Some interpretational issues connected with observational studies

D.R. Cox

Nuffield College, Oxford, UK

and

Nanny Wermuth

Chalmers/Gotherburg University, Sweden

Types of study

- secondary analysis
- cross-sectional observational study
- retrospective obervational study
- prospective observational study
- (randomized) experiment

Mixtures

Key ideas

- several features (variables) on each study individual
- represent each feature by node of graph
- for any two features
  - one response to other as explanatory, or
  - on an equal footing

- in graph if variables connected
  - joined by directed edge, or
  - joined by undirected edge
- variables on equal footing in same box
- absence of edge implies conditional independence, subject to rules specifying nature of conditioning set

Objective

To develop understanding of potential data-generating process

Example

Y,
Glucose control (GHb)

X,
Knowledge about illness

Types of attribution:

Z, Fatalistic externality

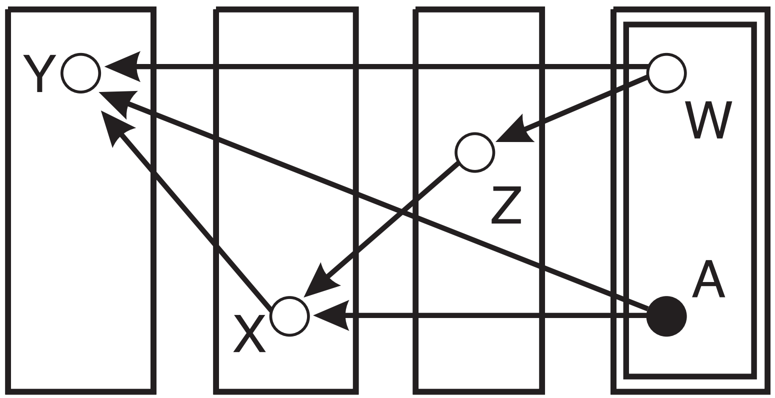
U, Social externality

V, Internality

W,
Duration of illness

A,
Duration of schooling

B,
Gender



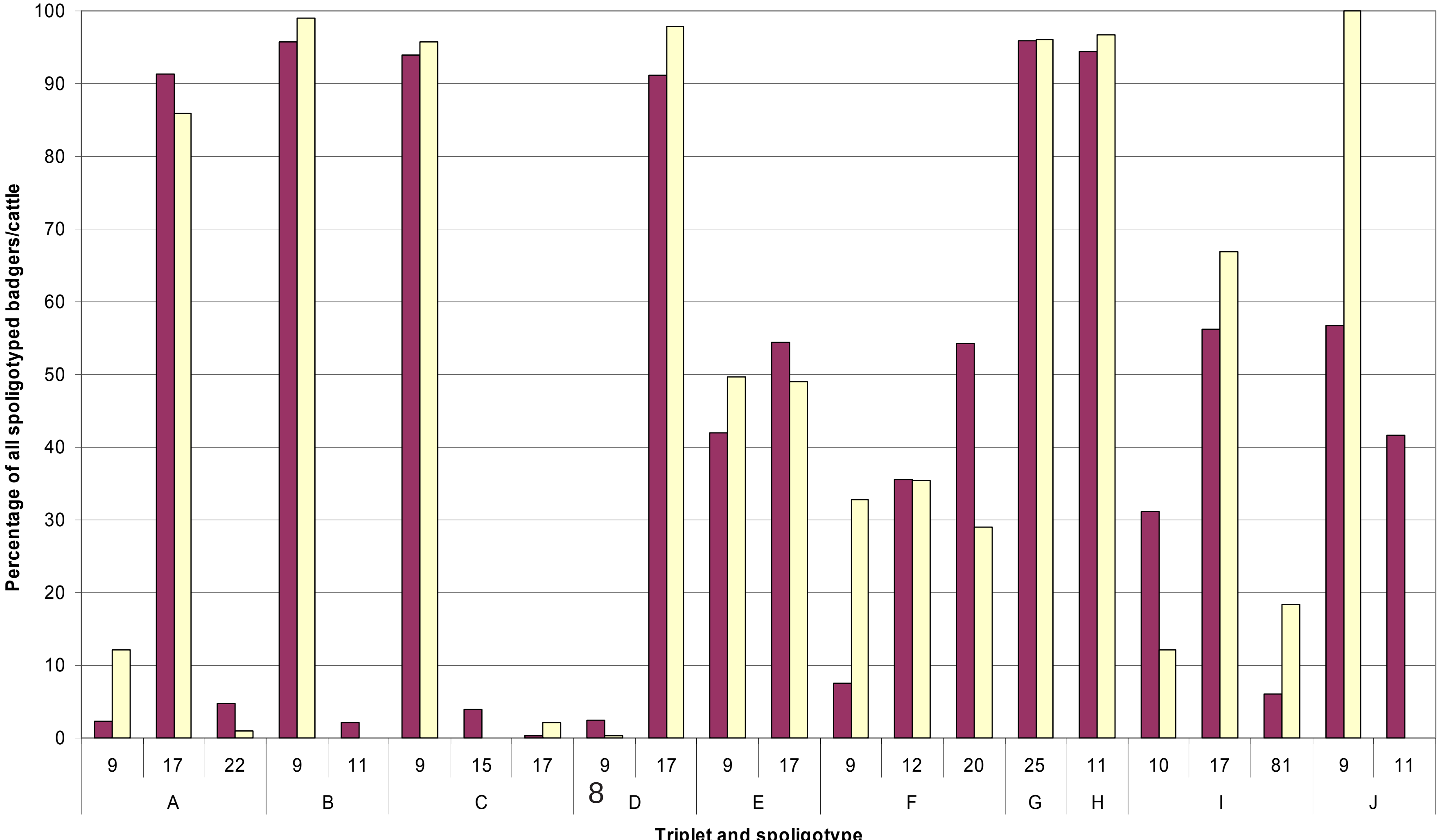
Another example

Infectious disease in two species

Causative organism can be genetically typed

Cross-sectional data

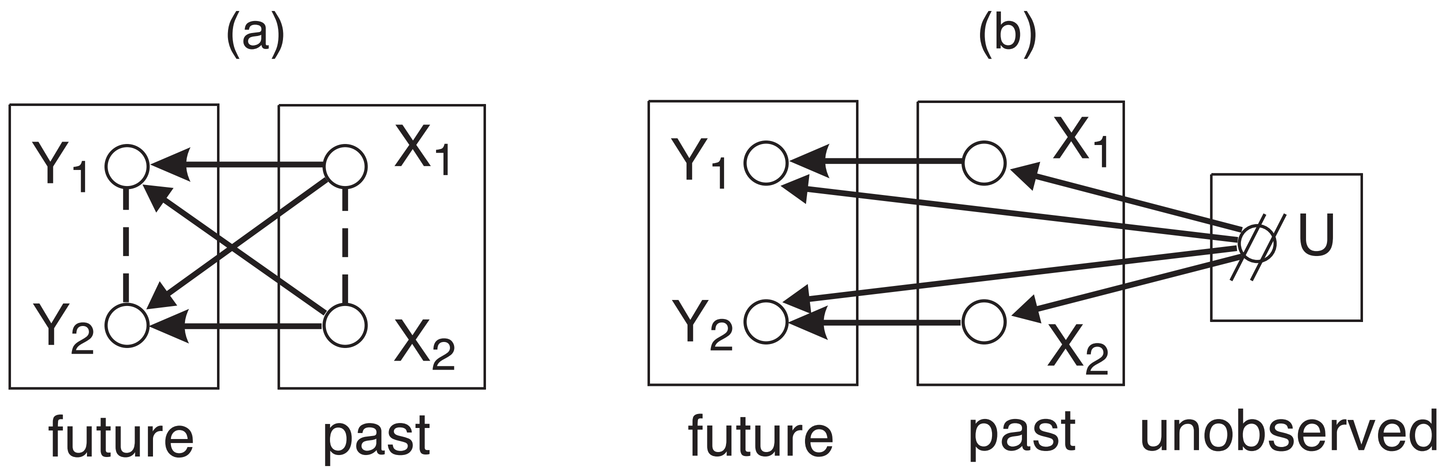
Percentage of culture positive animals by triplet and spoligotype from trial data.



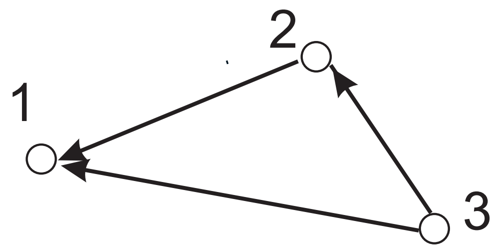
Interpretation

Difficulties of interpretation

How can interpretation be extended?

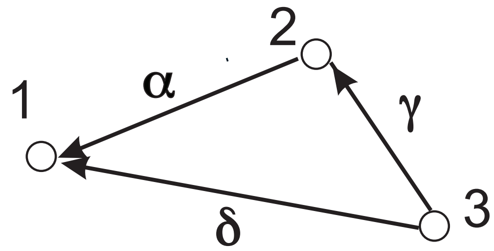

(a)
Y1
X1
Y2
X2
future
past
(b)
Y1
X1
U
Y2
X2
future
past
unobserved

A simple stepwise data generating process



for a joint density $\boldsymbol{f_{123}}$

$$\boldsymbol{f_{123} = f_{1|23} f_{2|3} f_{3}}$$



for a linear system in standardized variables

$$
\begin{aligned}
\mathbf{E}(\mathbf{Y_1}|\mathbf{Y_2},\mathbf{Y_3}) &= \alpha\mathbf{Y_1} + \delta\mathbf{Y_2} \\
\mathbf{E}(\mathbf{Y_2}|\mathbf{Y_3}) &= \gamma\mathbf{Y_3} \\
\mathbf{E}(\mathbf{Y_3}) &= \mathbf{0}
\end{aligned}
$$

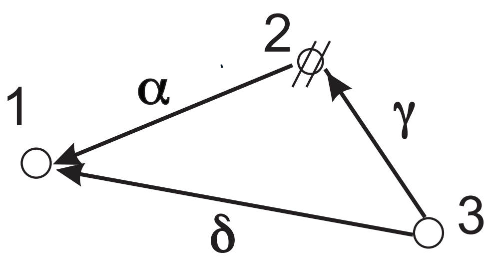
Distortions of effects

due to

 marginalizing over a variable

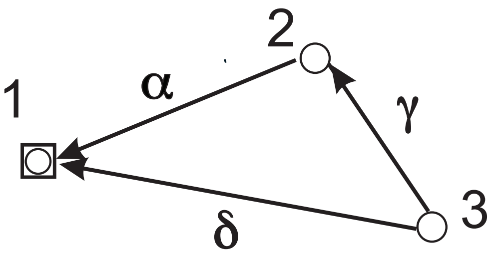
 conditioning on a variable

Distortion due to under-conditioning



$$\mathrm{E}(\mathbf{Y_1}|\mathbf{Y_3}) = (\delta + \alpha\gamma)\mathbf{Y_3}$$

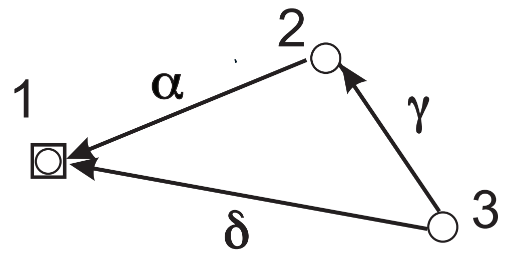
Distortion due to over-conditioning



With the simple correlation $\rho_{13} = \delta + \alpha\gamma$

$$\mathbf{E}(\mathbf{Y_2}|\mathbf{Y_3}, \mathbf{Y_1}) = (\gamma - \{(1-\gamma^2)/(1-\rho_{13}^2)\}\alpha\rho_{13})\mathbf{Y_3} + \ldots$$

Distortion due to over-conditioning



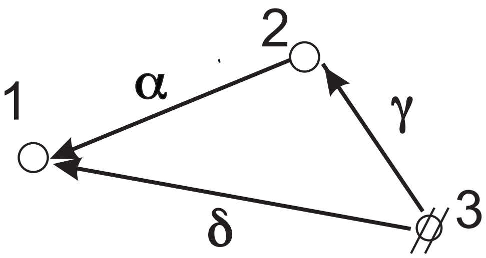
With the simple correlation $\rho_{13} = \delta + \alpha\gamma$

$$\mathrm{E}(\mathbf{Y_2}|\mathbf{Y_3}, \mathbf{Y_1}) = (\gamma - \{(1 - \gamma^2)/(1 - \rho_{13}^2)\}\alpha\rho_{13})\mathbf{Y_3} + \ldots$$

induced partial dependence in the case $\gamma = 0$

$$\mathrm{E}(\mathbf{Y_2}|\mathbf{Y_3}, \mathbf{Y_1}) = (-\alpha\delta/\{1 - \delta^2\})\mathbf{Y_3} + \ldots$$

Distortion due to direct confounding



$$\mathrm{E}(\mathbf{Y_1}|\mathbf{Y_2}) = (\alpha + \delta\gamma)\mathbf{Y_2}$$

To avoid both over- and under-conditioning

regress $\mathbf{Y}_i$ only all on all those observed variables

which are in the generating process

directly or indirectly explanatory for $\mathbf{Y}_i$

To avoid direct confounding

of $\boldsymbol{i} \prec\!\!- \boldsymbol{j}$

randomly allocate individuals to the levels of $\mathbf{Y}_{\boldsymbol{j}}$

To check on direct confounding

induce an ij-dashed line into the generating graph

– whenever i and j have an unobserved common parent

$$i \prec\!\!- \not\!\not\circ -\!\!\succ j$$

–or i and j have an unobserved common ancestor path

$$i \prec\!\!- \not\!\not\circ \prec\!\!- \ldots \prec\!\!- \not\!\not\circ -\!\!\succ ,\,,\,, \not\!\not\circ -\!\!\succ j$$

Conclude:

no direct confounding of $i \prec\!\!- j$ if no $i - - - j$ is induced

But, distortions due to indirect confounding

may be present and severe

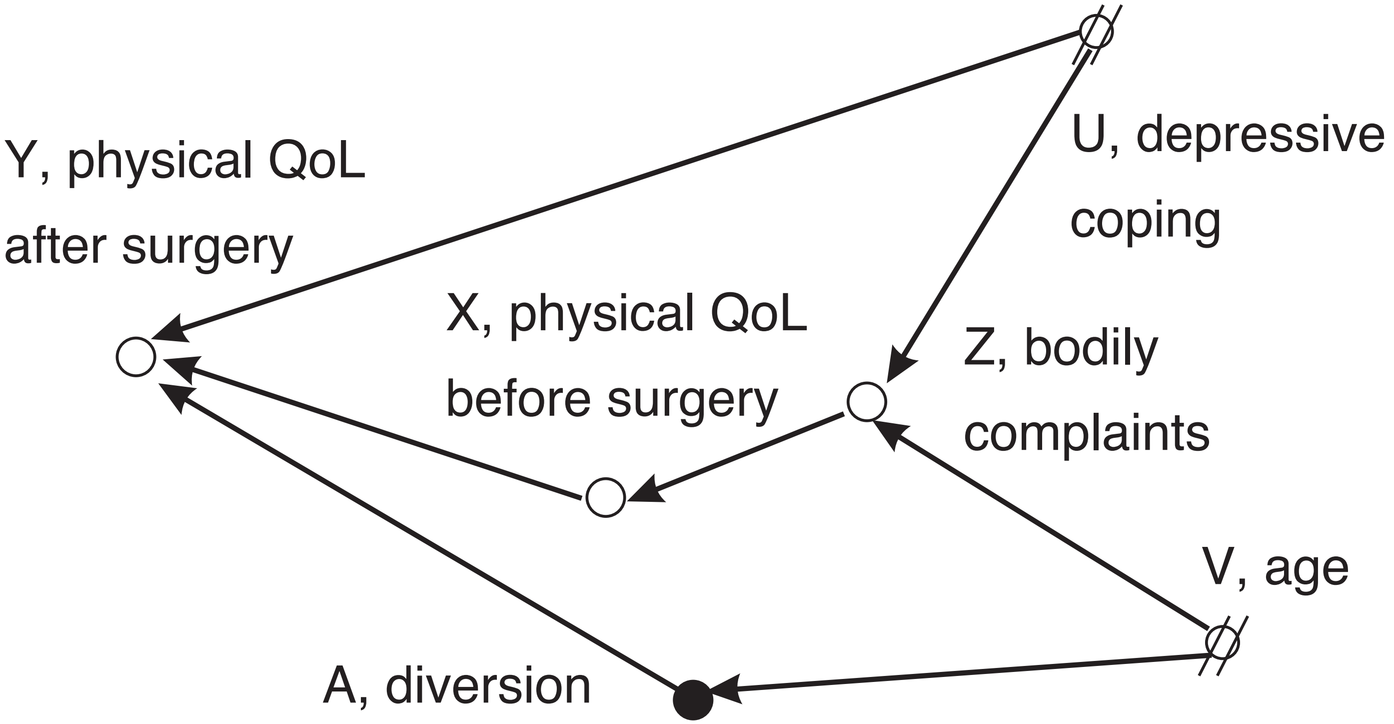
– **both in observational and in randomized studies**

– **and in the absence**

of direct confounding, of over- and of under-conditioning

Example

Even direction of the dependence of Y on A could be reversed in



via Y ←— U —→ Z ←— V —→ A being Y ←— ⦸ —→ ⧇ ←— ⦸ —→ A

References

Cox & Wermuth (1996). *Multivariate Dependencies – Models, Analysis and Interpretation.* London: Chapman & Hall

Cox & Wermuth (2004). Causality: a statistical view. *International Statistical Review*, **72**, 285-305

Lindley (2002). Seeing and doing: the concept of causation. *Int. Statist. Rev.* **70**, 191-214

Wermuth, Wiedenbeck & Cox (2006). Partial inversion for linear systems and partial closure of independence graphs. *BIT, Numerical Mathematics*, **46**, 883-901

Wermuth & Cox (2007). Distortion of effects. Submitted