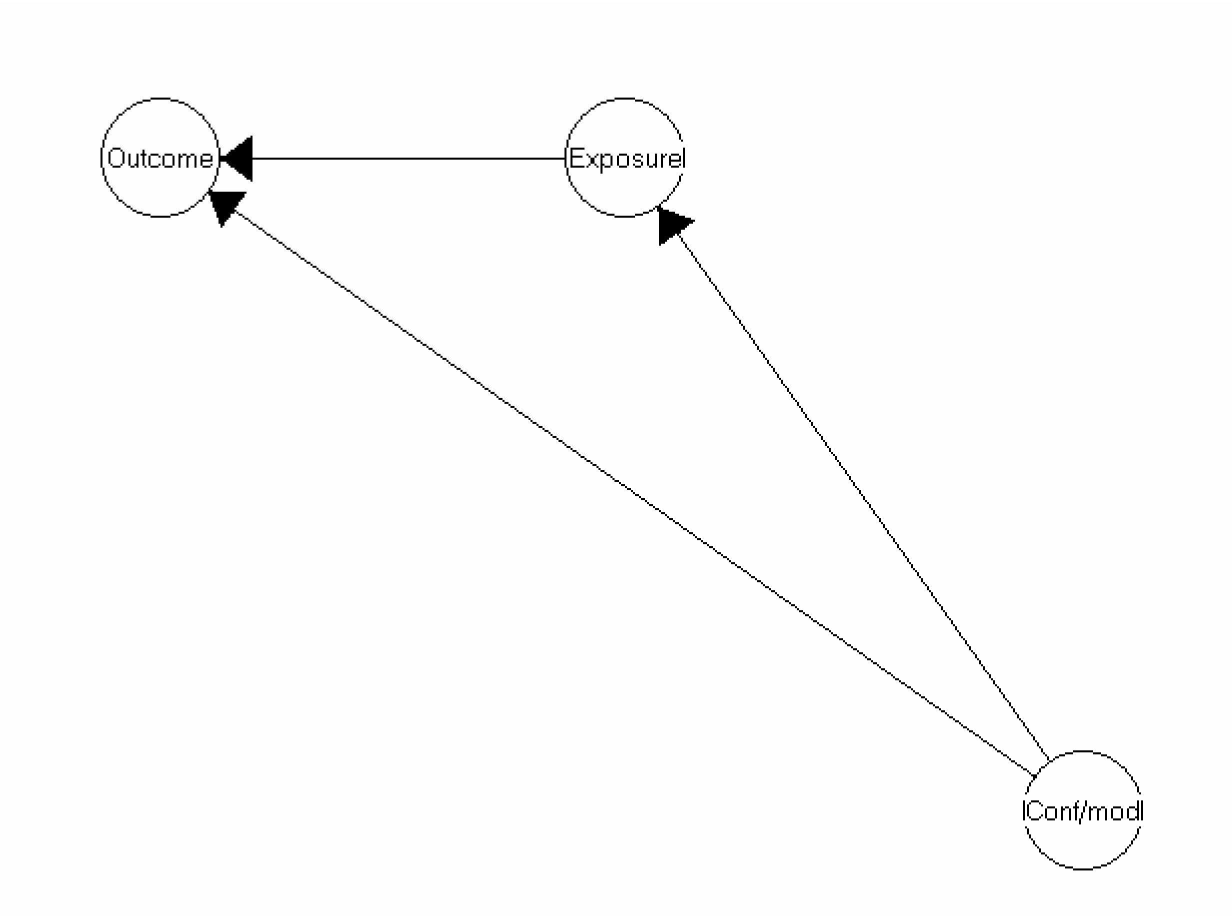
Statistical models for causality in observational studies

Svend Kreiner

Dept. of Biostatistics, Univ. of Copenhagen

The typical epidemiological situation



Arrows indicate causal relationships

The main problem: To estimate the causal effect of exposure on outcome taking confounding and effect modification into account.

Causality

What do we mean when we say that there is a causal relationship between two variables?

What is the exact meaning of causal effect?

How do we avoid or adjust for the confounding effect of other variables?

Hill's causality criteria (Rothman & Greenland, 1998):

- **Strength**
- **Consistency (repeated observations)**
- **Specificity**
- **Temporality (cause before effect)**
- **Biologic gradient**
- **Plausibility (subject matter arguments)**
- **Coherence (subject matter arguments)**
- **Experimental experience**
- **Analogy (subject matter arguments)**

Much epidemiological research is observational. Experimental evidence of causality therefore rarely exists.

Philosophy of science (Suppes)

Discusses causality relative to events rather than statistical variables.

Causality may be deterministic or probabilistic?

(A probabilistic causal effect of A an B means that there is a deterministic effect of A on P(B))

Primae facie probabilistic causes are

Spurious if cause and effect are conditionally independent given events occurring before the cause

Genuine – if not spurious

What about statisticians?

Randomized experiments are required if we need evidence supporting a causal statement.

Current theory about causal models in statistics insists that:

1) **A causal model has to be a directed acyclic graph (DAG).**

2) **Causal effects may be estimated from causal models of observational data.**

Can causal order be determined by analysis of data?

Freedman's (1997) law of conservation of rabbits:

If you want to pull a rabbit out of the hat,
you have to put a rabbit into the hat.

Remember,

Statisticians are making causal *models*, not causal theories.

Davis' (1985) rules for causal modeling:

Rule 1a: Run the arrow from X to Y if Y starts after X freezes.

Rule 1b: Run the arrow from X to Y if X is linked to an earlier step in a well-known sequence.

Rule 1c: Run the arrow from X to Y if X never changes and Y sometimes changes.

Rule 1d: Run the arrow from X to Y if X is relatively stable, hard to change, or fertile while Y is relatively volatile, easy to change, or has few consequences.

Rule 2: If there is a path starting from X and returning to it without retracing any steps, X and all the variables on the path form a loop. Variables in a loop have no order.

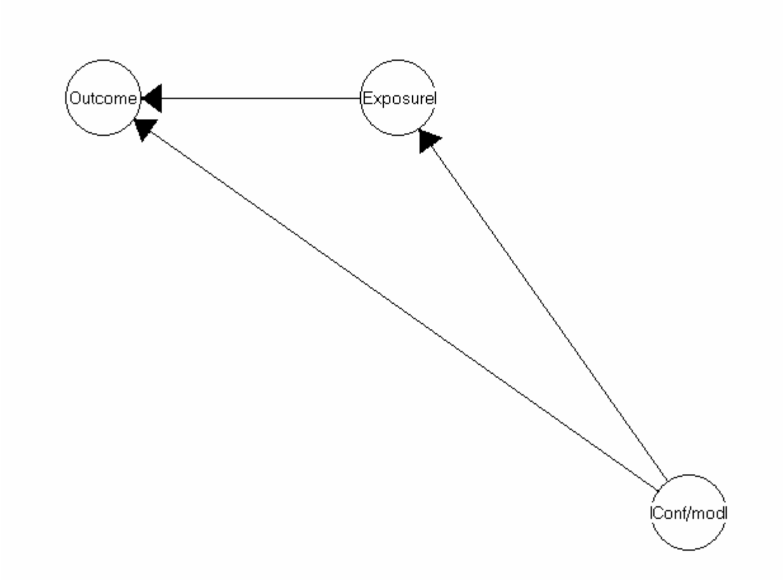
Rule 3: Confounding. If a prior variable has a causal path to the independent variable and a causal path to the dependent variable, it will contribute a statistical association between them that is causally spurious

Rule 4: Reversing poles for one variable reverses the signs of each of its relationships. Reversing polarities for both variables leaves the sign of their relationship unchanged.

Rule 5: The sign of a path is given by multiplying the sign of its arrows. A path of nonzero arrows will be positive unless it contains an odd number of negative arrows.

Rule 6: A System is inconsistent if at least one pair of variables has both positive and negative signs among its direct, indirect and spurious effects. Otherwise it is consistent.

Causal effects



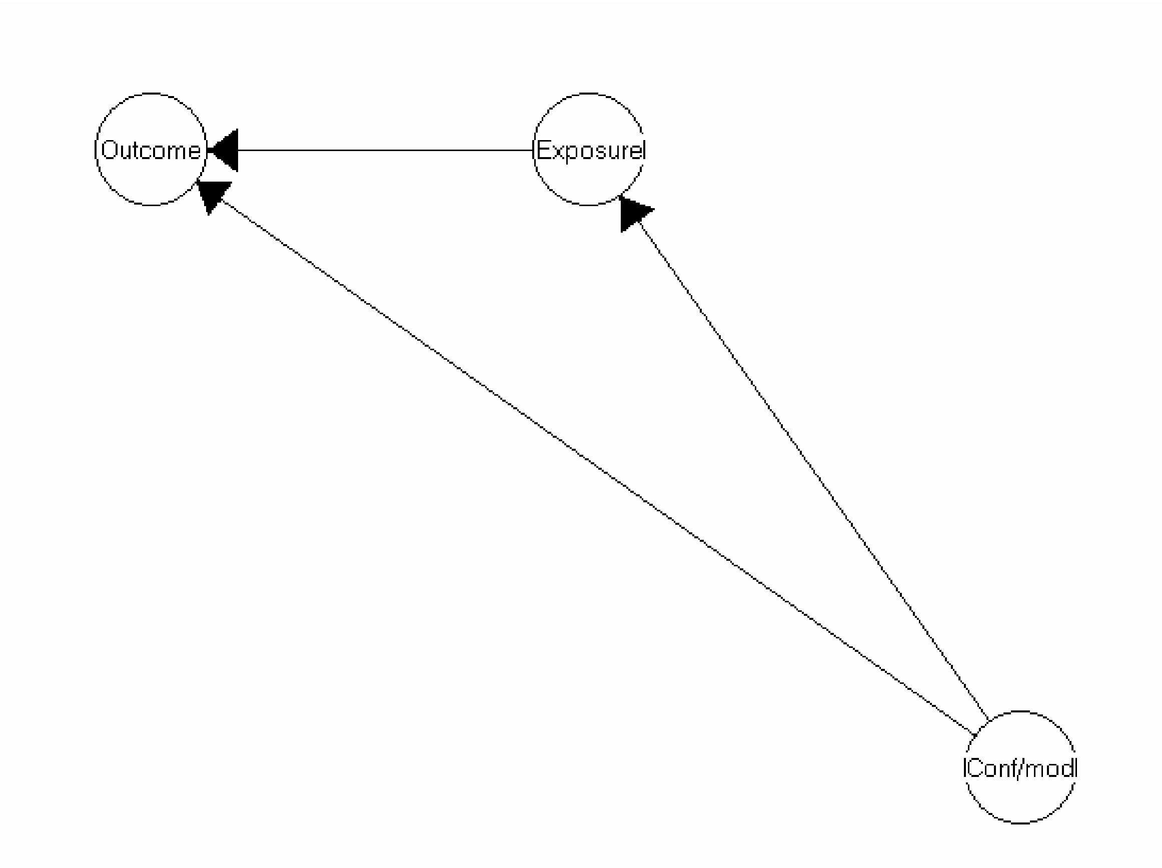
How do we measure the causal effect of exposure on outcome?

Three types of causal effects

Total effects.

Direct effects

Indirect effects



The total effect of the confounder/mediator

=

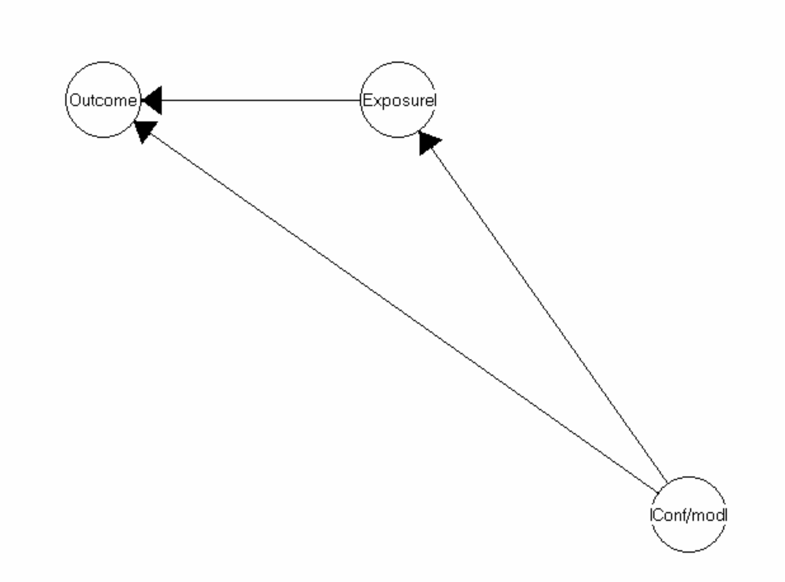
Indirect effect mediated through Exposure

“+”

Direct effect

Common wisdom of epidemiologists tells us to disregard mediating variables

Causal effects



How do we measure the causal effect of exposure on outcome?

Two types of causal effects

Individual local effects.

Average causal effects (ACE)

Local (individual) effects

are defined by the conditional distribution of outcome (Y) given exposure (X) and all other risk factors:

$$P(Y|X, Z_1, \ldots , Z_k)$$

The local effect of is a measure of the "distance" between two distributions

$$\beta_{loc} = \text{dist}(P(Y|X=1, Z), P(Y|X=0, Z))$$

The measure of effect may be confounded if the list of risk factors is incomplete, whether or not X is associated to the missing Zs

If the local effect differs across different values of Z then we say that Z modifies the causal effect.

The average causal effect

If X is independent of all other risk factors then ACE is as a measure of the "distance" between the conditional distributions of Y given different values of X

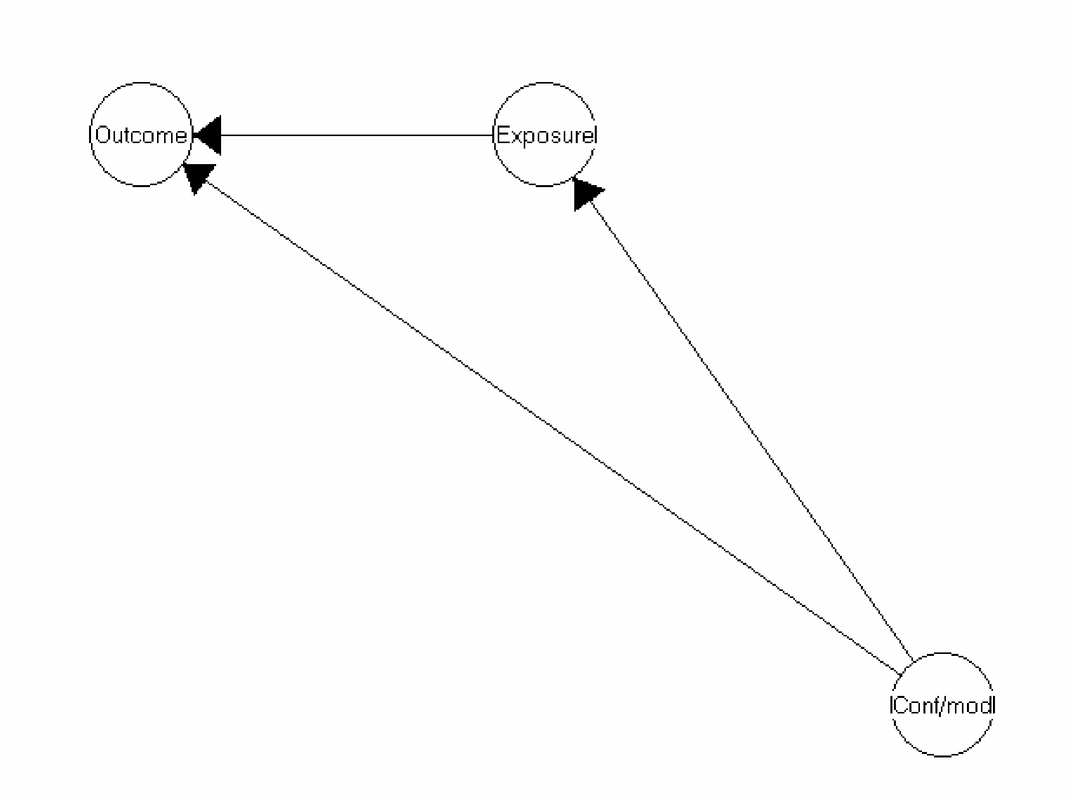
$$\beta_{ACE} = \mathrm{dist}(P(Y|X=1), P(Y|X=0))$$

ACE defined in this way is confounded if some risk factors are associated with X.

Calculation of ACE therefore in principle requires randomized experiments assuming that the local effects are the same

In practice, ACE may be estimated from observational studies if local effects are unconfounded.

Causal statistical models



What kind of statistical model should we use to test and estimate the causal effect?

Local individual effects:

Dichotomous outcome: Logistic regression

Continuous outcome: Linear regression/ANOVA

There are, however, complications

Complications 1

The nature of the outcome variable:

- Dichotomous
    - Frequent/infrequent events
- Waiting times
    - Censored/not censored
- Counts

- Ordinal categorical variables
- Summated scales
- Quantitative measures

The model structure

- Multiple outcomes
- Multiple exposures
- Multiple confounders/modifacators
- Intermediate variables

Complications 2

Design problems

Typical epidemiological studies are observational

Longitudinal studies are often not practical. Instead we use

Retrospective cross sectional surveys

Case control studies (consisting of retrospective surveys of Cases and controls)

Panel studies (repeated measurements)

Measurement problems

Many risk factors are measured with errors

Multivariate causal statistical models

Two options:

Structural equation models

Graphical models