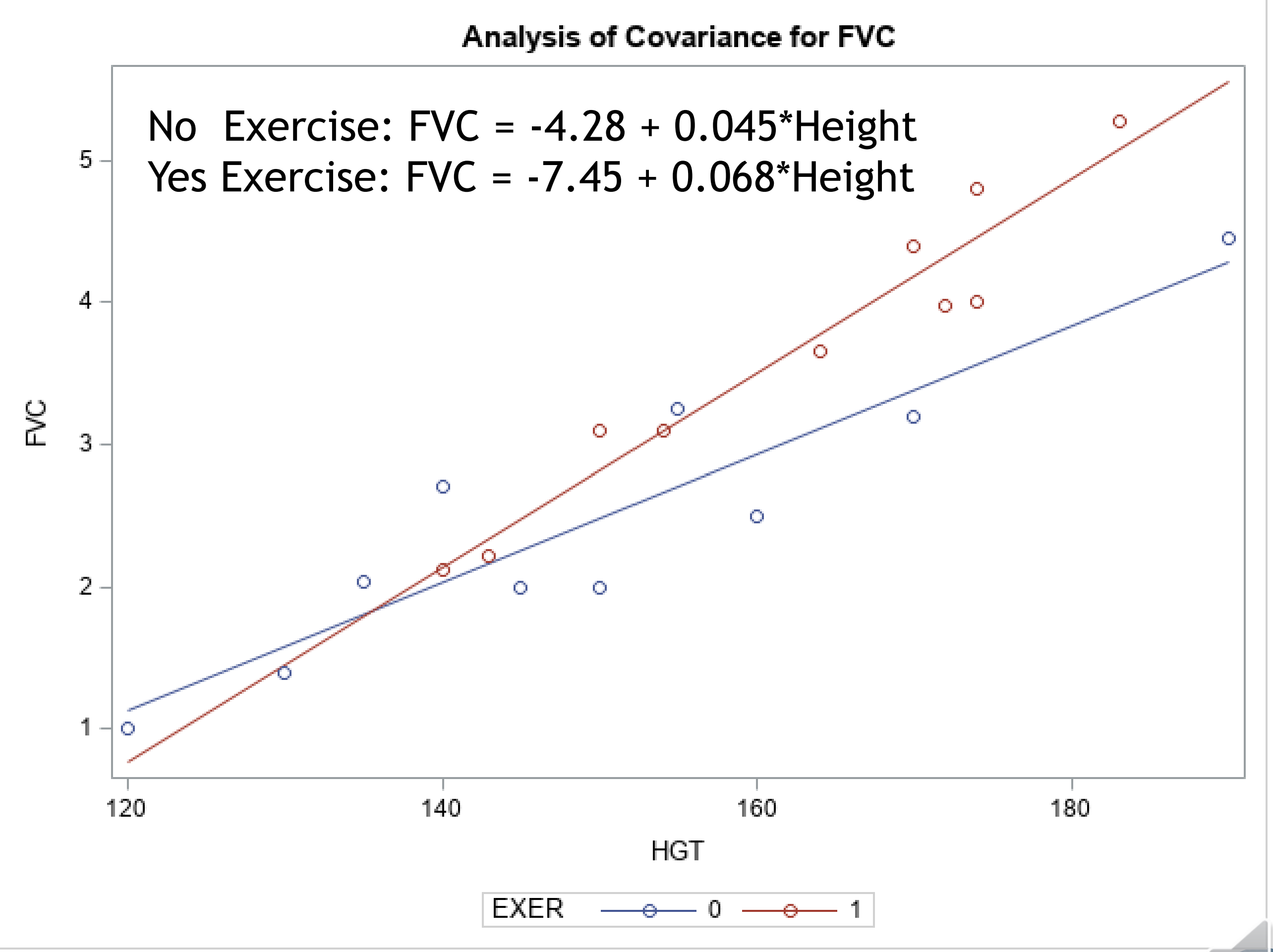
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	24.13382743	8.04460914	63.56	<.0001
Error	16	2.02497257	0.12656079		
Corrected Total	19	26.15880000			

R-Square	Coeff Var	Root MSE	FVC Mean
0.922589	11.62594	0.355754	3.060000

Source	DF	Type III SS	Mean Square	F Value	Pr > F
EXER	1	0.51258325	0.51258325	4.05	$\boxed{0.0613}$
HGT	1	16.55618219	16.55618219	130.82	<.0001
HGT*EXER	1	0.70549665	0.70549665	5.57	$\boxed{0.0312}$

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	-7.454282488	B	1.31369593	-5.67	<.0001
EXER 0	3.174540938	B	1.57742351	2.01	0.0613
EXER 1	0.000000000	B	.	.	.
HGT	0.068474646	B	0.00805954	8.50	<.0001
HGT*EXER 0	$\boxed{-0.023432896}$	B	0.00992494	-2.36	0.0312
HGT*EXER 1	0.000000000	B	.	.	.

```
PROC GLM DATA = UNPAIRED_I ;
CLASS EXER ;
MODEL FVC = EXER HGT EXER*HGT/ SOLUTION SS3 ;
RUN ;
```



So in summary:

- Because the interaction term is significant, we must leave it in the model
- The difference in mean FVC between those who exercise and those who do not remains NOT statistically significant (for this sample)
- The interesting finding is that as height increase its positive association with FVC differs significantly if you exercise or not with a much stronger association with the former (see slopes).

Confounding

- *Most important problem* in observational studies
- Results from the complex inter-relationships between exposure and outcome
- Can lead to overestimate or underestimate of the true association

Confounding

- Occurs when two factors are associated (travel together) and the effect of one is confused with or distorted by the effect of the other

  - e.g. age and many age-related medical conditions
  - e.g. smoking and other adverse lifestyle factors

How do we Select Confounders?

- Selection of potential confounders *must* occur at the design stage
- Based on:
  - Clinical experience
  - Biological plausibility
  - Literature review of previous studies
- Controlling for factors that are *not* confounders may introduce bias (intermediate variable)

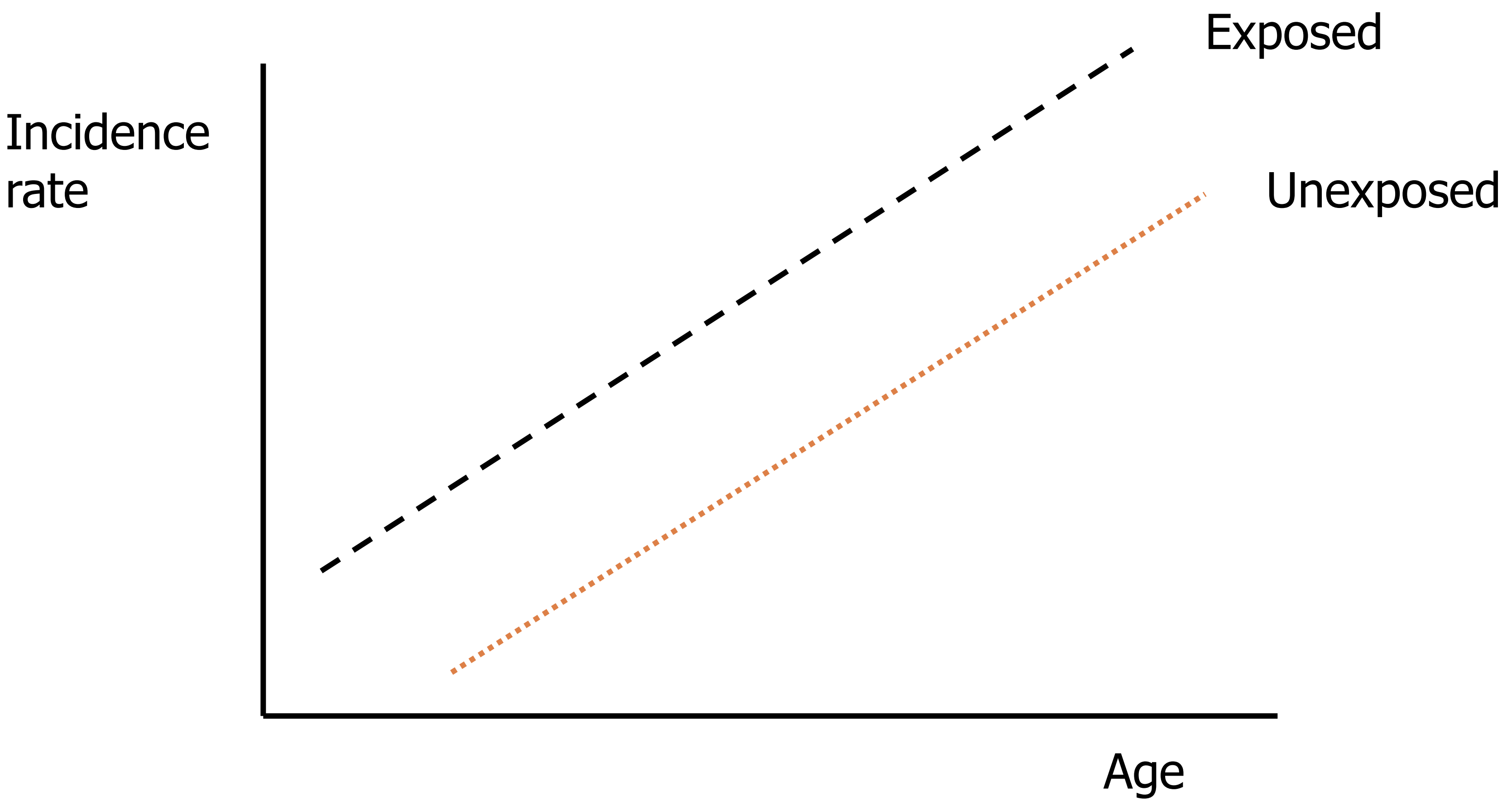
Interaction

- Also called “Effect Modification”
- Two or more risk factors modify the effect of each other on the outcome

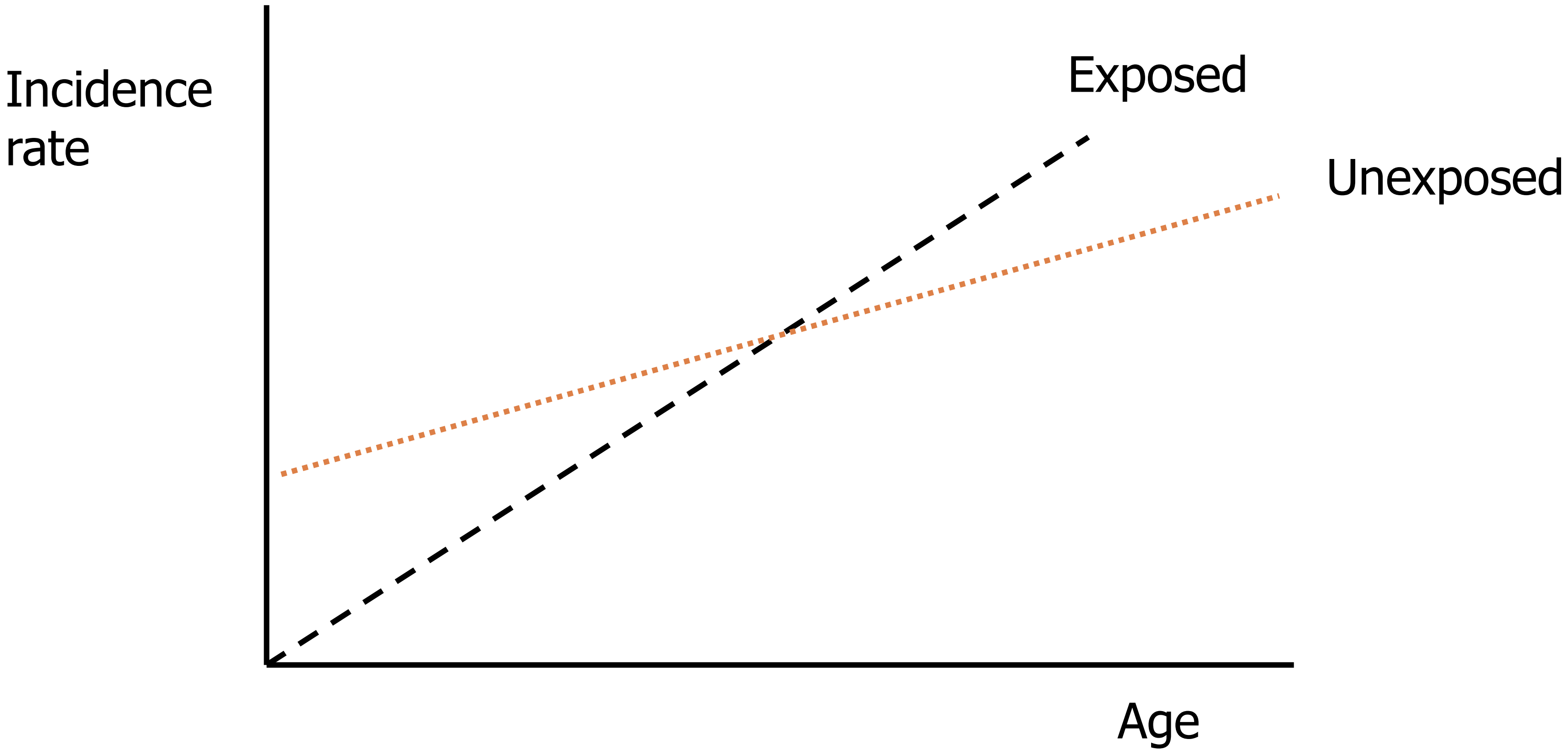
Interaction

- NOT the same as confounding
  - <u>Confounding</u> - when one variable partly or wholly explains the relationship between the exposure and outcome
  - <u>Interaction</u> – the association between the exposure and outcome varies by levels of a third variable

Confounding



Interaction



Checking for Interaction

- Effect modification should be stated *a priori* (as a research hypothesis) and be biologically plausible

- Results should be reported separately for each level of the effect modifier (group variable involved in the interaction)

19

End of Lecture 2

Next up in Part 3 Lecture 3: Recap on confounding

Medical Imaging
UNIVERSITY OF TORONTO

MiDATA