Statistical Interaction / Effect Modification

Dr Celia Gregson

www.theSAMSON.org

Day 2; Tuesday 12th

Learning Objectives

To be able to:

1. Define interaction

2. Understand the difference between confounding and interaction

3. Know how to test for interaction

4. Know how to deal with interaction in your analyses

Definitions

Interaction

- Interaction occurs when the presence of one factor ***modifies*** the effect of another on an outcome
- *i.e.* the effect of the exposure ***differs*** according to which category of the ***other factor*** is being examined

Confounding

- Occurs when an association between an exposure and an outcome is mixed up with the effect of another exposure on the outcome, and the two exposures are related to one another
- For something to be a confounder, it must be associated with the exposure ***and independently*** associated with the outcome

Confounding *vs*. Interaction

Confounding

- Concerned with ***'alternative explanations'*** for an effect of an exposure on outcome
- We aim to remove the influence of a confounder in order to get nearer the '***truth'***
- You ***control*** for confounding factors
- There is no statistical test for confounding

Interaction

- An ***important property*** of the relationship between two factors, and their influence on an outcome
- You do not try to eliminate this effect, instead you want to ***detect and describe interaction*** in the greatest possible ***detail***
- You ***stratify*** by effect modifiers
- There is a statistical test for interaction

University of BRISTOL

SAMSON Sub-Saharan African MuSculOskeletal Network

An example: Confounding
Coffee consumption and Cancer

Exposure → **Outcome**

Coffee consumption → *Cancer*

An example: Confounding
Coffee consumption and Cancer

Exposure → **Outcome**

Coffee consumption → *Cancer*

	Coffee	**No coffee**
Cases (Cancer)	450	300
Controls (No cancer)	200	250

An example: Confounding
Coffee consumption and Cancer

Exposure → **Outcome**

Coffee consumption → *Cancer*

	Coffee	No coffee
Cases (Cancer)	450	300
Controls (No cancer)	200	250

$$= \frac{450/300}{200/250} = \frac{1.5}{0.8}$$

Odds Ratio = 1.9

An example: Confounding
Coffee consumption and Cancer

Exposure

Coffee consumption

Outcome

Cancer

Confounder

Smoking

An example: Confounding
Coffee consumption and Cancer

Exposure

Coffee consumption

Outcome

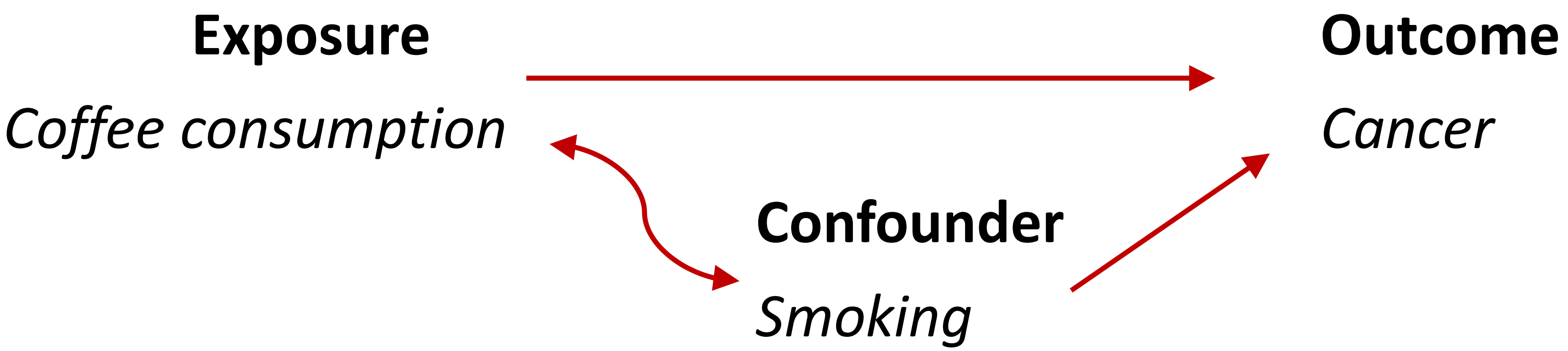
Cancer

Confounder

Smoking

	Smokers		**Non-smokers**	
	Coffee	No coffee	Coffee	No coffee
Cases (Cancer)	400	200	50	100
Controls (No cancer)	100	50	100	200

An example: Confounding
Coffee consumption and Cancer



	Smokers		Non-smokers	
	Coffee	No coffee	Coffee	No coffee
Cases (Cancer)	400	200	50	100
Controls (No cancer)	100	50	100	200

Odds Ratio = 1.0 Odds Ratio = 1.0

An example: Confounding
Coffee consumption and Cancer

Exposure

Coffee consumption

Outcome

Cancer

Confounder

Smoking

	Smokers		**Non-smokers**	
	Coffee	No coffee	Coffee	No coffee
Cases (Cancer)	400	200	50	100
Controls (No cancer)	100	50	100	200

Odds Ratio = 1.0 Odds Ratio = 1.0

Conclusion:

Smoking is totally confounding the association between coffee drinking and cancer

An example: Interaction
Coffee consumption and Cancer

Exposure

Coffee consumption

Outcome

Cancer

Confounder

Smoking

An example: Interaction
Coffee consumption and Cancer

Exposure
Coffee consumption

Outcome
Cancer

Confounder
Smoking

	Smokers		**Non-smokers**	
	Coffee	No coffee	Coffee	No coffee
Cases (Cancer)	400	100	50	200
Controls (No cancer)	100	50	100	200

An example: Interaction
Coffee consumption and Cancer

Exposure

Coffee consumption

Outcome

Cancer

Confounder

Smoking

	Smokers		**Non-smokers**	
	Coffee	No coffee	Coffee	No coffee
Cases (Cancer)	400	100	50	200
Controls (No cancer)	100	50	100	200

Odds Ratio = 2.0

Odds Ratio = 0.5

An example: Interaction
Coffee consumption and Cancer

Exposure

Coffee consumption

Outcome

Cancer

Confounder

Smoking

	Smokers		**Non-smokers**	
	Coffee	No coffee	Coffee	No coffee
Cases (Cancer)	400	100	50	200
Controls (No cancer)	100	50	100	200

Odds Ratio = 2.0

Odds Ratio = 0.5

Conclusion:

Smoking modifies the effect of coffee drinking on cancer

University of BRISTOL

SAMSON Sub-Saharan African MuSculOskeletal Network

An example: Sexual behaviour and risk of HIV infection among men in Kenya

- A study of sexual behaviours and risk of HIV infection
- 400 men with HIV recruited from general medical clinic
- 400 men coming to clinic and testing negative for HIV also recruited
- All were asked about number of sexual partners and condom use
- Study design?

An example: Sexual behaviour and risk of HIV infection among men in Kenya

- A study of sexual behaviours and risk of HIV infection
- 400 men with HIV recruited from general medical clinic
- 400 men coming to clinic and testing negative for HIV also recruited
- All were asked about number of sexual partners and condom use
- Study design?

Case-control study design

An example: Sexual behaviour and risk of HIV infection among men in Kenya

		HIV +	HIV -
Number of sexual partners in past 5 years	≥5	200	100
	<5	200	300

An example: Sexual behaviour and risk of HIV infection among men in Kenya

		HIV +	**HIV -**
Number of sexual partners in past 5 years	≥5	200	100
	<5	200	300

Odds ratio: 3.0

An example: Sexual behaviour and risk of HIV infection among men in Kenya

		HIV +	**HIV -**
Number of sexual partners in past 5 years	≥5	200	100
	<5	200	300

$= \dfrac{200/100}{200/300}$

Odds ratio: 3.0

An example: Sexual behaviour and risk of HIV infection among men in Kenya

		HIV +	HIV -
Number of sexual partners in past 5 years	≥5	200	100
	<5	200	300

$= \frac{200/100}{200/300}$

Odds ratio: 3.0

Men who report *always* using a condom

Men who report *not always* using a condom

An example: Sexual behaviour and risk of HIV infection among men in Kenya

		HIV +	HIV -
Number of sexual partners in past 5 years	≥5	200	100
	<5	200	300

$= \frac{200/100}{200/300}$

Odds ratio: 3.0

Men who report *always* using a condom

		HIV +	HIV -
Number of sexual partners in past 5 years	≥5	50	60
	<5	80	140

Men who report *not always* using a condom

		HIV +	HIV -
Number of sexual partners in past 5 years	≥5	150	40
	<5	120	160

An example: Sexual behaviour and risk of HIV infection among men in Kenya

		HIV +	HIV -
Number of sexual partners in past 5 years	≥5	200	100
	<5	200	300

$= \frac{200/100}{200/300}$

Odds ratio: 3.0

Men who report *always* using a condom

		HIV +	HIV -
Number of sexual partners in past 5 years	≥5	50	60
	<5	80	140

Odds ratio: 1.45

Men who report *not always* using a condom

		HIV +	HIV -
Number of sexual partners in past 5 years	≥5	150	40
	<5	120	160

Odds ratio: 5.0

An example: Sexual behaviour and risk of HIV infection among men in Kenya

		HIV +	HIV -
Number of sexual partners in past 5 years	≥5	200	100
	<5	200	300

$= \frac{200/100}{200/300}$

Odds ratio: 3.0

Men who report *always* using a condom

		HIV +	HIV -
Number of sexual partners in past 5 years	≥5	50	60
	<5	80	140

Odds ratio: 1.45

Men who report *not always* using a condom

		HIV +	HIV -
Number of sexual partners in past 5 years	≥5	150	40
	<5	120	160

Odds ratio: 5.0

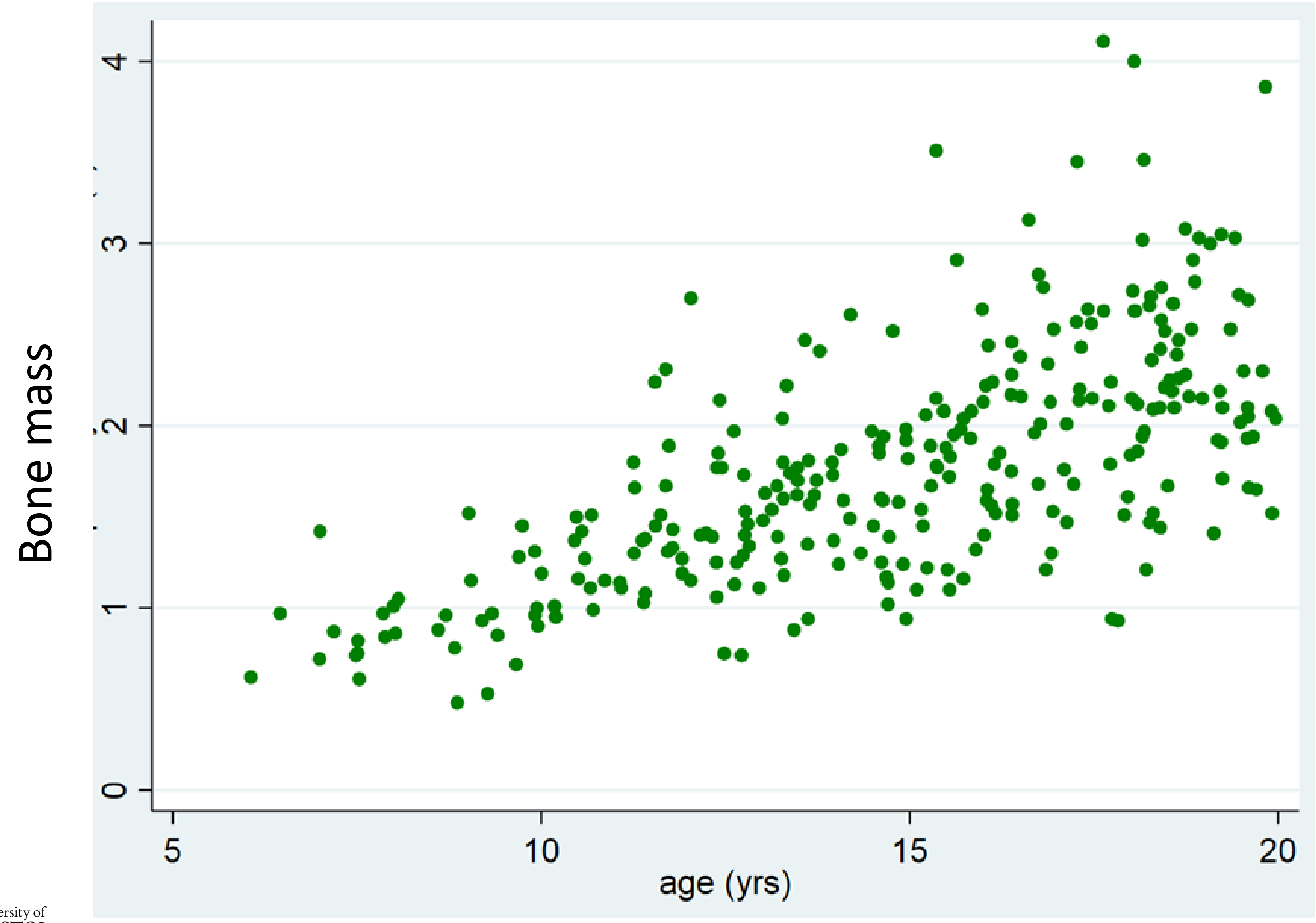
Conclusion: there is an interaction between the number of sexual partners and condom use on the odds of HIV infection

i.e. reporting using a condom ***modifies the effect*** of the number of sexual partners on the odds of HIV infection, so that reporting condom use lowers the effect of higher sexual partner number on the odds of HIV infection

Summary

		Crude odds/ rate/ risk ratio	Odds/ rate/ risk ratio in Stratum 1	Odds/ rate/ risk ratio in Stratum 2	Adjusted odds/ rate/ risk ratio
Example 1	No confounding No interaction	3.0	3.0	3.0	3.0
Example 2	Confounding No interaction	3.0	2.0	2.0	2.0
Example 3	Interaction	3.0	0.8	5.5	Should not be calculated

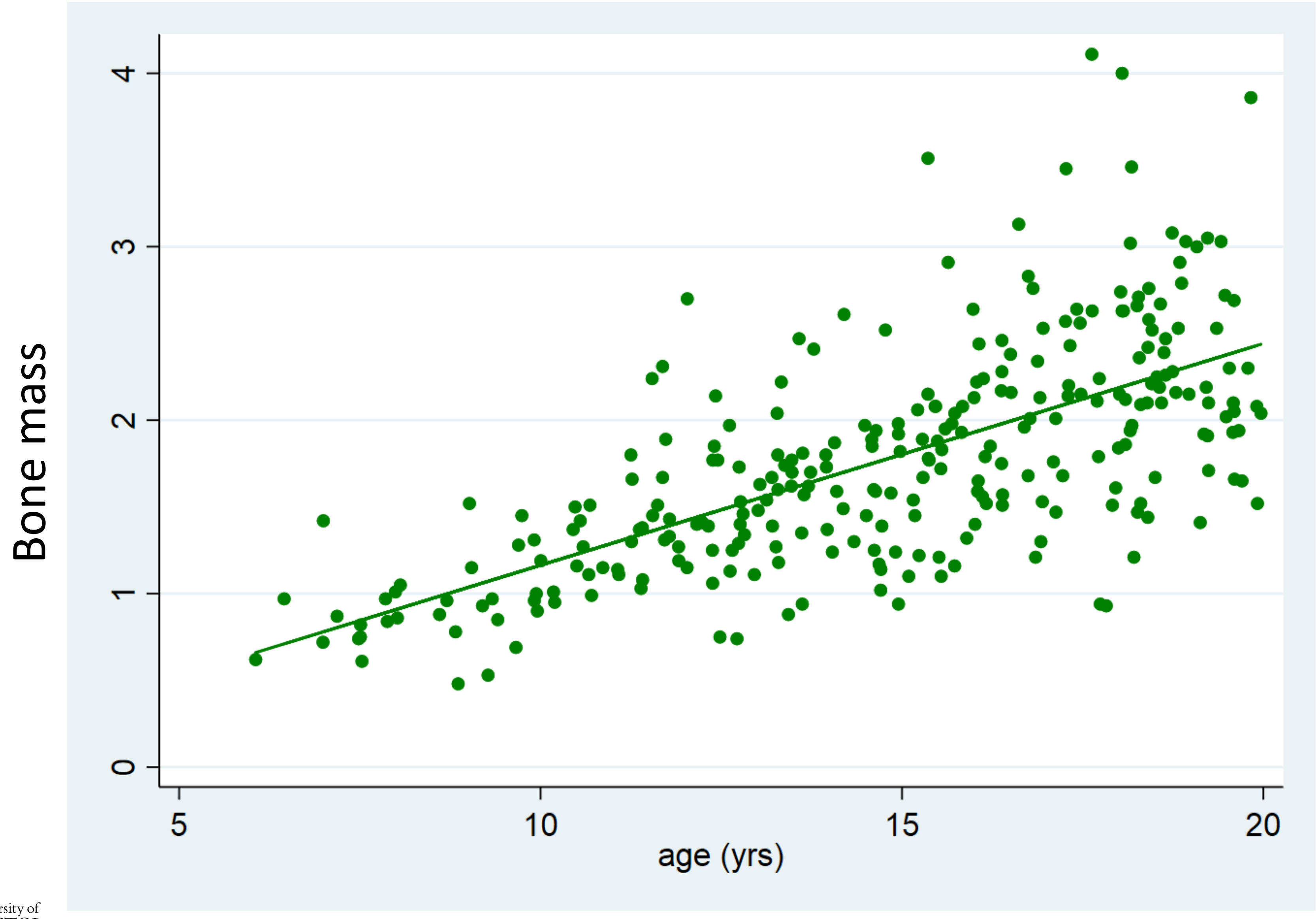
Now thinking about continuous data



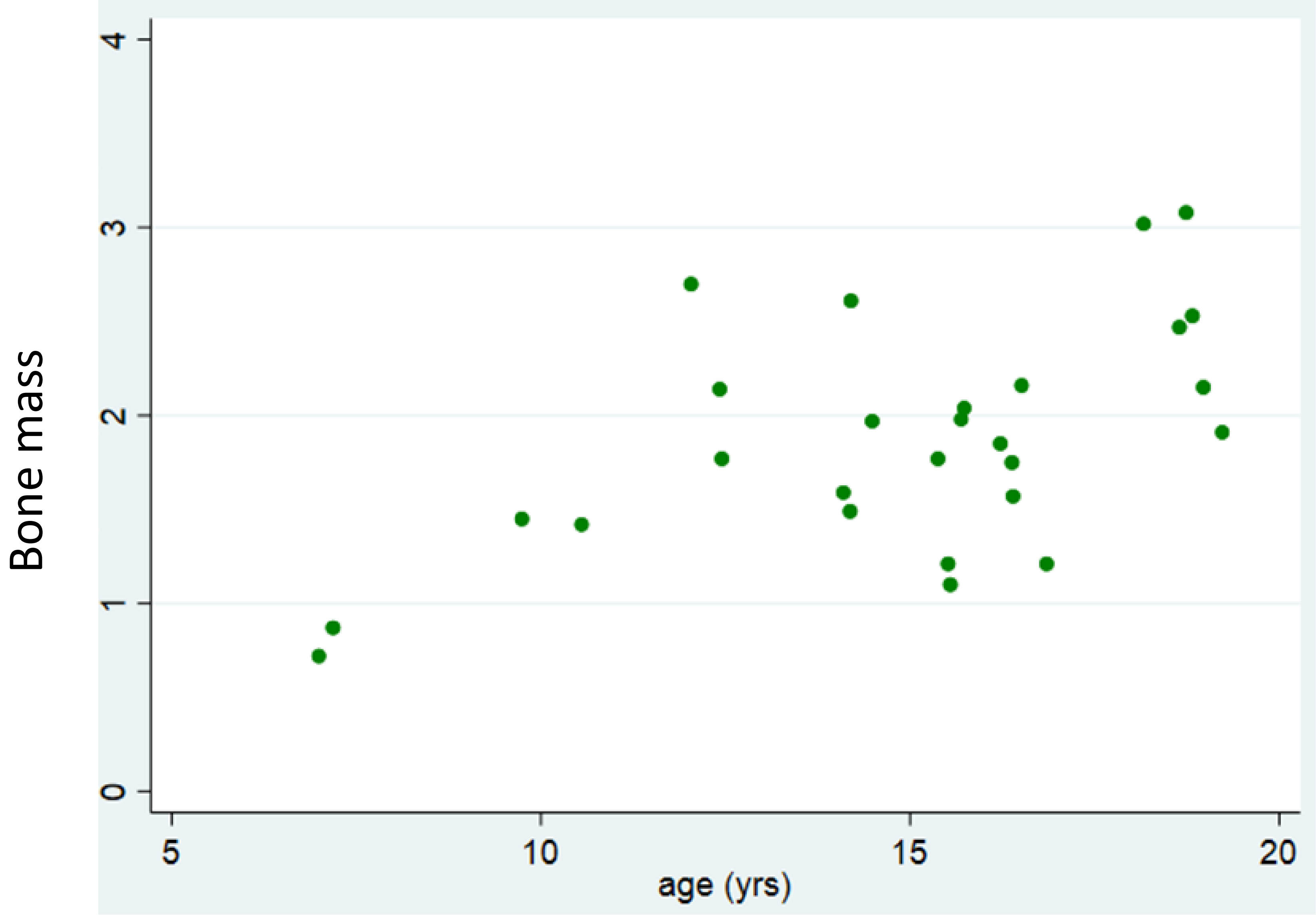
University of BRISTOL

SAMSON Sub-Saharan African MuSculOskeletal Network

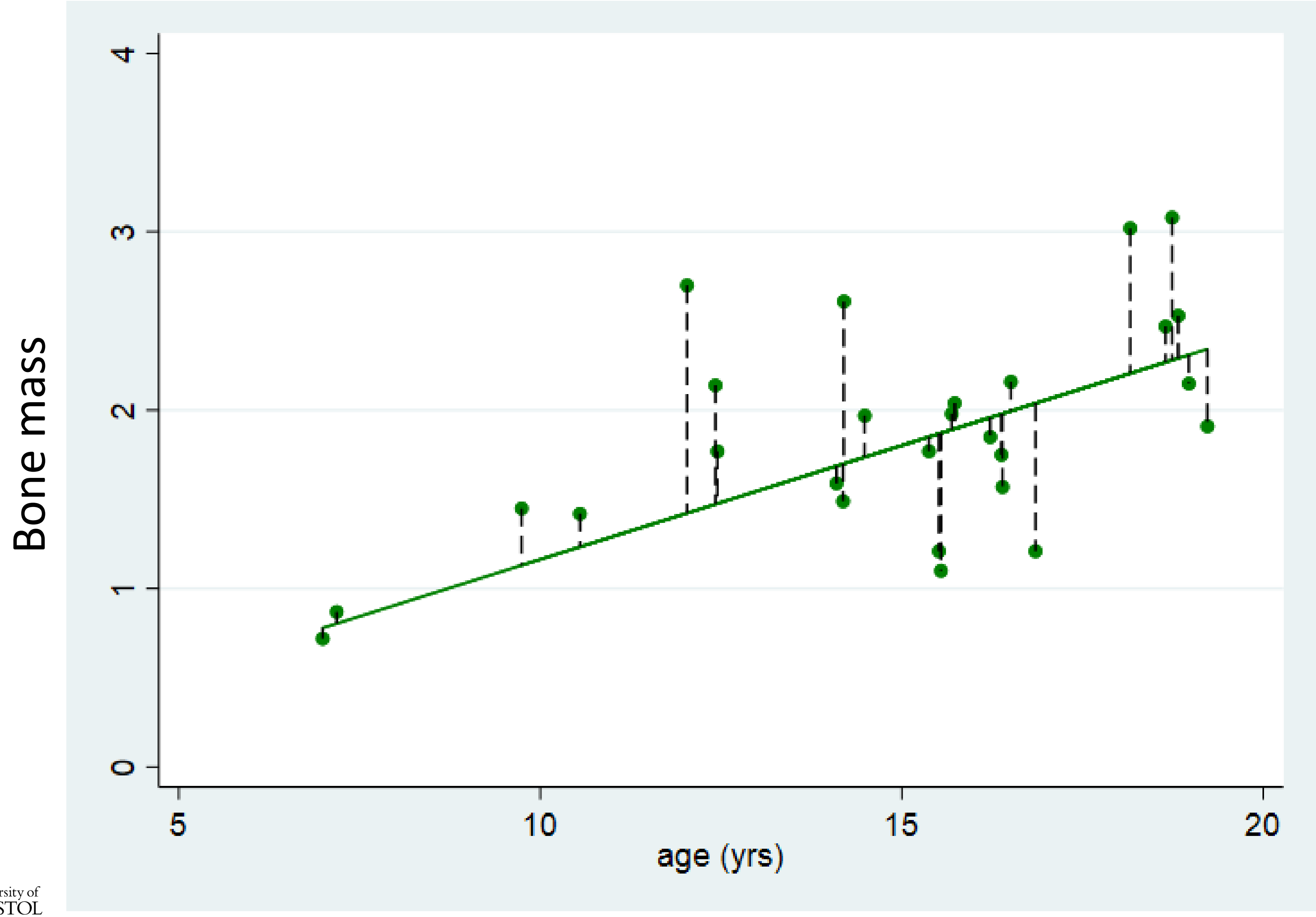
Adding a line of best fit



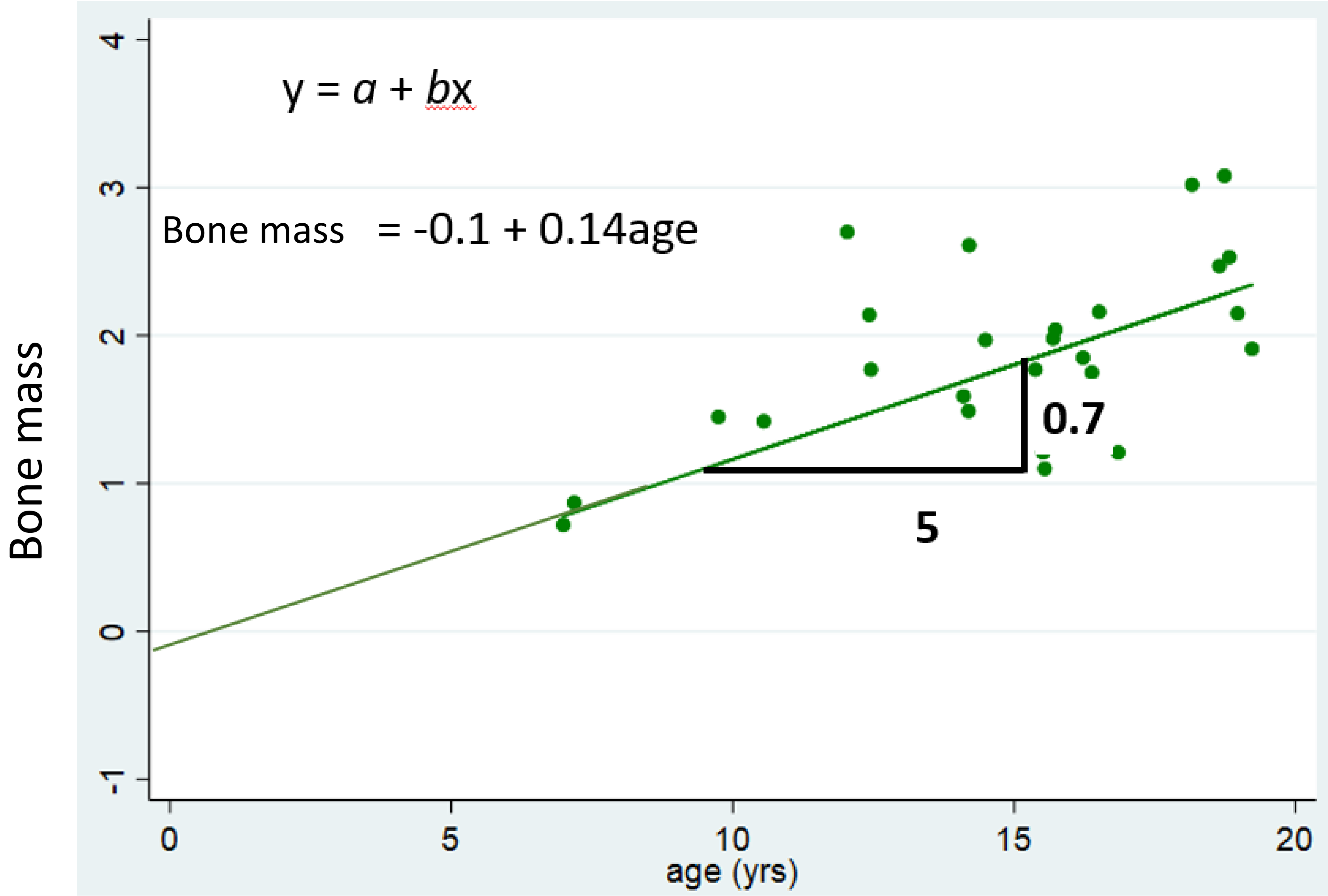
Adding a line of best fit



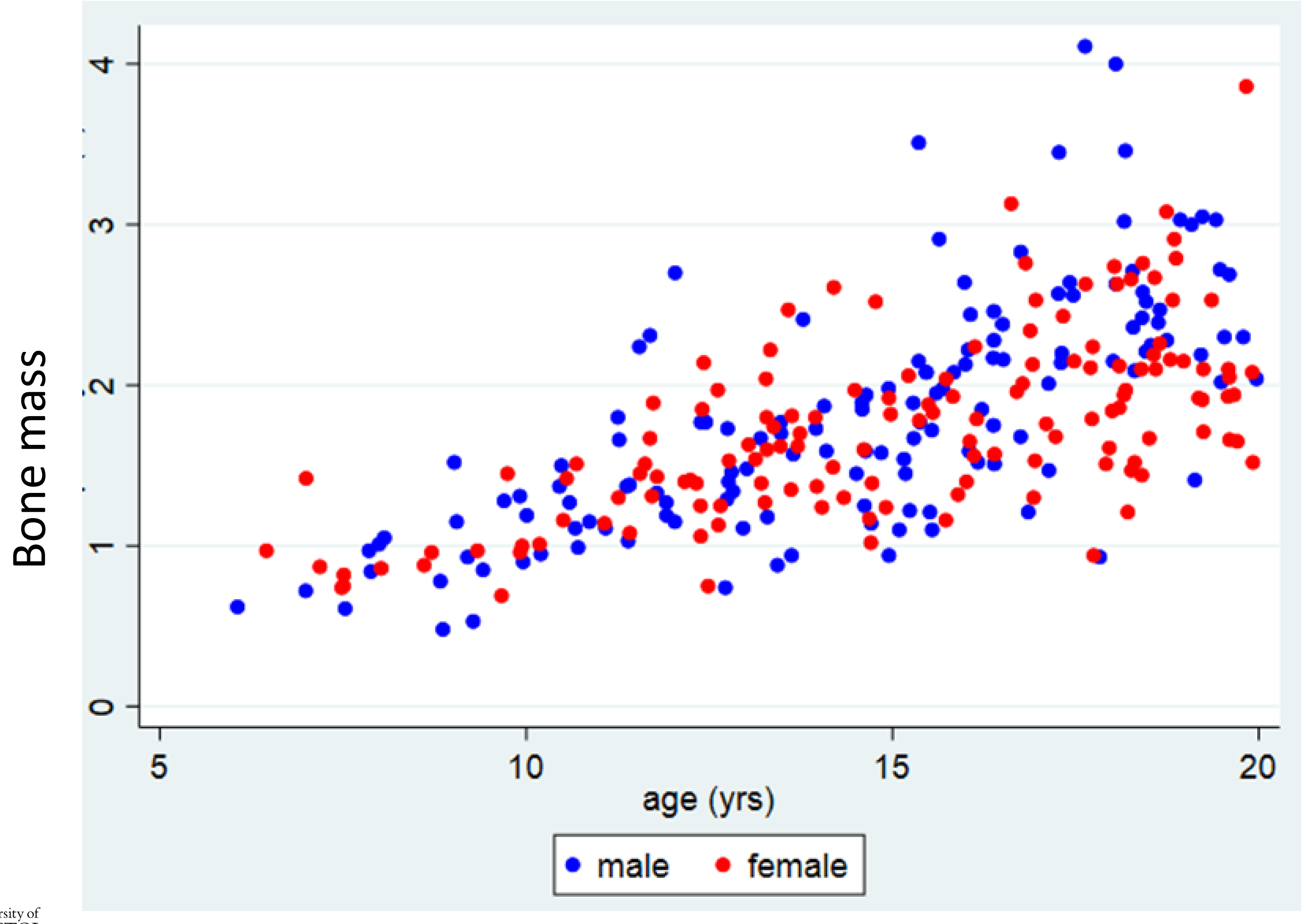
Adding a line of best fit



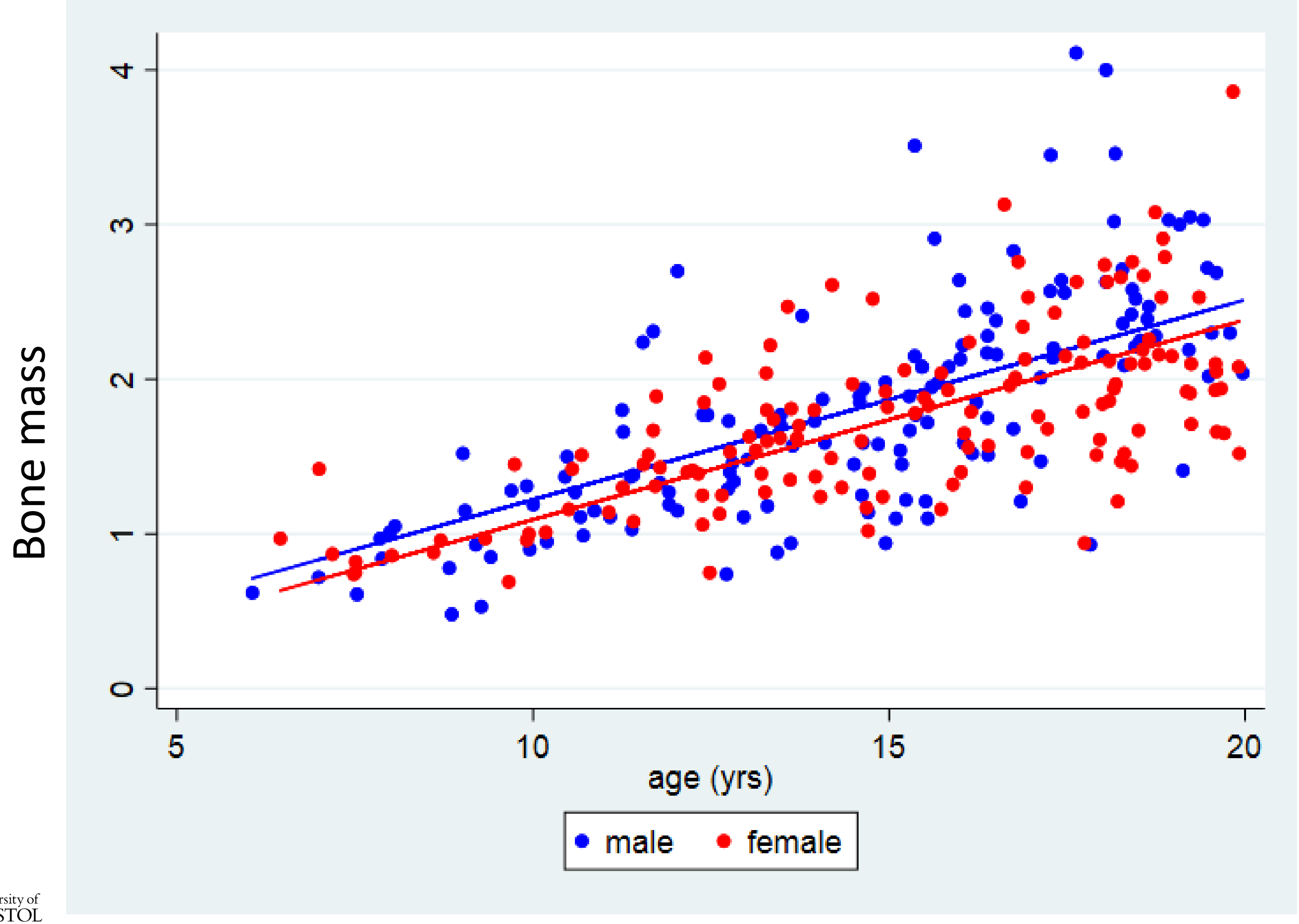
The intercept and the slope



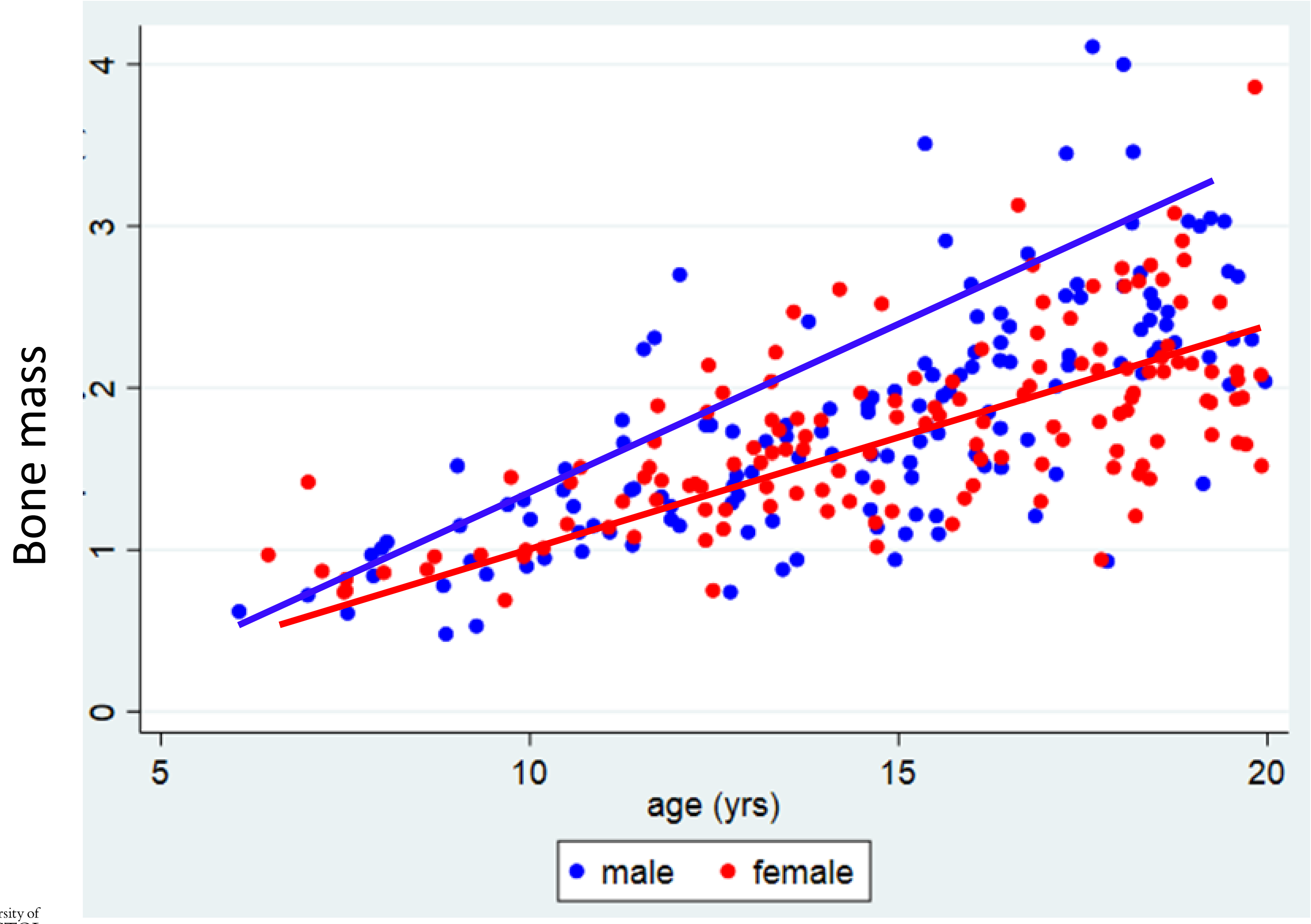
Two populations within this distribution



Each with lines of best fit with the same slopes



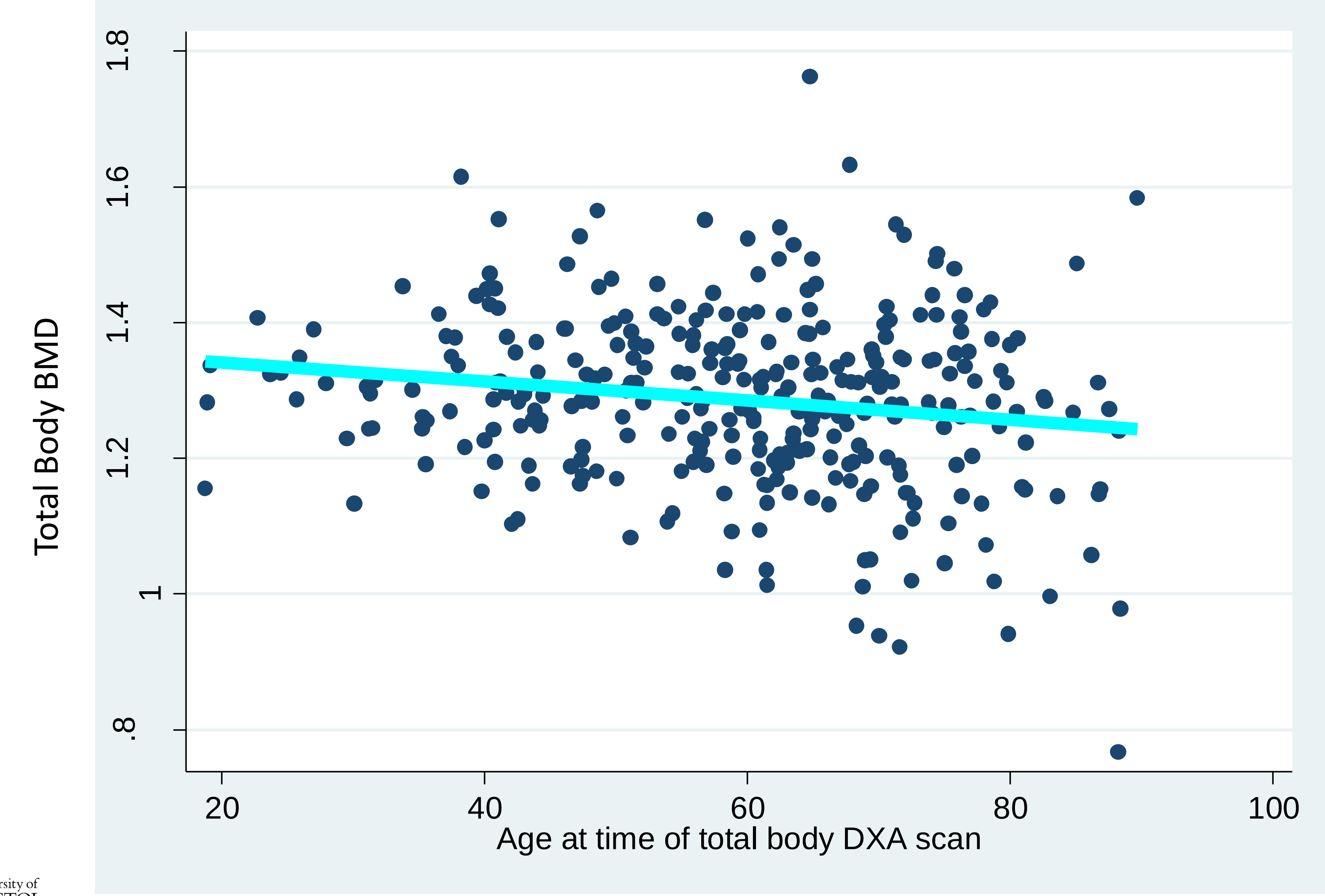
…or different slopes… This is interaction



University of BRISTOL
SAMSON Sub-Saharan African MuSculOskeletal Network

An example: Total Body BMD by Age

Adding in a regression line:



regress outcome exposure

```
reg tb_bmd age
```

Source	SS	df	MS
Model	.145290357	1	.145290357
Residual	5.24668332	326	.016094121
Total	5.39197368	327	.016489216

Number of obs	=	328
F(1, 326)	=	9.03
Prob > F	=	0.0029
R-squared	=	0.0269
Adj R-squared	=	0.0240
Root MSE	=	.12686

tb_bmd	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	-.0014089	.0004689	-3.00	0.003	-.0023313	-.0004864
_cons	1.369698	.0286457	47.82	0.000	1.313344	1.426051

```
reg tb_bmd age

      Source |       SS           df       MS      Number of obs   =       328
-------------+----------------------------------   F(1, 326)       =      9.03
       Model |  .145290357         1  .145290357   Prob > F        =    0.0029
    Residual |  5.24668332       326  .016094121   R-squared       =    0.0269
-------------+----------------------------------   Adj R-squared   =    0.0240
       Total |  5.39197368       327  .016489216   Root MSE        =    .12686

------------------------------------------------------------------------------
      tb_bmd |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.0014089   .0004689    -3.00   0.003    -.0023313   -.0004864
       _cons |   1.369698   .0286457    47.82   0.000     1.313344    1.426051
------------------------------------------------------------------------------
```

Amount of variance explained by model

Variables in the model

Proportion of variance explained by model

```
reg tb_bmd age

      Source |       SS           df       MS      Number of obs   =       328
-------------+----------------------------------   F(1, 326)       =      9.03
       Model |  .145290357         1  .145290357   Prob > F        =    0.0029
    Residual |  5.24668332       326  .016094121   R-squared       =    0.0269
-------------+----------------------------------   Adj R-squared   =    0.0240
       Total |  5.39197368       327  .016489216   Root MSE        =    .12686

------------------------------------------------------------------------------
      tb_bmd |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.0014089   .0004689    -3.00   0.003    -.0023313   -.0004864
       _cons |   1.369698   .0286457    47.82   0.000     1.313344    1.426051
------------------------------------------------------------------------------
```

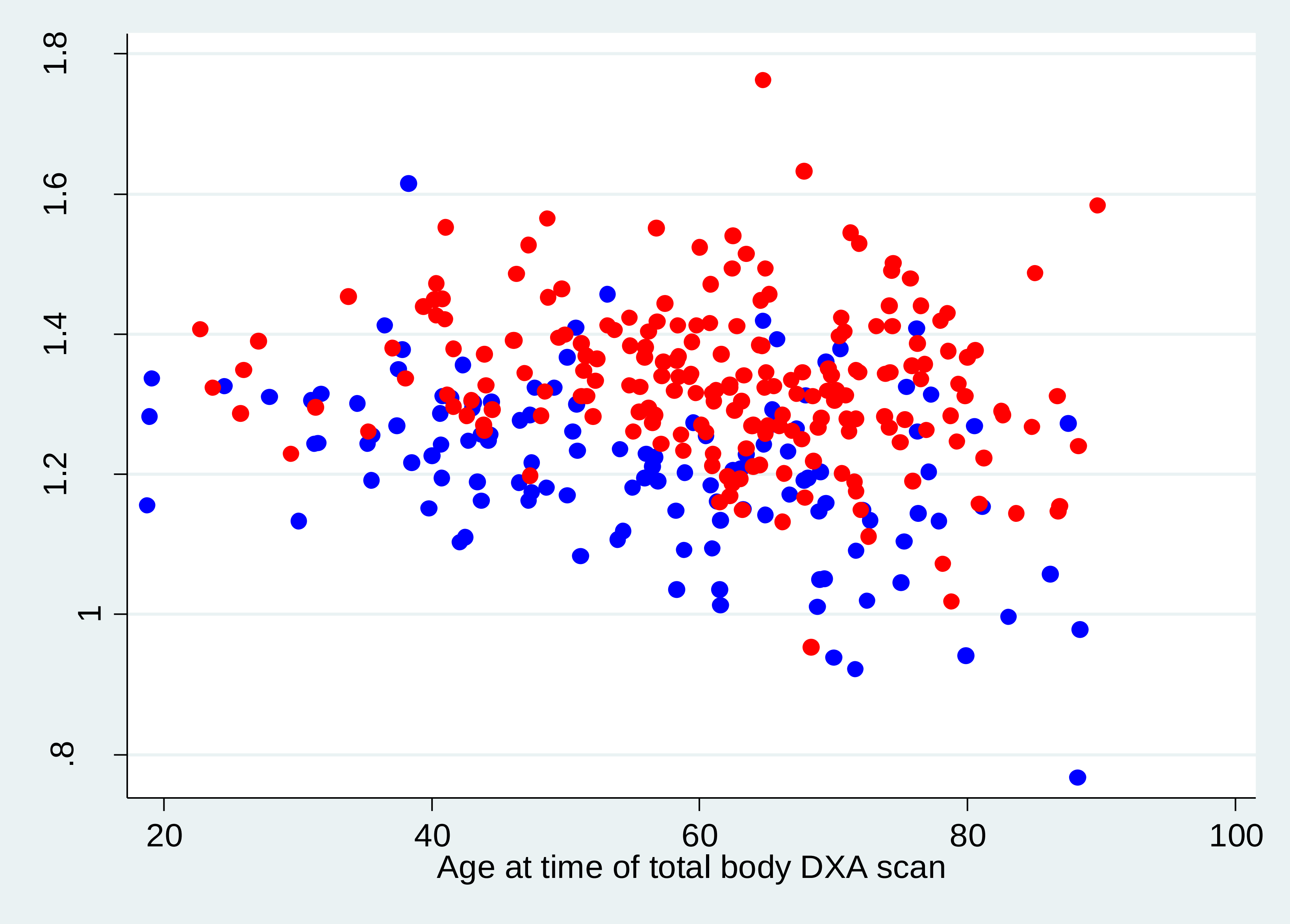
Slope of the line

Used to calculate 95% CI

Probability a result like this would have come about if in reality there is no association with outcome

95% confident the slope is within this range

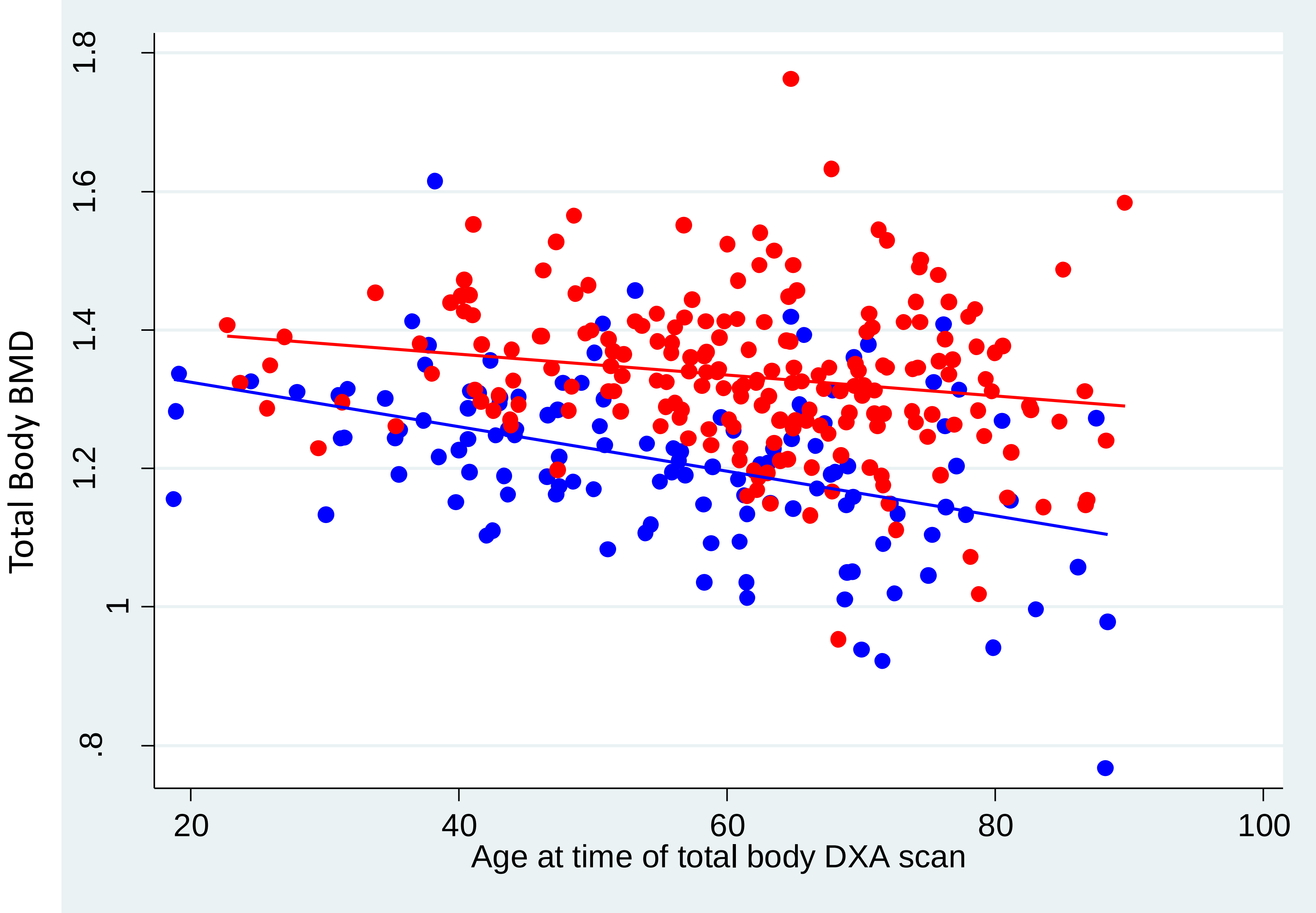
But there are 2 populations – with and without high bone mass



HBM cases in red

Normal individuals in blue

Testing for interaction:
Is the slope of the regression lines different?



HBM cases in red

Normal individuals in blue

HBM 1

Non-HBM 0

Test for interaction: reg tb_bmd i.hbm##c.age

```
. reg tb_bmd i.hbm##c.age

      Source |       SS           df       MS      Number of obs   =       328
-------------+----------------------------------   F(3, 324)       =     44.37
       Model |  1.57011575         3  .523371916   Prob > F        =    0.0000
    Residual |  3.82185793       324  .011795858   R-squared       =    0.2912
-------------+----------------------------------   Adj R-squared   =    0.2846
       Total |  5.39197368       327  .016489216   Root MSE        =    .10861

------------------------------------------------------------------------------
      tb_bmd |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       1.hbm |  -.0368218   .0495197    -0.74   0.458    -.1342425    .0605989
         age |  -.0015079   .0005569    -2.71   0.007    -.0026035   -.0004123
             |
   hbm#c.age |
          1  |  -.0017025   .0008221    -2.07   0.039    -.0033198   -.0000853
             |
       _cons |   1.425571   .0351261    40.58   0.000     1.356467    1.494675
------------------------------------------------------------------------------
```

Test for interaction: reg tb_bmd i.hbm##c.age

```
. reg tb_bmd i.hbm##c.age

      Source |       SS           df       MS      Number of obs   =       328
-------------+----------------------------------   F(3, 324)       =     44.37
       Model |  1.57011575         3  .523371916   Prob > F        =    0.0000
    Residual |  3.82185793       324  .011795858   R-squared       =    0.2912
-------------+----------------------------------   Adj R-squared   =    0.2846
       Total |  5.39197368       327  .016489216   Root MSE        =    .10861

------------------------------------------------------------------------------
      tb_bmd |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       1.hbm |  -.0368218   .0495197    -0.74   0.458    -.1342425    .0605989
         age |  -.0015079   .0005569    -2.71   0.007    -.0026035   -.0004123
             |
   hbm#c.age |
           1 |  -.0017025   .0008221    -2.07   0.039    -.0033198   -.0000853
             |
       _cons |   1.425571   .0351261    40.58   0.000     1.356467    1.494675
------------------------------------------------------------------------------
```

Slope of line for baseline group (non-HBM)

Y axis intercept for baseline group (non-HBM)

Test for interaction: reg tb_bmd i.hbm##c.age

```
. reg tb_bmd i.hbm##c.age

      Source |       SS           df       MS      Number of obs   =       328
-------------+----------------------------------   F(3, 324)       =     44.37
       Model |  1.57011575         3  .523371916   Prob > F        =    0.0000
    Residual |  3.82185793       324  .011795858   R-squared       =    0.2912
-------------+----------------------------------   Adj R-squared   =    0.2846
       Total |  5.39197368       327  .016489216   Root MSE        =    .10861

------------------------------------------------------------------------------
      tb_bmd |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       1.hbm |  -.0368218   .0495197    -0.74   0.458    -.1342425    .0605989
         age |  -.0015079   .0005569    -2.71   0.007    -.0026035   -.0004123
             |
 hbm#c.age |
           1 |  -.0017025   .0008221    -2.07   0.039    -.0033198   -.0000853
             |
       _cons |   1.425571   .0351261    40.58   0.000     1.356467    1.494675
------------------------------------------------------------------------------
```

Slope of line for baseline group (non-HBM)

Y axis intercept for baseline group (non-HBM)

age + hbm#c.age = Slope of line for HBM group

Test for interaction: reg tb_bmd i.hbm##c.age

```
. reg tb_bmd i.hbm##c.age
```

Source	SS	df	MS
Model	1.57011575	3	.523371916
Residual	3.82185793	324	.011795858
Total	5.39197368	327	.016489216

Number of obs	=	328
F(3, 324)	=	44.37
Prob > F	=	0.0000
R-squared	=	0.2912
Adj R-squared	=	0.2846
Root MSE	=	.10861

tb_bmd	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
1.hbm	-.0368218	.0495197	-0.74	0.458	-.1342425	.0605989
age	-.0015079	.0005569	-2.71	0.007	-.0026035	-.0004123
hbm#c.age						
1	-.0017025	.0008221	-2.07	0.039	-.0033198	-.0000853
_cons	1.425571	.0351261	40.58	0.000	1.356467	1.494675

Slope of line for baseline group (non-HBM)

1.hbm + _cons = Y axis intercept for HBM group

Y axis intercept for baseline group (non-HBM)

age + hbm#c.age = Slope of line for HBM group

Test for interaction: reg tb_bmd i.hbm##c.age

```
. reg tb_bmd i.hbm##c.age

      Source |       SS           df       MS      Number of obs   =       328
-------------+----------------------------------   F(3, 324)       =     44.37
       Model |  1.57011575         3  .523371916   Prob > F        =    0.0000
    Residual |  3.82185793       324  .011795858   R-squared       =    0.2912
-------------+----------------------------------   Adj R-squared   =    0.2846
       Total |  5.39197368       327  .016489216   Root MSE        =    .10861

------------------------------------------------------------------------------
      tb_bmd |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       1.hbm |  -.0368218   .0495197    -0.74   0.458    -.1342425    .0605989
         age |  -.0015079   .0005569    -2.71   0.007    -.0026035   -.0004123
             |
   hbm#c.age |
           1 |  -.0017025   .0008221    -2.07   0.039    -.0033198   -.0000853
             |
       _cons |   1.425571   .0351261    40.58   0.000     1.356467    1.494675
------------------------------------------------------------------------------
```

Slope of line for baseline group (non-HBM)

Wald p value for interaction term hbm#c.age

1.hbm + _cons = Y axis intercept for HBM group

Y axis intercept for baseline group (non-HBM)

age + hbm#c.age = Slope of line for HBM group

Post regression command: lincom

```
. lincom age + 1.hbm#c.age

 ( 1)  age + 1.hbm#c.age = 0
```

age + hbm#c.age =
Slope of line for
HBM group

tb_bmd	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
(1)	-.0032104	.0006047	-5.31	0.000	-.0044	-.0020208

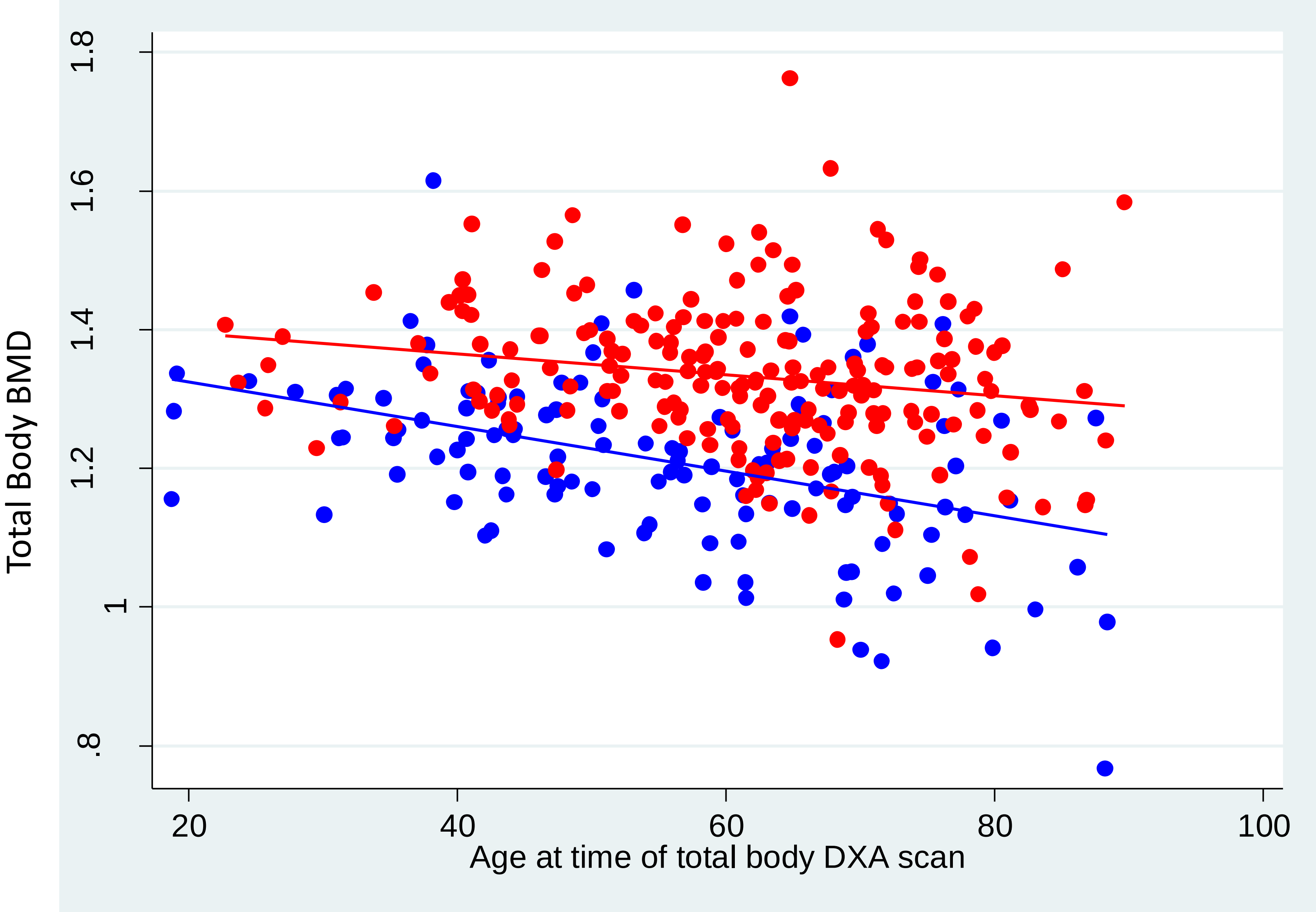
```
. lincom _cons + 1.hbm

 ( 1)  1.hbm + _cons = 0
```

1.hbm + _cons =
Y axis intercept for
HBM group

tb_bmd	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
(1)	1.388749	.034905	39.79	0.000	1.32008	1.457418

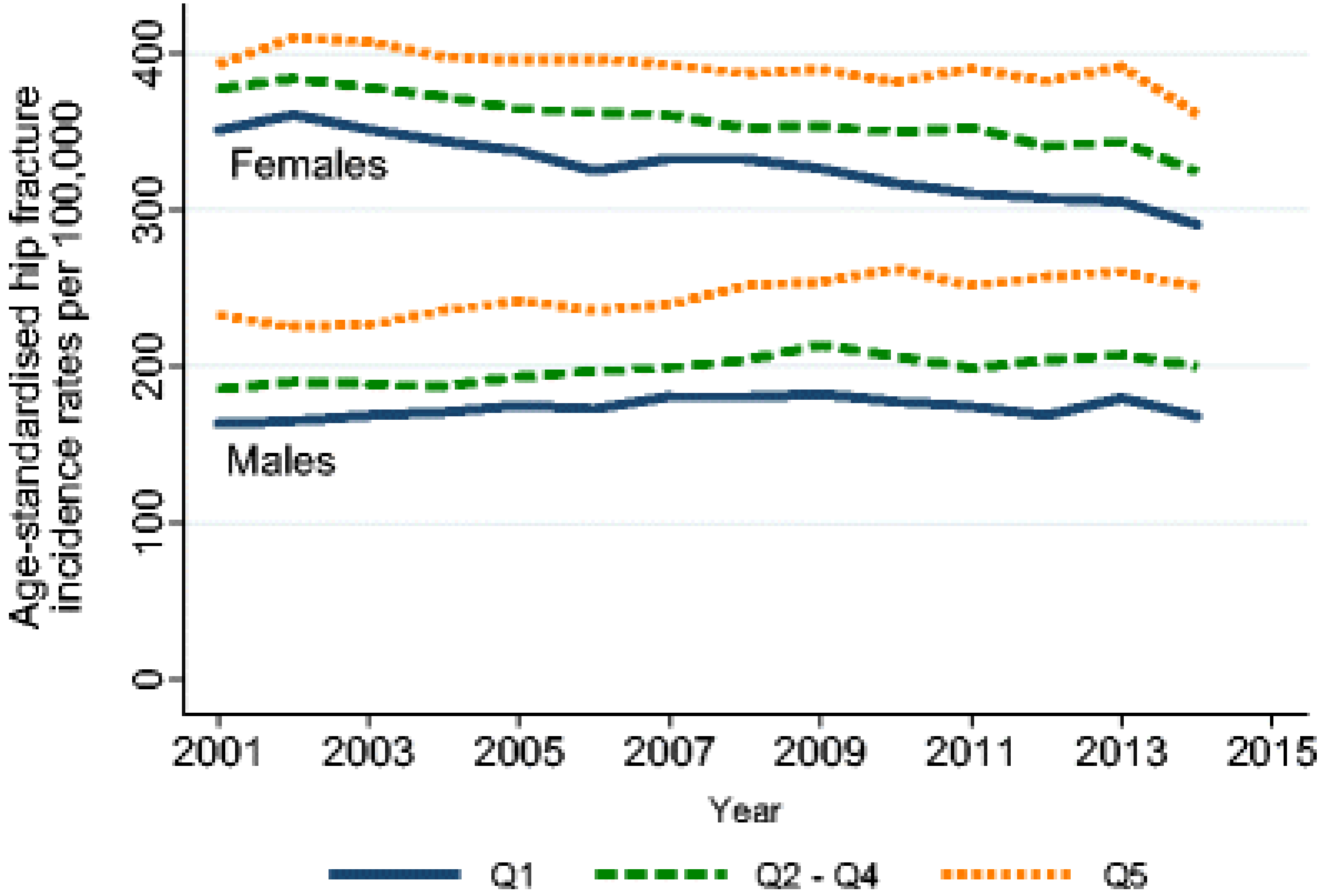
Testing for interaction:
Is the slope of the regression lines different? – Yes!



HBM cases in red
Normal individuals in blue

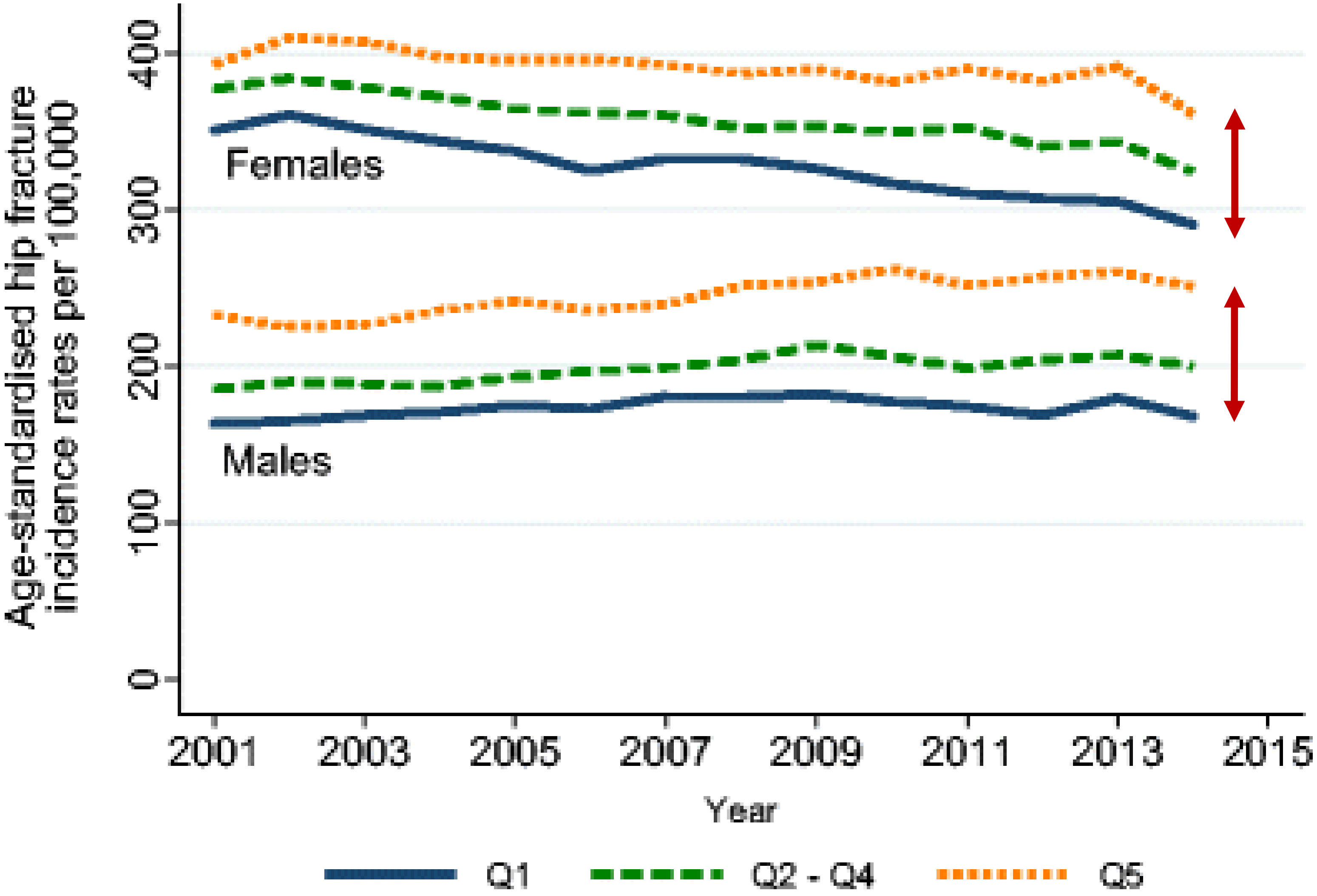
HBM 1
Non-HBM 0

Another example: The effect of social deprivation (quintiles) on hip fracture incidence in England



Bhimjiyani. Osteo Int. 2018. 29 (1), 115-24

Another example: The effect of social deprivation (quintiles) on hip fracture incidence in England



Bhimjiyani. Osteo Int. 2018. 29 (1), 115-24

University of BRISTOL

SAMSON
Sub-Saharan African MuSculOskeletal Network

Summary

		Crude odds/ rate/ risk ratio	Odds/ rate/ risk ratio in Stratum 1	Odds/ rate/ risk ratio in Stratum 2	Adjusted odds/ rate/ risk ratio
Example 1	No confounding No interaction	3.0	3.0	3.0	3.0
Example 2	Confounding No interaction	3.0	2.0	2.0	2.0
Example 3	Interaction	3.0	0.8	5.5	Should not be calculated

University of BRISTOL

SAMSON Sub-Saharan African MuSculOskeletal Network

Summary - Interaction

- Interaction occurs when the presence of one factor ***modifies*** the effect of another on an outcome. *i.e.* the effect of the exposure ***differs*** according to which category of the ***other factor*** is being examined

- Interaction is an ***important property*** of the relationship between two factors, and their influence on an outcome

- You do not try to eliminate this effect, instead you want to ***detect and describe interaction*** in the greatest possible ***detail***

- You ***stratify*** by effect modifiers

- There ***is*** a statistical test for interaction