Models In Epidemiology And Biostatistics
Gordon Hilton Fick
Session 6 : Modification and Confounding With Lines

We have seen how model based methods can enhance the interpretation available through a classic stratified analysis.

Model based methods enable us to see forms of confounding and modification not directly available from stratified analysis.

Next, we begin to explore the many scenarios where modifiers and confounders may be measured rather than placed into groups.

A potential modifier or confounder is measured.

We will see that our previous interpretations in terms of differences (and the differences between differences!) will now translate into the study of the slopes of lines and the differences between slopes of lines (and slopes of slopes!).

We will begin with the simplest of scenarios. For variety, let us now consider modeling the odds of disease.

Modifying A Modifier

Let us suppose that the log odds of disease varies linearly with age.

So, for the 4 exposure/gender groups, there could 4 different lines to describe the relationship between the log odds of disease and age.

Each line may have a different slope and intercept.

For example:

Modeling the odds of disease:

$$p = Pr(D)$$

$$\log(p/(1-p))$$
$$= \beta_0 + \beta_1 G + \beta_2 A + \beta_3 GA + \beta_4 E + \beta_5 GE + \beta_6 AE + \beta_7 GAE$$

We again see that β_7 measures how gender modifies the way age modifies.

The description of this phenomenon changes though... lets see how....

This model determines 4 lines: log odds versus age...

...based on gender and exposure

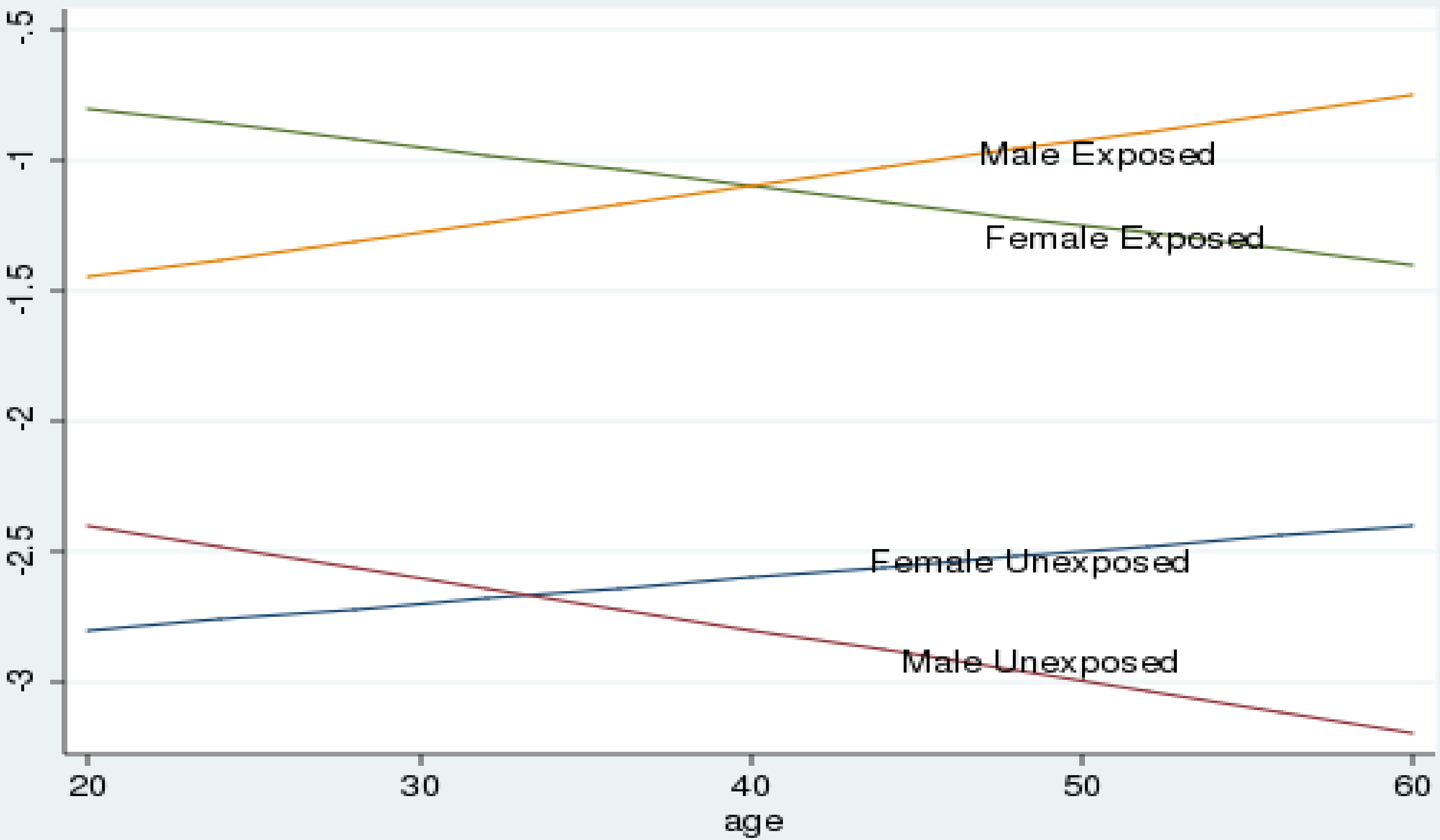
F E $\log(p/(1-p)) = \beta_0 + \beta_4 + (\beta_2 + \beta_6)A$

F $\bar{\mathrm{E}}$ $\log(p/(1-p)) = \beta_0 + \beta_2 A$

ME $\log(p/(1-p)) = \beta_0 + \beta_1 + \beta_4 + \beta_5 + (\beta_2 + \beta_3 + \beta_6 + \beta_7)A$

M$\bar{\mathrm{E}}$ $\log(p/(1-p)) = \beta_0 + \beta_1 + (\beta_2 + \beta_3)A$

Log odds versus age



Exposure

For the males, we consider the vertical distance between the male exposed and the male unexposed to get an age specific difference in log odds.

The exponent of the difference is an age specific OR

For the females, we consider the vertical distance between the female exposed and the female unexposed to get an age specific difference in log odds.

The exponent of the difference is an age specific OR

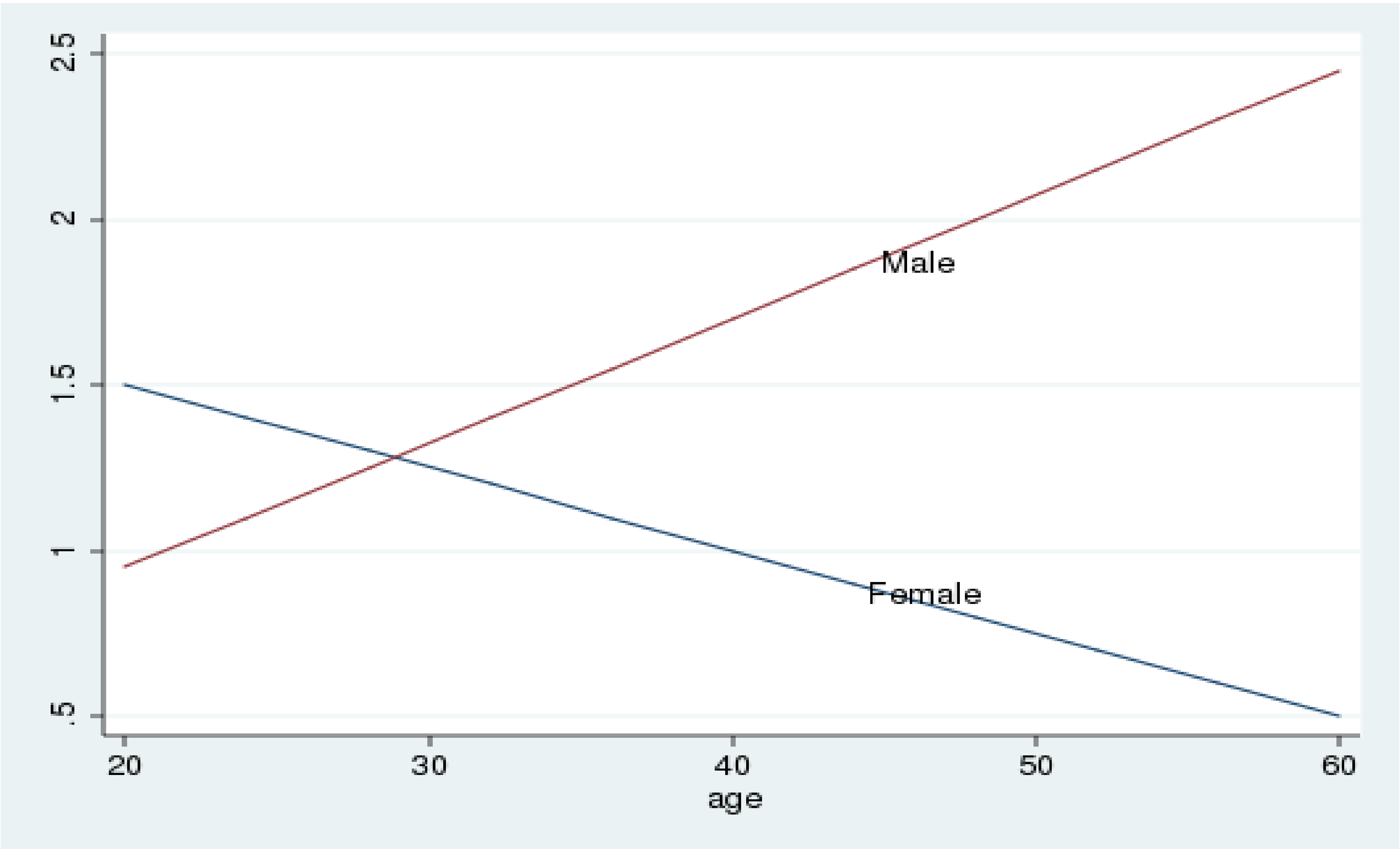
This model gives us 2 lines: log odds ratio versus age

One line for Females and One for Males:

F $\log(\mathrm{OR}) = \beta_4 + \beta_6 A$

M $\log(\mathrm{OR}) = \beta_4 + \beta_5 + (\beta_6 + \beta_7) A$

Log odds exposed minus Log odds unexposed versus age



Interpretation?

Among males, age modifies the disease/exposure relationship in that the OR increases with rising age.

Among females, age modifies the disease/exposure relationship in that the OR decreases with rising age.

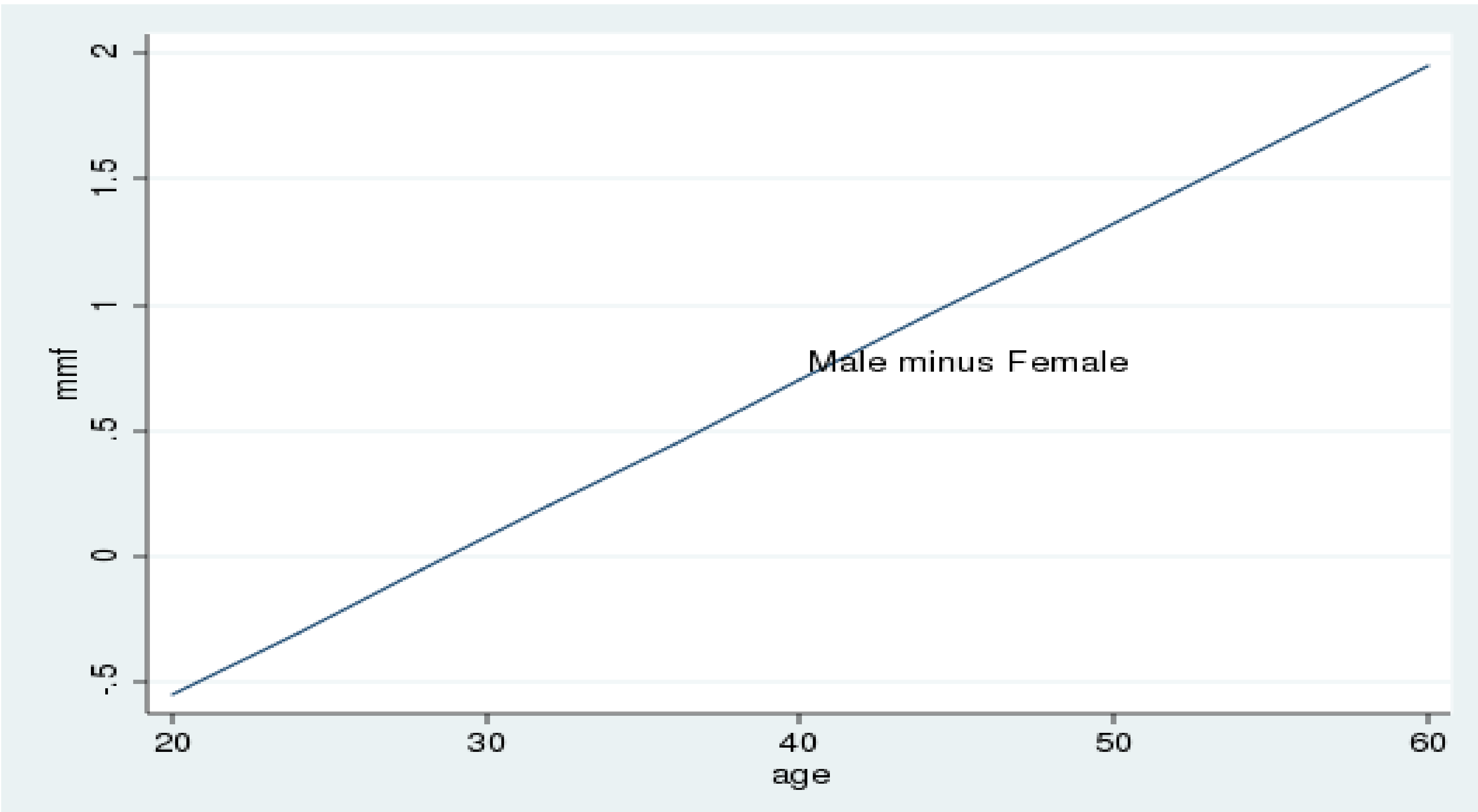
Age modification depends on gender.

This model gives us a line: The log of the ratio of odds ratios versus age

We get: $$\log\left(\frac{\mathrm{OR}_M}{\mathrm{OR}_F}\right)=\beta_5+\beta_7 A$$

If the slope of this line: β_7 is non-zero, then we know that gender modification depends on age.

Difference between the male difference and the female difference



Something new to report?

Age modifies the gender modification.

Gender modifies the age modification.

The appropriate logistic regression model can provide for a regression coefficient to estimate this effect.

The usual LR and Wald tests are then available.

No direct analogue with 'classic' methods

Other models that enable an assessment of modifying a modifier?

There are models that provide differing descriptions of the relationship between the log odds of disease and age among the unexposed.

For an example only, suppose it is reasonable to suppose that, among the unexposed, that the log odds of disease versus age does not depend on gender.

If $\beta_7 \neq 0$ and $\beta_6 \neq 0$ and $\beta_5 \neq 0$ and $\beta_6 + \beta_7 \neq 0$

Then the odds ratio depends on age for both males and females but the degree of the dependency is different between males and females

What if $\beta_7 \neq 0$ but $\beta_6 = 0$? Then the odds ratio depends on age for the males but not for the females.

If the odds ratio depends on age for females and not for the males, then $\beta_7 \neq 0$ and $\beta_6 \neq 0$ and $\beta_6 + \beta_7 = 0$

One could reconsider the model with gender reverse coded to have G=0 for men and G=1 for women.

The next model:

Continuing as before:

$$p = Pr(D)$$

$$\log(p/(1-p))$$
$$= \beta_0 + \beta_1 \mathrm{G} + \beta_2 \mathrm{A} + \beta_3 \mathrm{GA} + \beta_4 \mathrm{E} + \beta_5 \mathrm{GE} + \beta_6 \mathrm{AE}$$

There are many examples....

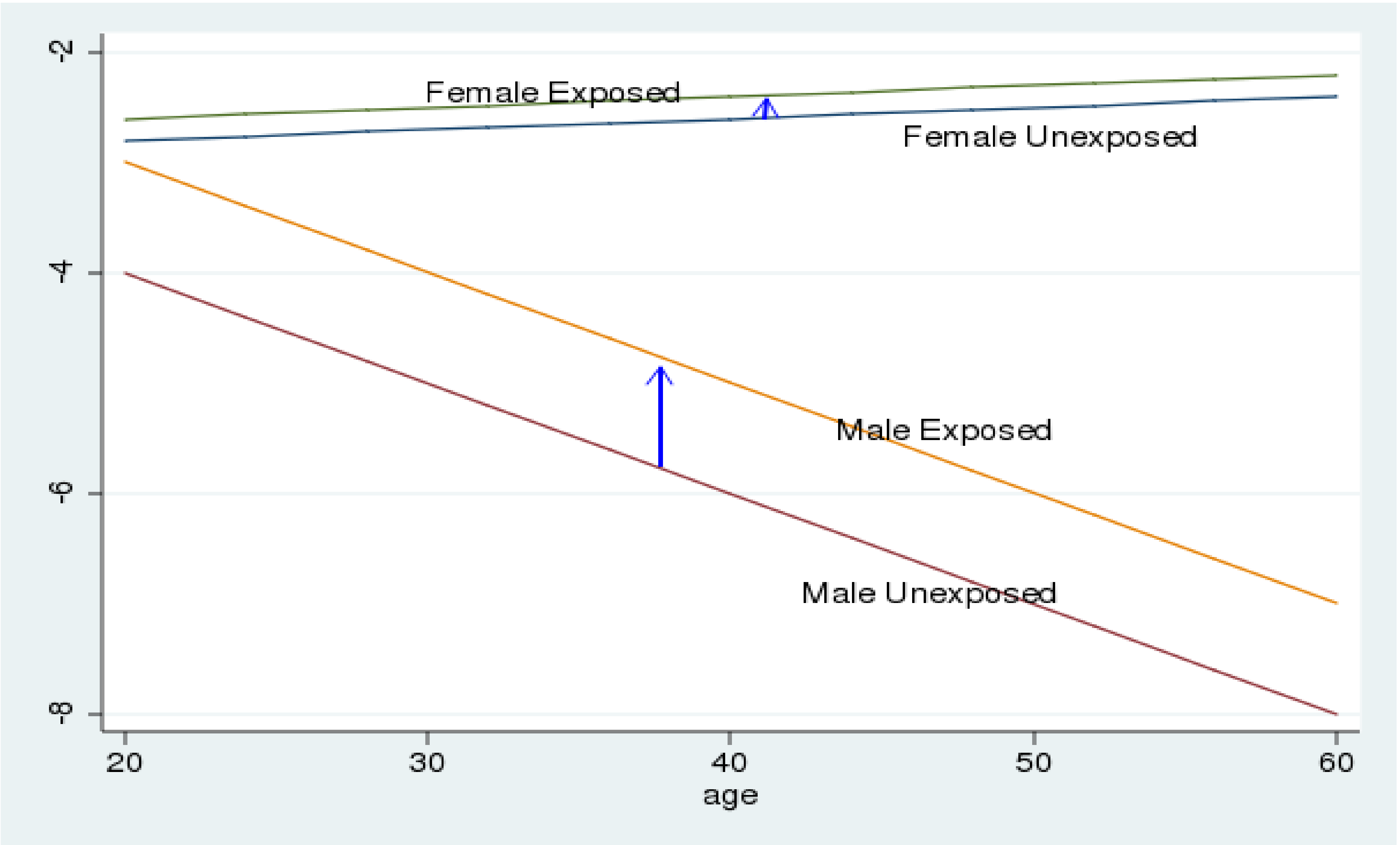
Confounding A Modifier

Let us again suppose that the log odds of disease varies linearly with age.

So, for the 4 exposure/gender groups, there could 4 different lines to describe the relationship between the log odds of disease and age.

Here is another scenario

Log Odds versus age



Gender modifies

The vertical distance (blue arrow) for males is much larger than the vertical distance (blue arrow) for the females

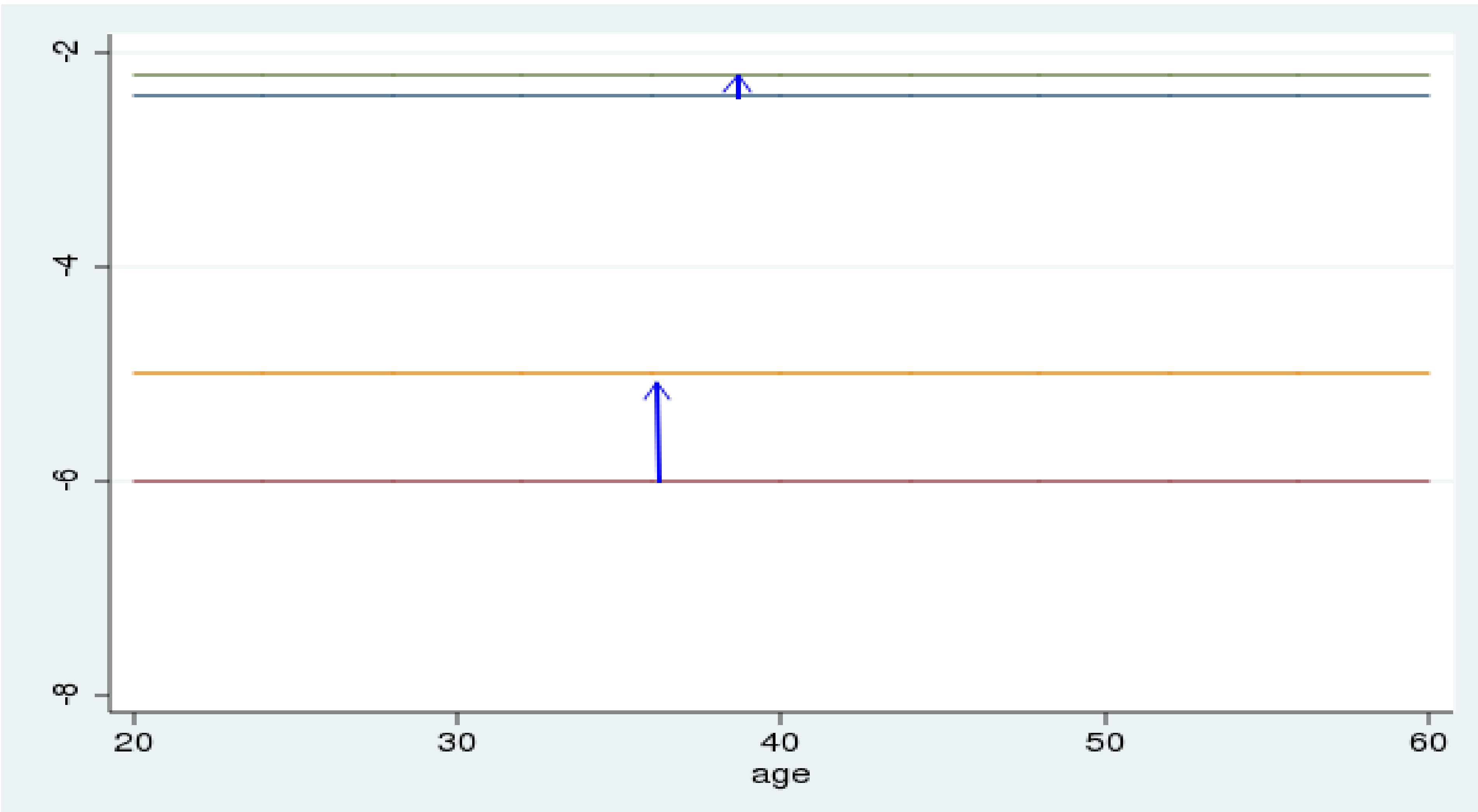
But age does not modify

This picture illustrates gender modification adjusted for age.

Suppose we had not adjusted for age

... and suppose the picture then looks like --->

Log Odds versus Age



Confounding?

Without adjustment for age, the 2 blue arrows are the same length as the 2 blue arrows with adjustment

We can see the gender modification with or without adjustment for age.

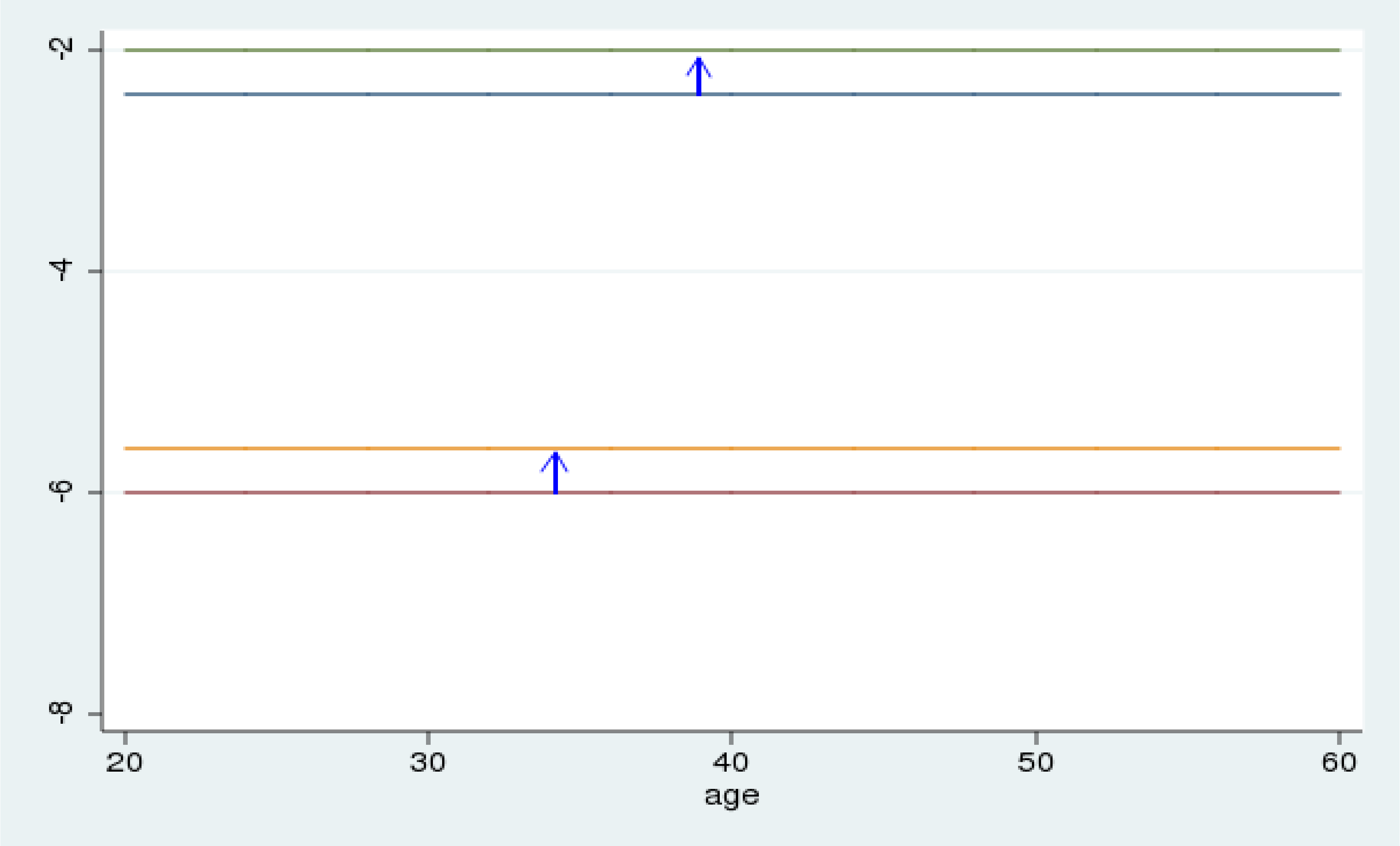
Age is not a confounder

There are (many) other situations...

Now suppose we had not adjusted for age

...and we saw the next picture

Log odds versus age



Gender modification is missed!

The lengths of the 2 blue arrows without adjustment are different from the lengths of the 2 blue arrows with adjustment

So age confounds the gender modification.

To correctly see the gender modification we must adjust for age.

2 Models: β_5 being compared

$$\log(p/(1-p)) = \beta_0 + \beta_1 G + \beta_2 A + \beta_3 GA + \beta_4 E + \beta_5 GE$$

$$\log(p/(1-p)) = \beta_0 + \beta_1 G \qquad\qquad\qquad + \beta_4 E + \beta_5 GE$$

A group of terms handles the adjustment for age

The correct adjustment for age may require both the terms: A and GA

Notice that the term GA enables a proper assessment of confounding and is not a term for assessment of modification

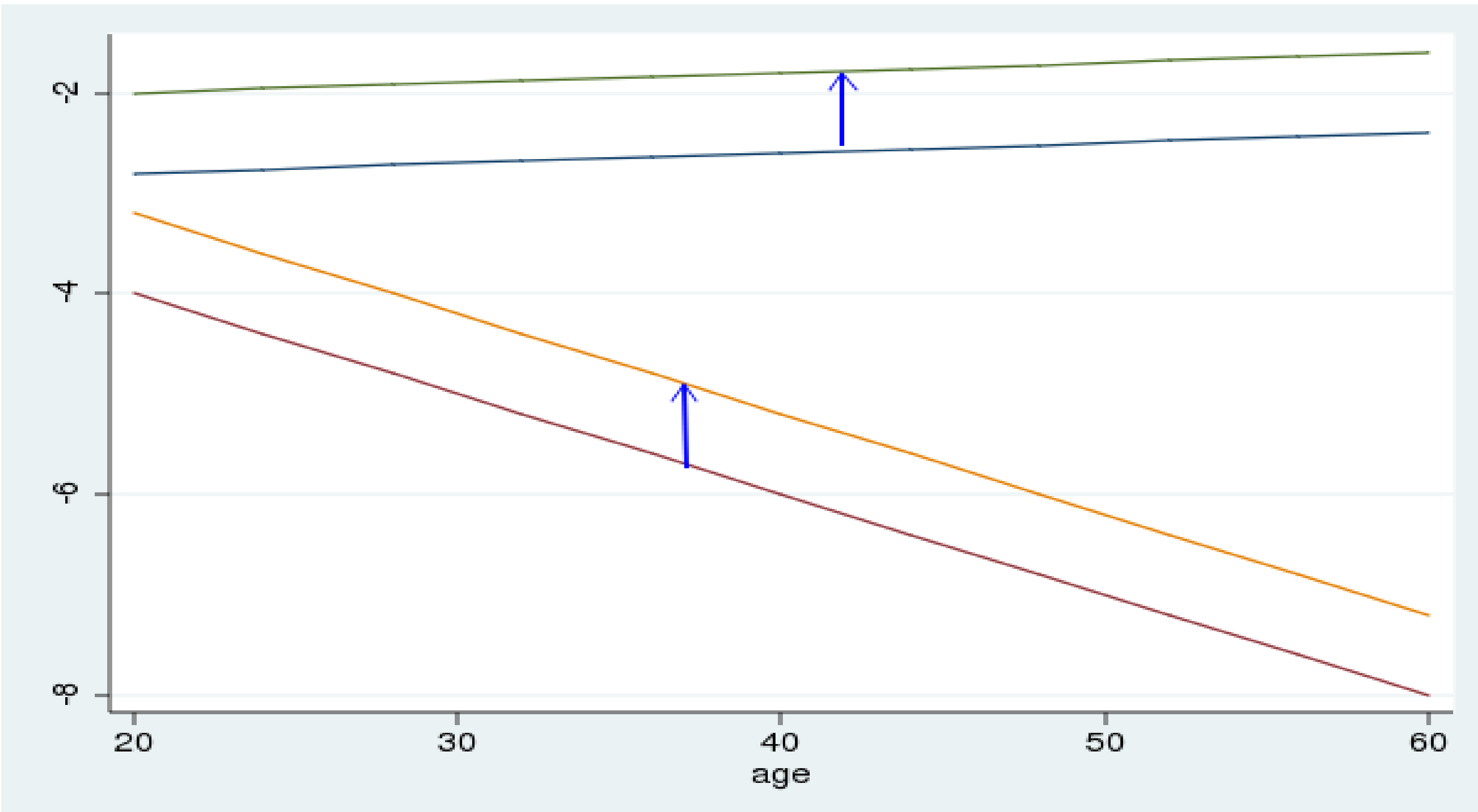
Confounding A Confounder

Let us again suppose that the log odds of disease varies linearly with age.

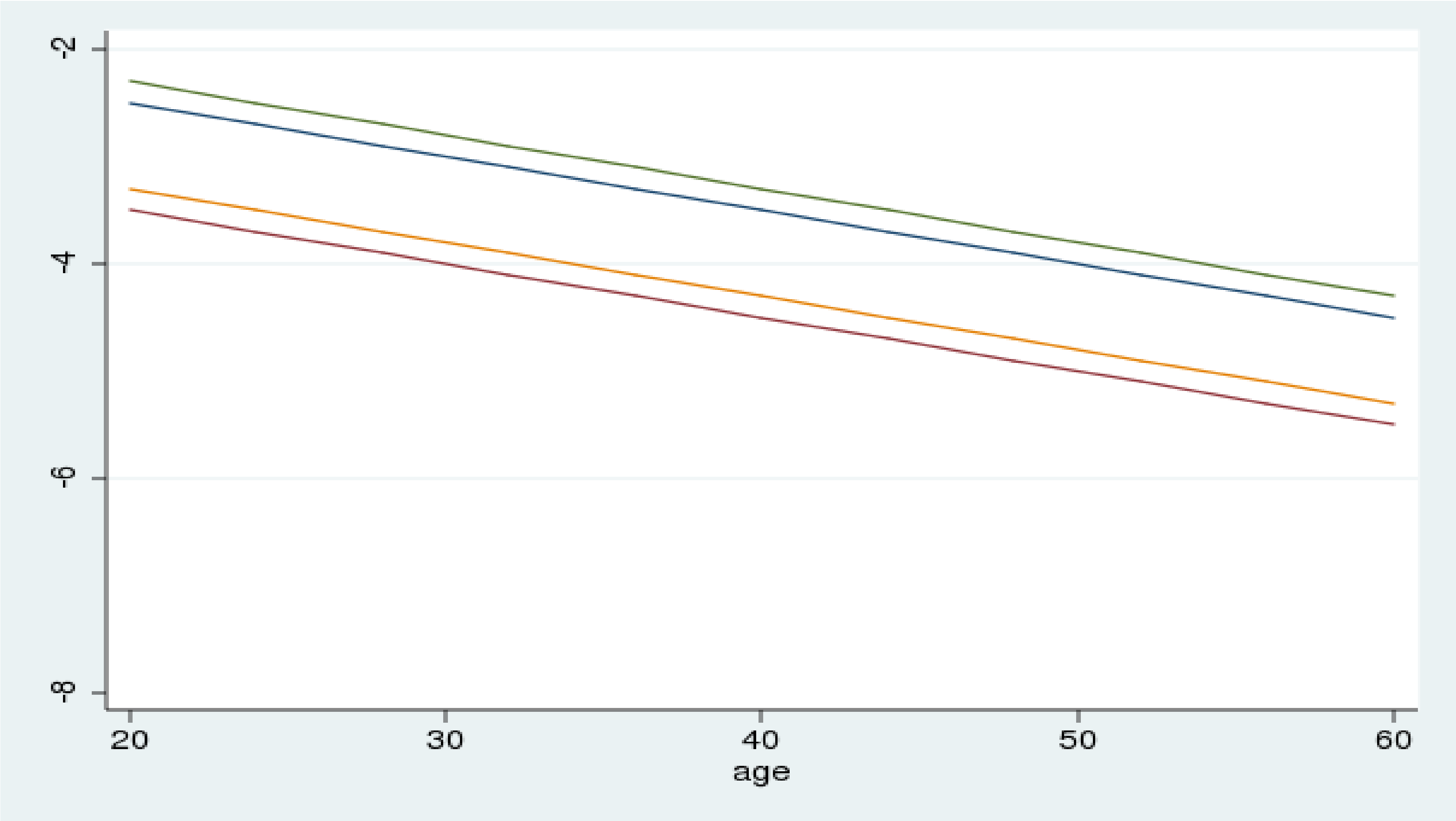
So, for the 4 exposure/gender groups, there could 4 different lines to describe the relationship between the log odds of disease and age.

Here is another scenario

Log Odds Versus Age



What we force all 4 lines to have the same slope?



Confounding?

With the correct adjustment for age and gender, we see a disease/exposure relationship

With an incorrect (too simple) adjustment for age and gender, we miss seeing the disease/exposure relationship

Log Odds versus Age

$$\log(p/(1-p)) = \beta_0 + \beta_1 G + \beta_2 A + \beta_3 GA + \beta_4 E$$

$$\log(p/(1-p)) = \beta_0 + \beta_1 G + \beta_2 A + \beta_4 E$$

If β_4 is different in these 2 models, then the 'simpler' model (that forces parallel lines) is incorrect.

To correctly 'see' that age and gender confound, one must adopt the more complex model.

Is simple enough?

$$\log(p/(1-p)) = \beta_0 + \beta_1 \mathrm{G} + \beta_2 \mathrm{A} + \beta_4 \mathrm{E}$$

Many investigators start (and stop) here.

OK?

Well.... what are the assumptions implicit in such a simple (looking) model?