Models In Epidemiology And Biostatistics
Gordon Hilton Fick

Two Modifiers/Confounders

Let us now consider Age (Young/Old) and Gender (Female/Male) as potential modifiers and/or confounders

Now there are 4 strata:

Young Females	Old Females
Young Males	Old Males

The 4 Strata

	Young	Old
Female	Strata 1	Strata 2
Male	Strata 3	Strata 4

The 8 Probabilities of Exposure
The Corresponding 8 Odds of Exposure

	Young	Old
Female	Case	Case
	Control	Control
Male	Case	Case
	Control	Control

The 4 Odds Ratios

	Young	Old
Female	Odds Ratio	Odds Ratio
Male	Odds Ratio	Odds Ratio

.... where 'Odds Ratio' here is

$$\frac{\text{Odds of Exposure for Cases}}{\text{Odds of Exposure for Controls}}$$

A Classical analysis?

A classical analysis of the 2x2 tables involves 3 assessments:

1) Assess the 4 strata: age and gender together

2) Assess age alone: 2 strata ignoring gender

3) Assess gender alone: 2 strata ignoring age

Sometimes the first analysis is called a 'joint' analysis

Sometimes the second and third analyses are called 'one-at-a-time' analyses

The 4 Strata: Test For Homogeneity of Odds Ratios

The Mantel Haentszel Test

With 3 degrees of freedom

Tests whether all 4 odds ratios are equal

Evidence against the null hypothesis is evidence that the 4 stratum specific odds ratios are not all the same

An omnibus[1] test

[1] omnibus - of, dealing with, or providing for many different things or cases

- providing for many things at once

The 4 Strata: The assumed common odds ratio

If there is no evidence that the 4 odds ratios are different, consider an odds ratio that assumes all 4 are the same

The Mantel-Haentszel estimate of the assumed common odds ratio : the 'adjusted' estimate

Compare with the 'crude' estimate

If the 'crude' is different from the 'adjusted', there is evidence of confounding,

If so, test the null hypothesis that the assumed common odds ratio is one.

One-at-a-time: Age assessment

Provisional assessment of age alone ignoring gender

Two 2x2 tables as in the previous class

Would this analysis alone have identified age as a modifier?

Would this analysis alone have identified age as a confounder?

One-at-a-time: Gender assessment

Provisional assessment of gender alone ignoring age

Two 2x2 tables as in the previous class

Would this analysis alone have identified gender as a modifier?

Would this analysis alone have identified gender as a confounder?

Interpretation that combines the 3 analyses

Does the first analysis based on the 4 strata provide information about modification or confounding not identified by the 2 'one-at-a-time' analyses?

If so, then the 2 'one-at-a-time' analyses are too simple. They involve assumptions that are discredited by the 'joint' analysis.

How does one then report the nature and form of the identified modification and/or confounding?

A model based approach can enhance this process

But... my, my... there are now SO MANY models

There are 8 unknowns:

the 8 probabilities :

2 with each of the four 2-by-2 tables

Hence there can be as many as 8 regression coefficients

And so there are, in principle, 256 = 2^8 models

Fortunately, we can exclude many of them from our attention

It may be best to start with the model that has all 8 regression coefficients

It looks like:

$$\log(p/(1-p)) =$$
$$\beta_0 + \beta_1 D + \beta_2 A + \beta_3 G + \beta_4 AG + \beta_5 AD + \beta_6 GD + \beta_7 AGD$$

... where p is the probability of exposure

This model gives us 8 log odds

		Y	O
F	D	$\beta_0+\beta_1$	$\beta_0+\beta_1+\beta_2+\beta_5$
	$\bar{D}$	β_0	$\beta_0+\beta_2$
M	D	$\beta_0+\beta_1+\beta_3+\beta_6$	$\beta_0+\beta_1+\beta_2+\beta_3+\beta_4+\beta_5+\beta_6+\beta_7$
	$\bar{D}$	$\beta_0+\beta_3$	$\beta_0+\beta_2+\beta_3+\beta_4$

This model then gives us 4 log odds ratios

	Young	Old	
Difference Female	β_1	$\beta_1+\beta_5$	β_5
Male	$\beta_1+\beta_6$	$\beta_1+\beta_5+\beta_6+\beta_7$	$\beta_5+\beta_7$
Difference	β_6	$\beta_6+\beta_7$	β_7

What if $\beta_7 \neq 0$?

Both age and gender modify.

But there is more:

The age modification depends on gender

AND the gender modification depends on age.

These are symmetrical statements. The first statement implies the second statement and vice versa.

What if $\beta_7 = 0$?

We can consider the model:

$$\log(p/(1-p)) = \beta_0 + \beta_1 D + \beta_2 A + \beta_3 G + \beta_4 AG + \beta_5 AD + \beta_6 GD$$

This model gives us 8 log odds

		Y	O
F	D	$\beta_0+\beta_1$	$\beta_0+\beta_1+\beta_2+\beta_5$
	$\bar{D}$	β_0	$\beta_0+\beta_2$
M	D	$\beta_0+\beta_1+\beta_3+\beta_6$	$\beta_0+\beta_1+\beta_2+\beta_3+\beta_4+\beta_5+\beta_6$
	$\bar{D}$	$\beta_0+\beta_3$	$\beta_0+\beta_2+\beta_3+\beta_4$

This model then gives us 4 log odds ratios

	Young	Old	
Difference Female	β_1	$\beta_1+\beta_5$	β_5
Male	$\beta_1+\beta_6$	$\beta_1+\beta_5+\beta_6$	β_5
Difference	β_6	β_6	

Now what?

The age modification is measured by β_5

If $\beta_5 \neq 0$, then the age is a modifier and this modification does not depend on gender.

The gender modification is measured by β_6

If $\beta_6 \neq 0$, then the gender is a modifier and this modification does not depend on age.

Modification description

If this model displays that both age and gender modify, we can then determine if the 2 'one-at-a-time' models (each with 4 terms) identify gender modification of the same form as the 'big model' with 8 terms.

If not, we sometimes say that age and gender jointly modify in that the age modification and the gender modification are only seen through the joint analysis.

Other possibilities?

The age modification is measured by β_5

If $\beta_5 \neq 0$ and $\beta_6 = 0$ then the age is a modifier and gender is not a modifier.

What if $\beta_5 \neq 0$ and $\beta_6 = 0$?

We can consider the model:

$$\log(p/(1-p)) = \\ \beta_0 + \beta_1 D + \beta_2 A + \beta_3 G + \beta_4 AG + \beta_5 AD$$

This model gives us 8 log odds

		Y	O
F	D	$\beta_0+\beta_1$	$\beta_0+\beta_1+\beta_2+\beta_5$
	$\bar{D}$	β_0	$\beta_0+\beta_2$
M	D	$\beta_0+\beta_1+\beta_3$	$\beta_0+\beta_1+\beta_2+\beta_3+\beta_4+\beta_5$
	$\bar{D}$	$\beta_0+\beta_3$	$\beta_0+\beta_2+\beta_3+\beta_4$

This model then gives us 4 log odds ratios

	Young	Old	
Difference Female	β_1	$\beta_1 + \beta_5$	β_5
Male	β_1	$\beta_1 + \beta_5$	β_5
Difference	zero	zero	

Now what?

Age is a modifier.

But what can we say about gender?

Is the age modification confounded by gender?

One way to address this question is to compare the modification determined from this model with the modification from the 'one-at-a-time' model:

$$\log(p/(1-p)) = \beta_0 + \beta_1 D + \beta_2 A + \beta_5 AD$$

What if $\beta_5 = 0$ and $\beta_6 \neq 0$?

We can consider the model:

$$\log(p/(1-p)) =$$
$$\beta_0 + \beta_1 D + \beta_2 A + \beta_3 G + \beta_4 AG + \beta_6 GD$$

This model then gives us 4 log odds ratios

	Young	Old	Difference
Female	β_1	β_1	
Male	$\beta_1 + \beta_6$	$\beta_1 + \beta_6$	
Difference	β_6	β_6	

Now what?

Gender is a modifier.

But what can we say about age?

Is the gender modification confounded by age?

One way to address this question is to compare the modification determined from this model with the modification from the 'one-at-a-time' model:

$$\log(p/(1-p))=\beta_0+\beta_1 D+\beta_3 G+\beta_6 GD$$

What if $\beta_5 = 0$ and $\beta_6 = 0$?

We can consider the model:

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

This model enables an assessment of the disease /exposure relationship 'adjusting' for age and gender.

Now we can ask whether age or gender confound.

Confounding now?

One way to assess whether age or gender confound is the compare β_1 from this model with β_1 from various other models. Perhaps the 3 first choices for comparison are the 2 'one-at-a-time' models and the 'crude' model:

$$\log(p/(1-p)) = \beta_0 + \beta_1 D + \beta_3 G$$

$$\log(p/(1-p)) = \beta_0 + \beta_1 D + \beta_2 A$$

$$\log(p/(1-p)) = \beta_0 + \beta_1 D$$

Confounding description

We can then determine if the way age and gender confound as seen by the model with 5 terms is seen in the same form by the 'one-at-a-time' models each with 3 terms.

If not, we sometimes say that age and gender jointly confound in that the way age and gender confound is only seen through the joint analysis.

Forms of Confounding

Returning to the model:

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

The Mantel-Haentzsel estimate of the assumed common odds ratio from a classical stratified analysis is analogous to the estimate of β_1 from this model above. The 2 estimates will 'typically' be very 'close'

Another model...

...that looks 'simpler' is:

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

The estimate of β_1 from this model is not directly comparable with the MH estimate from a classical analysis without an additional assumption.

This assumption can be stated as:

The difference between the log odds of exposure for the old and for the young does not depend on gender.

Another look at the 4 strata

As we saw earlier, age and gender determine 4 strata:

Strata

1: Young Females

2: Old Females

3: Young Males

4: Old Males

Indicators for the 4 Strata

Let S_1, S_2, S_3, and S_4 be indicators for the 4 strata:

	S_1	S_2	S_3	S_4
YF	**1**	**0**	**0**	**0**
OF	**0**	**1**	**0**	**0**
YM	**0**	**0**	**1**	**0**
OM	**0**	**0**	**0**	**1**

Then we could consider the model:

Using the 4 indicators for the strata:

$$\log(p/(1-p)) = \beta_0 + \beta_1 D + \beta_2 S_2 + \beta_3 S_3 + \beta_4 S_4 + \beta_5 S_2 D + \beta_6 S_3 D + \beta_7 S_4 D$$

This model gives exactly the same fitted values as...

...our first model:

$$\log(p/(1-p)) = \beta_0 + \beta_1 D + \beta_2 A + \beta_3 G + \beta_4 AG + \beta_5 AD + \beta_6 GD + \beta_7 AGD$$

The test: $\beta_5 = \beta_6 = \beta_7 = 0$ is the same with both models. This test is the same as the test for homogeneity of odds ratios in a classic stratified analysis.

But the individual coefficients in the 2 models address different roles.

More on this

The 2 models:

$$\log(p/(1-p)) = \beta_0 + \beta_1 D + \beta_2 S_2 + \beta_3 S_3 + \beta_4 S_4$$

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

Gives the same fitted values and the same estimate of β_1

More on testing and interpretation

Lets reconsider the first model:

$$\log(p/(1-p)) =$$
$$\beta_0 + \beta_1 D + \beta_2 A + \beta_3 G + \beta_4 AG + \beta_5 AD + \beta_6 GD + \beta_7 AGD$$

It is possible to consider any of the regression coefficients:

For example, if $\beta_4 = 0$, then, among the controls, the difference between the log odds of exposure for the old and the young does not depend on gender.