Confounder & effect modifier

What are extraneous variables?

- Extraneous variables are variables that occur outside of the exposure-disease relationship.
- Extraneous variables can be:
  - confounding variables
  - effect modifying variables

Two main complications of analysis of single exposure effect

(1) Effect modifier - useful information

(2) Confounding factor - bias

Confounding

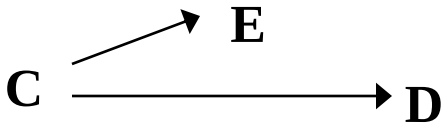
- A confounding variable is associated with the exposure and it affects the outcome, but it is not an intermediate link in the chain of causation between exposure and outcome.
- Should be prevented or Needs to be controlled for

Confounding - Causal Diagrams

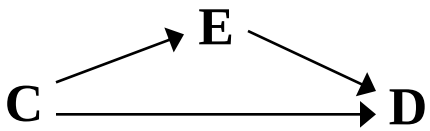
E = Exposure

D = Disease

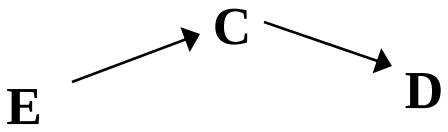
C = Potential Confounder



An apparent association between E and D is completely explained by C. C is a confounder.

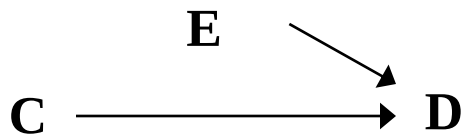


An association between E and D is partly due to variations in C. C is a confounder.

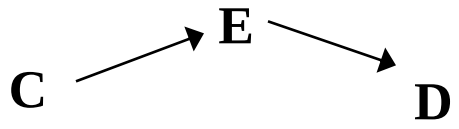


C is in the causal path between E and D, a confounder.

Confounding - Causal Diagrams



C has an independent effect on D. C is not a confounder.



The effect of C on D is completely contained in E. C is not a confounder

How to prevent/control confounding?

Prevention (Design Stage)

- Restriction to one stratum
- Matching

Control (Analysis Stage)

- Stratified analysis
- Multivariable analysis

Restriction

- The classical procedure for preventing confounding at the design stage is restriction.
- This simply means not allowing people that are exposed to a potential confounder to participate in the study.
- Since the restricted sample does not include people exposed to the confounder the effect of the confounder cannot distort the study's estimates.

Randomization

- In randomization, an investigator randomly assigns an exposure to the study participants and then looks for an effect.
- This is NOT to be confused with random selection of a study sample.
- When the exposure is randomly assigned, the law of large numbers ensures an equal distribution of the confounder in exposed and nonexposed groups—hence, no exposure-confounder association (the dotted line on the confounding triangle).
- There can therefore be no confounding.

Matching

- Matching involves making sure that groups being compared are exact matches in terms of the distribution of confounding variables.
- Matching is a type of partial restriction: it does not prevent all subjects exposed to a confounder from participating in a study, but it places some restrictions on who can participate.
- Like restriction and randomization, matching targets the left-hand side of the confounding triangle—eliminating the exposure-confounder association.

Regression models

- Regression models produce adjusted estimates (i.e., adjusted for the effects of a confounder, which is included as a variable in the model).
- In general, they involve developing a best-fitting linear equation (of the type introduced in chapter 10), or a transformed version of such a linear equation.
- Time-to-event or survival analysis methods are also used for this purpose.

Stratification

- Stratified analysis divides the contingency tables arising from a study into groups (strata).
- The groups are defined by levels of the confounding variable.
- Since confounding is caused by intermixing, unmixing through stratification will lead to a change in the estimated effect **if** the unstratified (crude) estimate was distorted by confounding.

Stratification: “Series of 2x2 tables”

Idea: Take a 2x2 table and break it into a series of smaller 2x2 tables (one table at each of *J* levels of the confounder yields *J* tables).

Example: in testing for an association between lung cancer and alcohol drinking (yes/no), separate smokers and non-smokers.

Stratification:“Series of 2xK tables”

Idea: Take a 2xK table and break it into a series of smaller 2xK tables (one table at each of *J* levels of the confounder yields *J* tables).

Example: In evaluating the association between lung cancer and being either a teetotaler, light drinker, moderate drinker, or heavy drinker (2x4 table), separate into smokers and non-smokers (two 2x4 tables).

Choosing Confounders for Statistical Adjustment

One school says choice should be based on **a priori** considerations.

Others say choice of confounders should be based on how much they affect RR (OR, RD) when included/ excluded from the model.

Confounding: example

	Lung cancer	No lung cancer
Drinker	**50**	**50**
Non-drinker	**50**	**150**
	100	**200**

50% of cases are drinkers, but only 25% of controls are drinkers.

Therefore, it appears that drinking is strongly associated with lung cancer.

Confounding: example

Smoker

	Lung cancer	No lung cancer
Drinker	45	15
Non-drinker	30	10
	75	25

Among smokers, 45/75=60% of lung cancer cases drink and

15/25=60% of controls drink.

Non-smoker

	Lung cancer	No lung cancer
Drinker	5	35
Non-drinker	20	140
	25	175

Among non-smokers 5/25=20% of lung cancer cases drink and

35/175=20% of controls drink.

Mantel Haenszel Methods

Mantel Haenszel Methods-Notations

	Outcome		
	Experienced event: D (Disease)	Did not experience event: H (Healthy)	Total
Group 1 (exposed)	d_{1i}	h_{1i}	n_{1i}
Group 0 (unexposed)	d_{0i}	h_{0i}	n_{0i}
Total	d_i	h_i	n_i

Mantel Haenszel Methods common odds ratio

(1) **The Mantel-Haenszel estimate of the odds ratio** assumes there is a **common** odds ratio:

$$OR_{pool} = OR_1 = OR_2 = \ldots = OR_K$$

To estimate the common odds ratio we take a weighted average of the stratum-specific odds ratios:

$$OR_{MH} = \frac{\Sigma(w_i \times OR_i)}{\Sigma w_i} = \frac{\Sigma \frac{d_{1i} \times h_{0i}}{n_i}}{\Sigma \frac{d_{0i} \times h_{1i}}{n_i}}$$

Mantel Haenszel Methods common odds ratio

$$95\%\ \mathrm{CI} = \mathrm{OR}_{MH}/\mathrm{EF}\ \text{to}\ \mathrm{OR}_{MH} \times \mathrm{EF},$$

$$\text{where the error factor EF} = \exp(1.96 \times \mathrm{s.e.}_{MH})$$

$$\mathrm{s.e.}_{MH} = \sqrt{[V/(Q \times R)]},$$

$$Q = \Sigma \frac{d_{1i} \times h_{0i}}{n_i},\ R = \Sigma \frac{d_{0i} \times h_{1i}}{n_i},\ V = \Sigma V_i = \Sigma \frac{d_i \times h_i \times n_{0i} \times n_{1i}}{n_i^2 \times (n_i - 1)}$$

Mantel Haenszel Methods common odds ratio

$$\chi^2_{MH} = \frac{(\Sigma d_{1i} - \Sigma E_{1i})^2}{\Sigma V_i} = \frac{(O - E)^2}{V} = \frac{U^2}{V}; \text{ d.f.} = 1$$

$$E_{1i} = \frac{d_i \times n_{1i}}{n_i}$$

$$O = \Sigma d_{1i}, \; E = \Sigma E_{1i} \text{ and } U = O - E$$

Controlling for confounding by stratification

- Example: Gender Bias at Berkeley?

(From: Sex Bias in Graduate Admissions: Data from Berkeley, *Science* 187: 398-403; 1975.)

	Female	Male
Denied	1276	1486
Admitted	559	1195
	1835	2681

Crude RR =
(1276/1835)/(1486/2681)
=1.25 (1.20 - 1.32)

Program A

- Stratum 1 = only those who applied to program A

	Female	Male
Denied	19	314
Admitted	89	511
	108	825

Stratum-specific
RR = .46 (.30-.70)

Program B

- Stratum 2 = only those who applied to program B

	Female	Male
Denied	8	208
Admitted	17	352
	25	560

Stratum-specific
RR = 0.86 (.48-1.54)

Program C

- Stratum 3 = only those who applied to program C

	Female	Male
Denied	391	205
Admitted	202	120
	593	325

Stratum-specific
RR = 1.05 (.94-1.16)

Program D

- Stratum 4 = only those who applied to program D

	Female	Male
Denied	248	265
Admitted	127	142
	375	407

Stratum-specific
RR = 1.02 (.92-1.12)

Program E

- Stratum 5 = only those who applied to program E

	Female	Male
Denied	289	147
Admitted	104	44
	393	191

Stratum-specific
RR = 0.96 (.87-1.05)

Program F

- Stratum 6 = only those who applied to program F

	Female	Male
Denied	321	347
Admitted	20	26
	341	373

Stratum-specific
RR = 1.01 (.97-1.05)

Summary

Crude RR = 1.25 (1.20 – 1.32)

Stratum specific RR's:
.46 (.30-.70)
0.86 (.48-1.54)
1.05 (.94-1.16)
1.02 (.92-1.12)
0.96 (.87-1.05)
1.01 (.97-1.05)

Maentel-Haenszel Summary RR: .97

Cochran-Mantel-Haenszel Test is NS. Gender and denial of admissions are conditionally independent given program.

The apparent association (RR=1.25) was due to confounding.

Cochran-Mantel-Haenszel Test of Conditional Independence

The (Cochran)-Mantel-Haenszel statistic tests the null hypothesis that exposure and disease are independent when conditioned on the confounder.

advantages and limitations

advantages…

- Mantel-Haenszel summary statistic is easy to interpret and calculate
- Gives you a hands-on feel for the data

disadvantages…

- Requires categorical confounders or continuous confounders that have been divided into intervals
- Cumbersome if more than a single confounder

  To control for ≥ 1 and/or continuous confounders, a multivariate technique (such as logistic regression) is preferable.

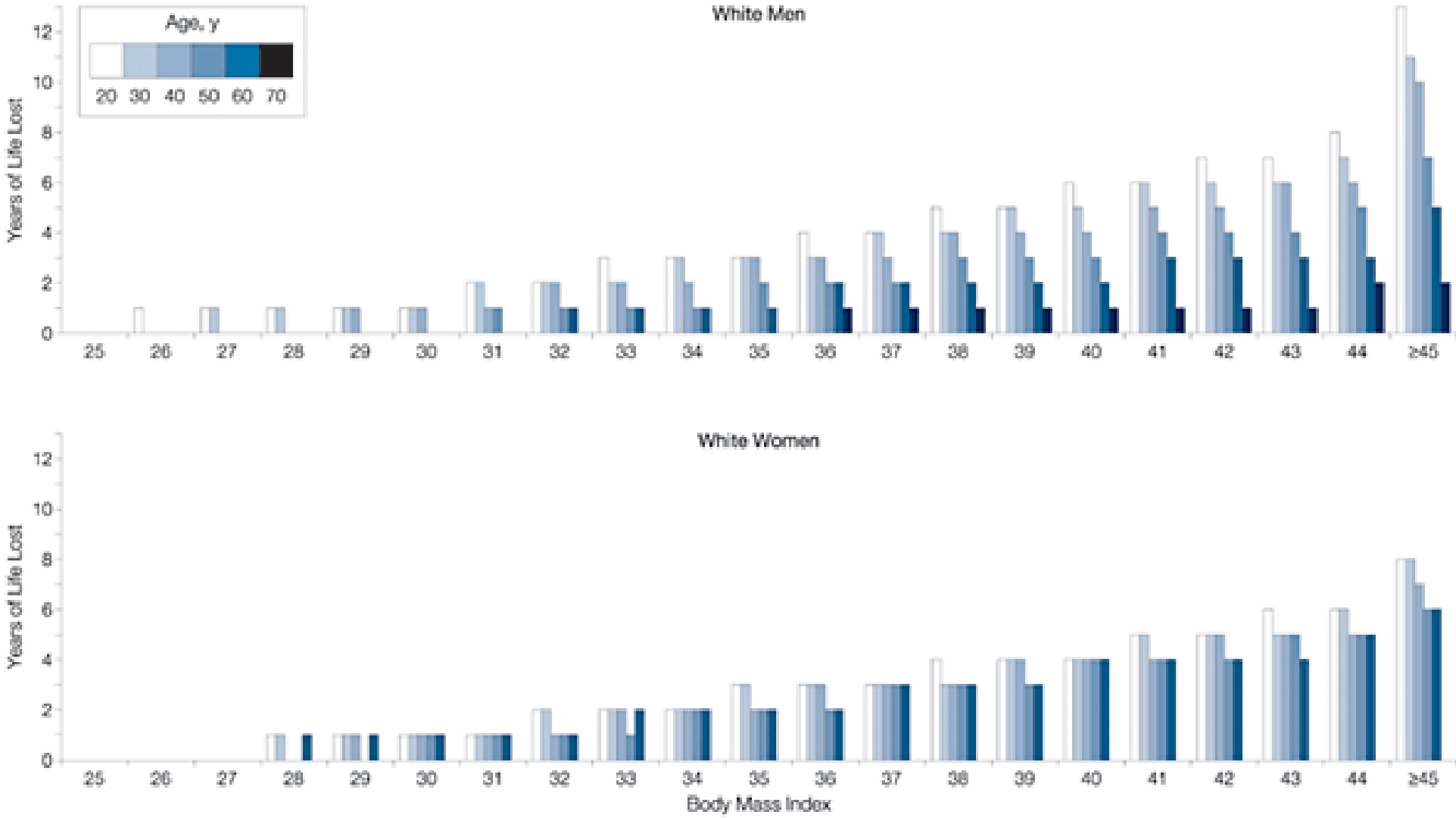
Effect Modification

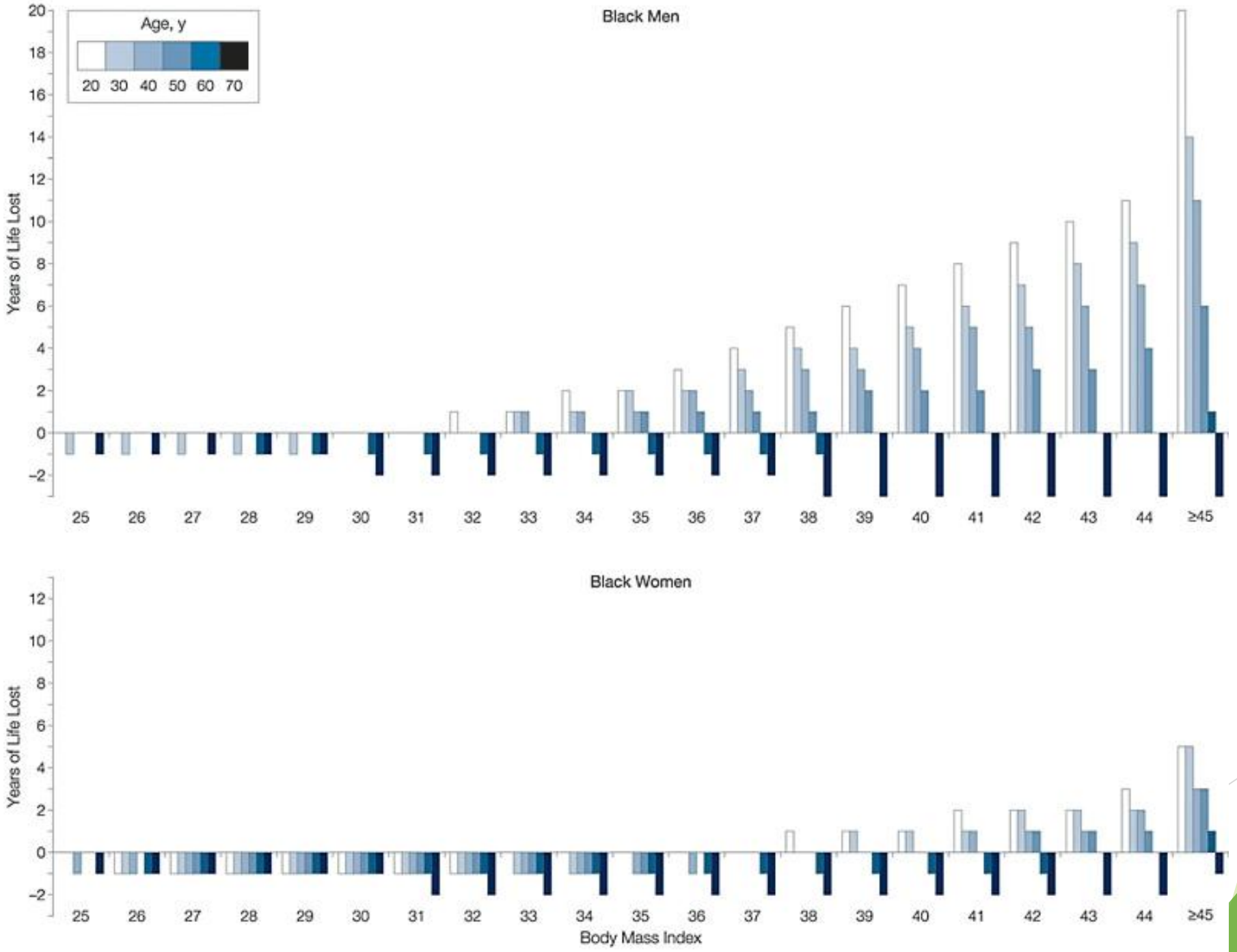
- Effect modification occurs when the effect of an exposure is different among different subgroups. (Variation in the magnitude of measure of effect across levels of a third variable).
- Effect modification is not a bias but useful information

Happens when RR or OR is different between strata (subgroups of population)

Years of Life Lost Due to Obesity (*JAMA*. Jan 8 2003;289:187-193)

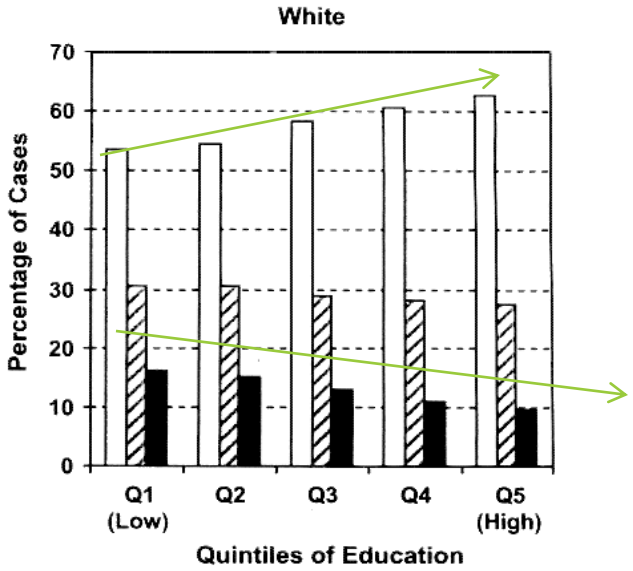
Data from US Life Tables and the National Health and Nutrition Examination Surveys (I, II, III).

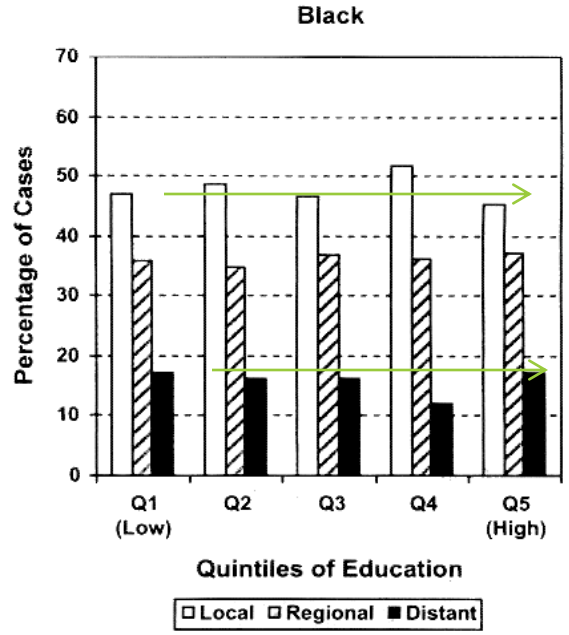

White Men
Age, y
20 30 40 50 60 70
Years of Life Lost
0 2 4 6 8 10 12
25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 ≥45
White Women
Years of Life Lost
0 2 4 6 8 10 12
25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 ≥45
Body Mass Index


Black Men
Age, y
20 30 40 50 60 70
Years of Life Lost
20
18
16
14
12
10
8
6
4
2
0
−2
25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 ≥45
Black Women
Years of Life Lost
12
10
8
6
4
2
0
−2
25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 ≥45
Body Mass Index

Conclusion

Race and gender modify the effect of obesity on years-of-life-lost.





Note. Quintiles of education are based on the percentage of individuals with at least a high school degree: Q1 (<59.7), Q2 (59.7–67.4), Q3 (67.5–73.5), Q4 (73.6–81.0), and Q5 (≥81.1). Chi-square tests for association and trend yielded $P < .05$.

Among white women, stage of breast cancer at detection is associated with education.

However, no clear pattern among black women.

Statistical tests for homogeneity

- Statistical tools can help determine whether 2 stratum-specific estimates are homogeneous.
- In stratified analysis, a test called the Mantel-Haenszel test for homogeneity is commonly employed.
- In modelling, tests for statistical interaction between exposure and potentially modifying variables are often used. These are called tests of interaction.
- The identification of effect modification is a key task in epidemiological analysis. While good procedures exist to test for it, judgement is also needed.

Statistical tests for homogeneity

$$\chi^2 = \Sigma \frac{(d_{1i} \times h_{0i} - \mathrm{OR}_{MH} \times d_{0i} \times h_{1i})^2}{\mathrm{OR}_{MH} \times V_i \times n_i^2}, \text{ d.f.} = c - 1$$

summarize

Effect modifier

- **Belongs to nature**
- **Different effects in different strata**
- **Simple**
- **Useful**
- **Increases knowledge of biological mechanism**
- **Allows targeting of public health action**

Confounding factor

- **Belongs to study**
- **Adjusted OR/RR different from crude OR/RR**
- **Distortion of effect**
- **Creates confusion in data**
- **Prevent (design)**
- **Control (analysis)**

summarize

Randomization breaks any links
between treatment and prognostic factors