EUROPEAN RENAL ASSOCIATION

EUROPEAN DIALYSIS AND TRANSPLANT ASSOCIATION

# ERA-EDTA Registry:

# Evaluation and Pitfalls in Epidemiology

*Confounding: what it is and how to deal with it*

***Kitty Jager, Carmine Zoccali, Alison MacLeod and Friedo W. Dekker***

***14th Budapest Nephrology School***

***Budapest – August 30, 2007***

# Confounding

- 'Mixing' or 'blurring' of effects
- Occurs when an investigator tries to determine the effect of an exposure on the occurrence of a disease/outcome, but then actually measures the effect of another factor, a confounding variable.
- In studies investigating disease etiology and causal relationships, confounding is regarded as undesirable, as it obscures the 'real' effect of an exposure.

AFTER CORRECTION FOR DIABETES

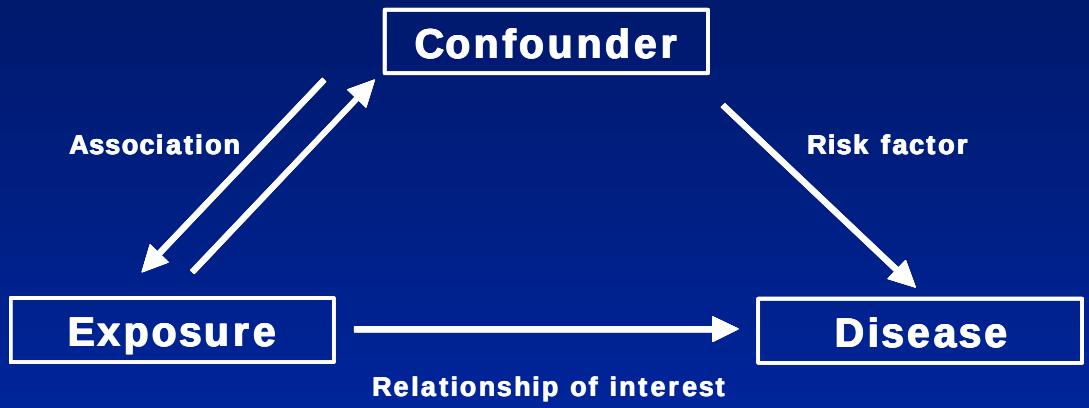
*Independent effect of Kt/V*

*Real effect of ....*

Fig. 7. Graft survival in first renal allograft recipients in different time periods (adjusted for age, gender, diabetes mellitus, and donor type).

- This presentation will explain the concept of confounding and describe a number of ways in which it can be addressed:
  - randomization, restriction, matching, stratification and multivariate analysis

# When are variables potential confounders?




Properties of a potential confounder

(1) the variable must have an association with the disease,
*i.e. it should be a risk factor for the disease;*

(2) it must be associated with the exposure,
*i.e. it must be unequally distributed between the exposed and non-exposed groups; and*

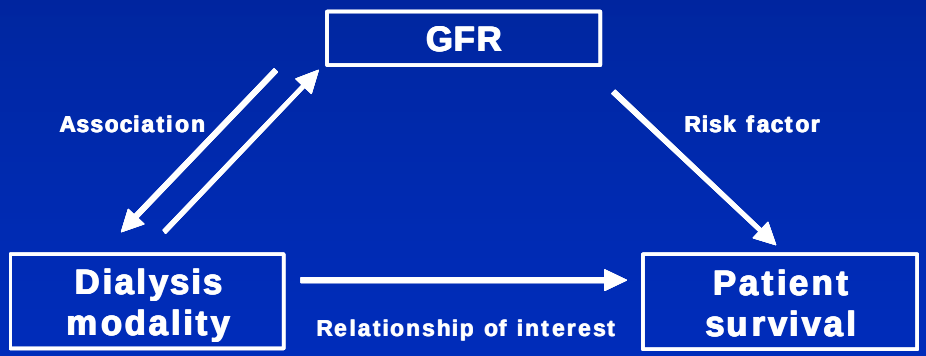
(3) it must not be an effect of the exposure, nor (linked to this) be a factor in the causal pathway of the disease

# Association between initial dialysis modality and patient survival

## *Is GFR a potential confounder?*

*Example 1 – Association between treatment choice and outcome in the elderly with end-stage renal disease (ESRD).*

Couchoud et al.[1] studied the association between initial dialysis modality and 2-year patient survival in a cohort of 3512 elderly ESRD patients. After adjustment for eGFR at dialysis initiation and a number of other factors, unplanned HD was associated with a 50% increased risk of death and PD with a 30% increased risk of death compared with planned HD.




***Yes, GFR is a potential confounder***

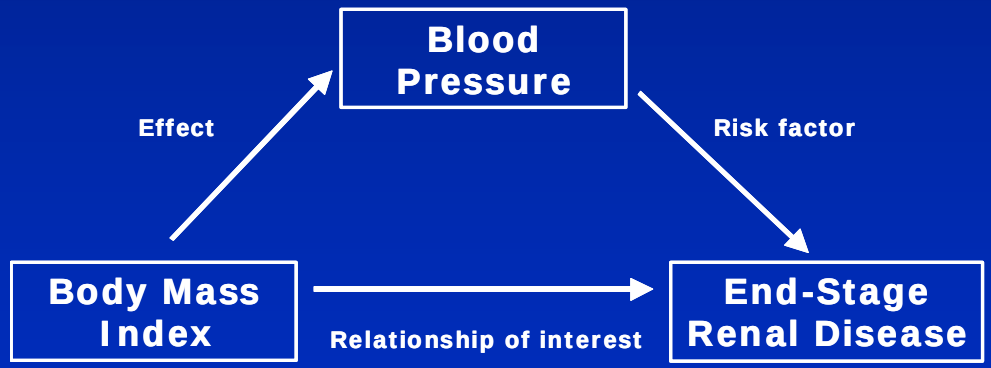
[1] *Couchoud C, Moranne O, Frimat L, Labeeuw M, Allot V, Stengel B. Associations between comorbidities, treatment choice and outcome in the elderly with end-stage renal disease. Nephrol Dial Transplant Advance Access published July 5, 2007.*

# Association between body mass index and the risk of ESRD

## *Is blood pressure a potential confounder?*

*Example 2 - BMI and the risk for ESRD*

Hsu et al.[2] investigated the relationship between BMI and the risk for ESRD using data of more than 320,000 members of Kaiser Permanente. They were able to show that, adjusted for a number of confounders like age, sex and race (but not for blood pressure), increased BMI was strongly associated with an increased risk for ESRD.




***NO, Blood pressure is not a potential confounder***

[2] *Hsu C, McCulloch CE; Iribarren C, Darbinian J, Go AS. Body Mass Index and Risk for End-Stage Renal Disease. Ann Intern Med 2006;144:21-28.*

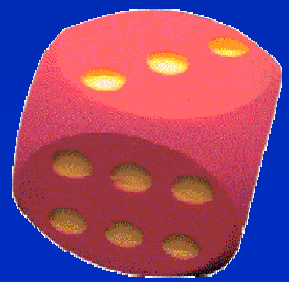
# How to address confounding?

- During study design by
  - randomization
  - restriction
  - matching

- During data analysis by adjustment for confounding using
  - Stratification
  - multivariate analysis

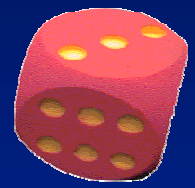
# Randomization

- Random assignment of patients to the experimental group or to a control group
- Helps to prevent selection bias / ‘confounding by indication’ by the clinician
- Any remaining differences between the groups are due to chance



# Randomization

- Large study size helps randomization process to be successful; small studies may however run into problems



*Example 3 - Dialysis modality choice and survival*

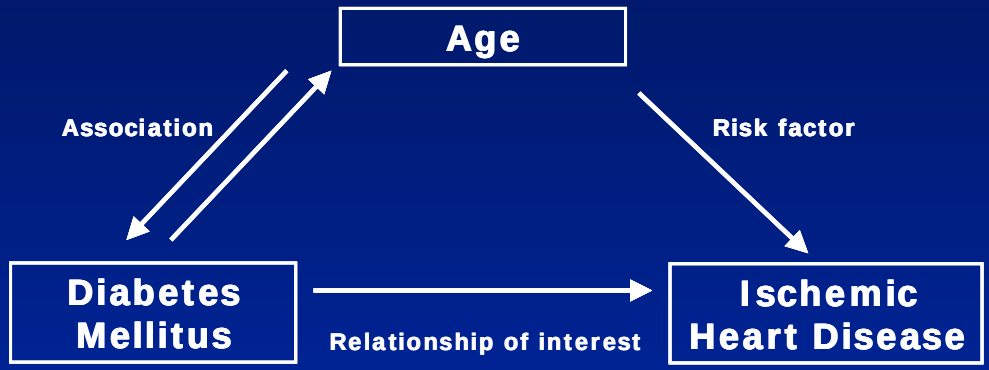
In order to investigate the relationship between dialysis modality and patient survival Korevaar et al.[3] performed an RCT within the NEtherlands COoperative Study on the Adequacy of Dialysis (NECOSAD). Due to problems with patient recruitment this RCT was very small (only including 38 patients).

As a result there were important differences between the groups: patients randomized to PD were about 7 years younger and had less DM and other co-morbidity than those assigned to HD.

- If important differences remain, investigators may adjust for these confounders in their analysis

[3] *Korevaar JC, Feith GW, Dekker FW et al. Effect of starting with hemodialysis compared with peritoneal dialysis in patients new on dialysis treatment: a randomized controlled trial. Kidney Int 2003;64:2222–2228.*

# Other ways to address confounding

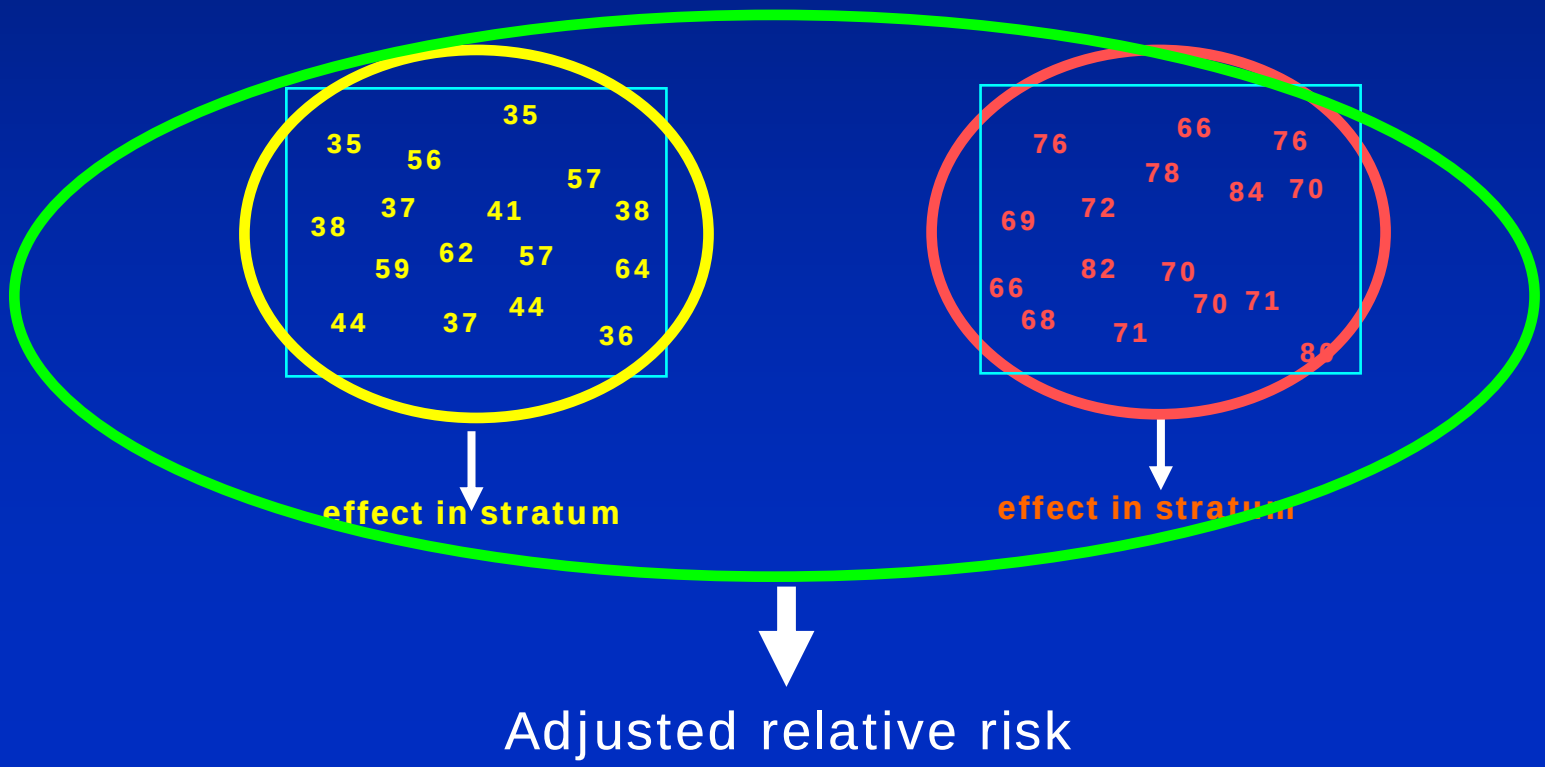



- Restriction: e.g. perform only in patients above 65 years of age
- Matching: e.g. in a cohort study for each 'exposed' person with DM one may select an 'unexposed' person without DM of the same age
  - *Cave: in case-control studies the choice of matching variables requires careful attention*
- Stratification: e.g. calculate relative risks in subgroups according to age and then calculate an adjusted relative risk by pooling or standardization

# Stratification

Stratification: - calculate relative risks in subgroups according to age

- calculate an adjusted relative risk by pooling or standardization



# Stratification

**All RRT patients**

| | | Ischemic heart disease | | | |
|---|---|---|---|---|---|
| | | Yes | No | Total | Proportion with ischemic heart disease |
| Diabetes mellitus | Yes | 184 | 376 | 560 | 32.9% |
| | No | 278 | 1162 | 1440 | 19.3% |

*Crude RR = 32.9% / 19.3% = 1.7*

**RRT patients < 65 years of age**

| | | Ischemic heart disease | | | |
|---|---|---|---|---|---|
| | | Yes | No | Total | Proportion with ischemic heart disease |
| Diabetes mellitus | Yes | 36 | 114 | 150 | 24.0% |
| | No | 136 | 714 | 850 | 16.0% |

*Stratum specific RR =24.0% / 16.0% = 1.5*

**RRT patients ≥ 65 years of age**

| | | Ischemic heart disease | | | |
|---|---|---|---|---|---|
| | | Yes | No | Total | Proportion with ischemic heart disease |
| Diabetes mellitus | Yes | 148 | 262 | 410 | 36.1% |
| | No | 142 | 448 | 590 | 24.1% |

*Stratum specific RR = 36.1% / 24.1% = 1.5*

Stratification into age groups

Pooling / Standardization

**Adjusted RR = 1.5**

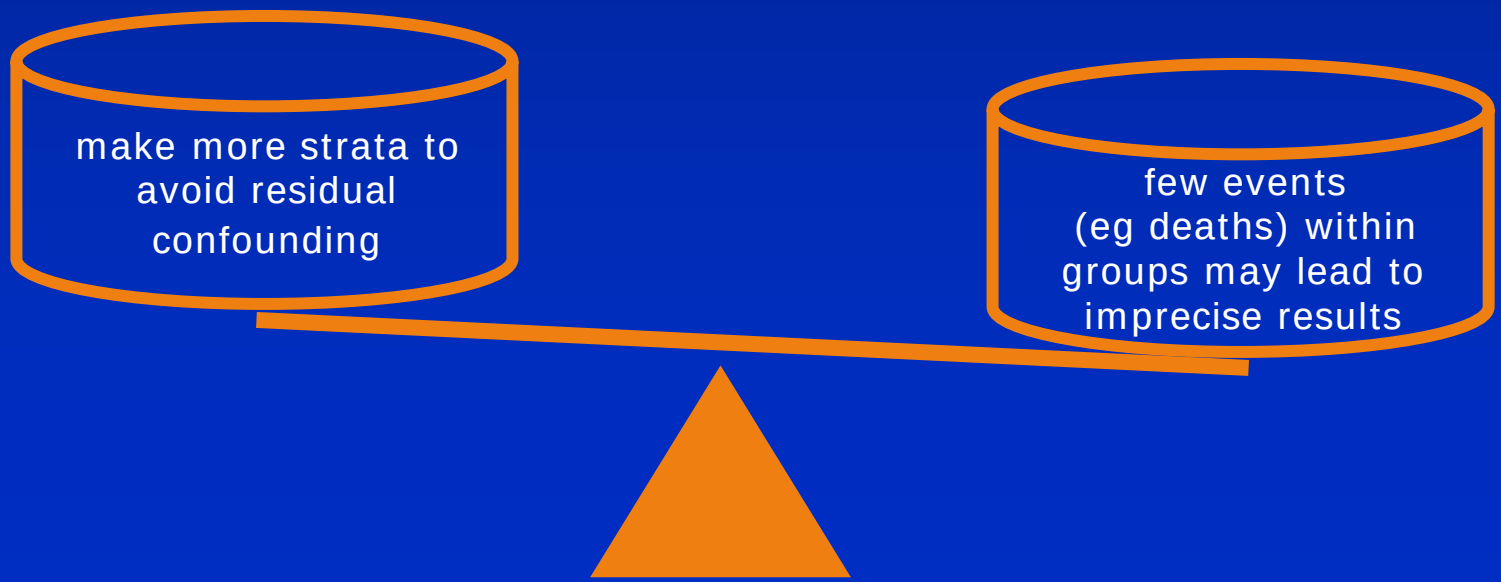
# Stratification and residual confounding

- A stratified analysis controls confounding only between strata

- When there are relatively few strata of continuous variables (e.g. age), there may be residual confounding within strata

…… the more strata we make the better will be

the adjustment for confounding …….



# Stratification and the number of variables

- Stratification is an effective means of adjusting for confounding

- Numbers of categories are multiplied;
  - by gender and by 5 age categories -› 10 strata;
  - 5 variables with 3 categories -> 243 strata

→ *When one wants to adjust for many confounding factors at the same time stratified analysis is <u>not</u> a practical method*

↓

**Multivariate regression is more practical**

# Regression analysis

## Regression

- Describes the relationship between two variables
- Predicts the value of one variable from a known value of the other(s)

**Exposure** → **Disease**

***Synonyms***

| | |
|---|---|
| Independent | Dependent |
| Predictor | Outcome |
| Explanatory | Response |

# Types of regression analysis

| | outcome/ = response/ = dependent variable | predictor/ = explanatory/ = independent variable |
|---|---|---|
| Linear regression | *continuous* | *continuous/categorical* |
| Logistic regression | *categorical* | *continuous/categorical* |
| Cox proportional hazards regression | *time to event* | *continuous/categorical* |

The type of regression analysis to use depends upon the type of outcome variable

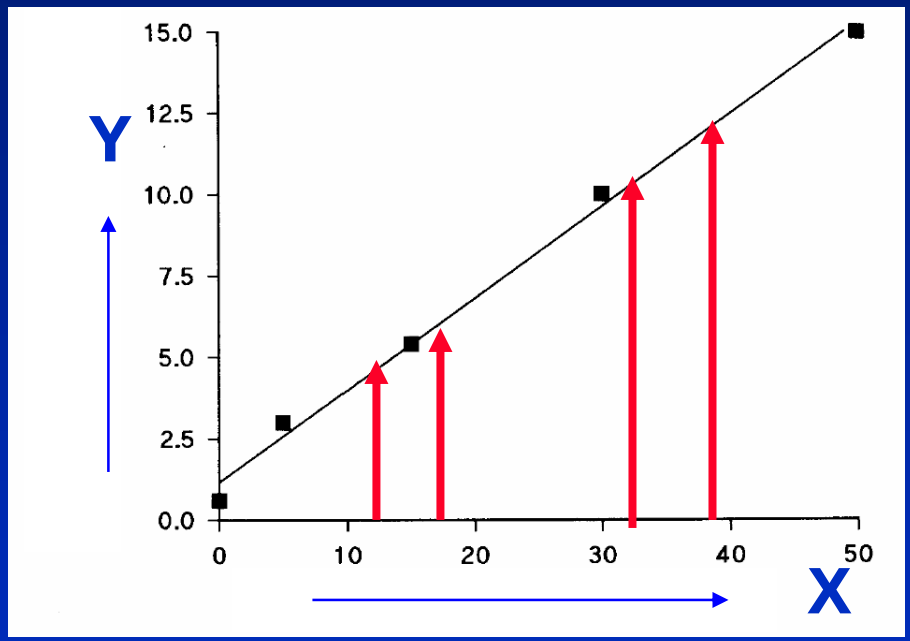
# Linear Regression

## Univariate (simple)

Mathematical equation $y = a + bx$

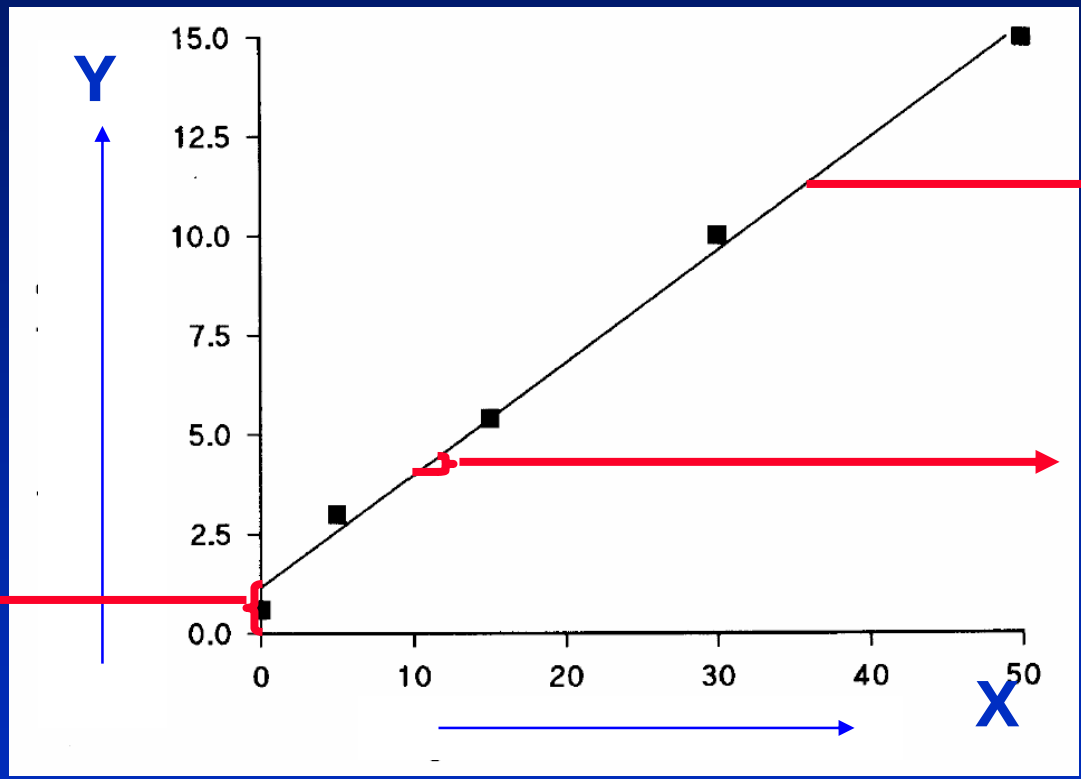
| DISEASE | EXPOSURE |
| --- | --- |
| Dependent | Independent |
| Outcome | Predictor |
| Response | Explanatory |

# Linear regression




- Determines the regression line that best describes the straight line relationship between x and y
- It estimates average values for y according to the different values of x

# Linear regression




Regression line

b = slope (regression coefficient)

a = intercept

mathematical equation $y = a + bx$

a = intercept  value of y when x = 0

b = slope  the mean increase in y for one unit increase in x

# Multivariate linear regression

**$y = a + bx$** — **Simple/univariate regression**

**$y = a + b_1X_1 + b_2X_2 + ..+ b_kx_k$** — **Multiple/multivariate regression**

**Provides estimates of effect that are mutually unconfounded = adjusted for each other**

For every x there is a regression coefficient b:

$b_1$, $b_2$ …… $b_k$ are partial regression coefficients

$b_1$ represents the amount by which y increases on average, if we increase $x_1$ by one unit, but KEEP ALL OTHER X'S CONSTANT (i.e. control or adjust for them)

$b_1$ represents the effect of $x_1$ on y that is INDEPENDENT OF ALL OTHER X's

# Commonly made errors - I

Nowadays over-adjustment is the most commonly made error!

- It takes away part of the real effect

*Example 2 - BMI and the risk for ESRD*[2]

| | Relative Risk of BMI 35.0-39.9 kg/m2 |
|---|---|
| adjusted model without blood pressure | 6.12 (CI, 4.97 to 7.54) |
| after additional 'adjustment' for blood pressure | 4.68 (CI, 3.79 to 5.79) |

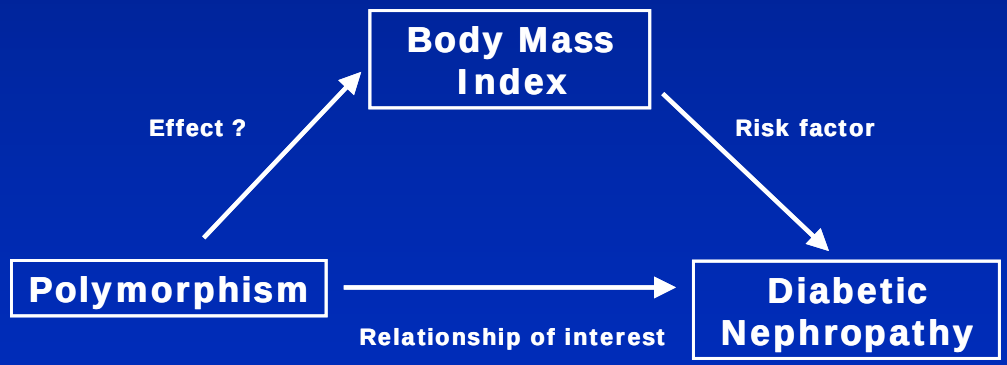
- In this example adjustment for blood pressure was incorrect from the perspective of confounding
- 'Adjustment' may however be useful to explore potential causal pathways…. *To see how much of the variance in ESRD is due to the effect of BMI via an increase of the blood pressure …*

# Commonly made errors - II

*Example 4 - Polymorphism (CNDP1 - Mannheim variant) and the susceptibility to diabetic nephropathy (DN)*

Janssen et al.[3] performed a case-control study using diabetic patients with DN as cases and diabetic patients without DN as controls. They showed that the CNDP1 - Mannheim variant was more common in the absence of DN (odds ratio 2.56 (CI, 1.36 to 4.84)).

***Would it have been useful if Janssen et al. would have matched for BMI?***




***Matching for body mass index would have been incorrect, as body mass index is <u>not</u> a potential confounder***

[4] *Janssen B, Hohenadel D, Brinkkoetter P et al. Carnosine as a protective factor in diabetic nephropathy. Association with a leucine repeat of the carnosinase gene CNDP1. Diabetes 2005;54:2320–2327.*

# Commonly made errors - III

Second most common error:

the use of statistical significance tests to detect confounding

- The amount of confounding is the result of the strength of the associations between the confounder on the one hand and the exposure and the disease on the other hand
- P-values will not provide information if a particular variable is a confounder
- The amount of confounding caused by a variable that satisfies all criteria for a potential confounder can be measured by looking at the difference between the crude and adjusted effect size:
  - If these are almost equal → there is no confounding
  - If the difference between is important → there is confounding

# Final remarks - I

- Confounding is a ‘mixing’ of effects distorting the real effect of an exposure. It is only an issue in studies investigating etiology
- Before adjusting for confounding all criteria for a possible confounder should be carefully checked in order to prevent the introduction of new bias through over-adjustment for variables that do not satisfy all criteria for confounding
- This requires sufficient knowledge on patho-physiology and being up-to-date with the nephrology literature
- Prevention of confounding during study design:
  - Randomization
  - Restriction
  - Matching

# Final remarks - II

- Adjustment for confounding during data analysis
  - –Stratification
  - –Multivariate analysis

- Most commonly made errors

  – ‘over-adjustment’ by adjusting for factors that do not satisfy the criteria for confounding, thereby taking away (at least a part of) the real effect

  – detection of confounding by the use of statistical significance tests instead of looking at the difference between crude and adjusted effect sizes

# Suggestions for further reading

Rothman KJ. Epidemiology: an introduction.
Oxford University Press 2002, New York

New series in Kidney International
- ABC of Epidemiology -