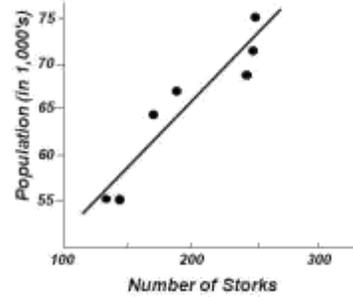


Using regression in combination with expert knowledge more correct and gain mechanisms

analysis with expert to get estimates, insight into mechanisms

Kathrine Frey Frøslie, statistician, PhD
Norwegian advisory unit for women's health, Oslo university hospital

Population of Oldenburg, Germany, at Year's End vs. Number of Storks Observed Each Year (1930 - 1936)



Source: Statistics for Experimenters, by Box, Hunter & Hunter

Understanding

Mechanisms

Expert knowledge

What is the best estimate for the association between the main exposure and the main outcome?

Etiology

Causality

Causality

What is causality?

The Counterfactual concept

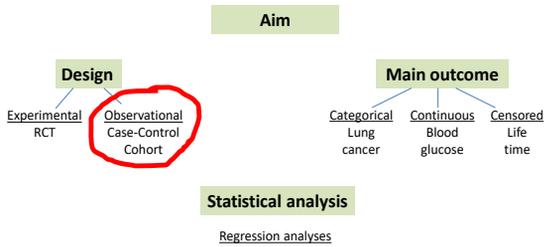
Philosophic background

Interventions and causality
Consequences of actions

Ultimate goal:
Action!



Problem of interest



Based on expert knowledge of the topic under investigation, we want to estimate the association between an exposure and an outcome as unbiasedly as possible.

Knowledge of the topic makes it plausible that the estimated association can be interpreted as an effect, i. e. causal, i.e. as a quantification of mechanisms.

The expert knowledge about the topic is formalised in a graph of the variables studied, a Directed Acyclic Graph (DAG).

In a DAG, one variable is defined as the main outcome, and one variable is defined as the main exposure.

Expert knowledge is used to define other variables as either a confounder, a mediator, or a collider.

DIRECTED ACYCLIC GRAPH (DAG)

See CAUSAL DIAGRAM.

CAUSAL DIAGRAM

(Syn: causal graph, path diagram) A graphical display of causal relations among variables, in which each variable is assigned a fixed location on the graph (called a node) and in which each direct causal effect of one variable on another is represented by an arrow with its tail at the cause and its head at the effect. Direct noncausal associations are usually represented by lines without arrowheads.

Graphs with only directed arrows (in which all direct associations are causal) are called directed graphs.

Graphs in which no variable can affect itself (no feedback loop) are called acyclic.

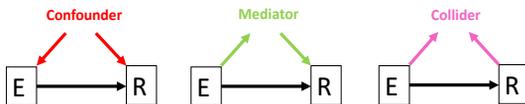
Algorithms have been developed to determine from causal diagrams which sets of variables are sufficient to control for confounding, and for when control of variables leads to bias.

Porta M: A Dictionary of Epidemiology, Fifth Edition, 2008

Exposure and response



Confounder, mediator, collider



Note: The presence of either of these may affect the association of interest. Hence, including either of these in regression analyses may change effect estimates.

Confounding

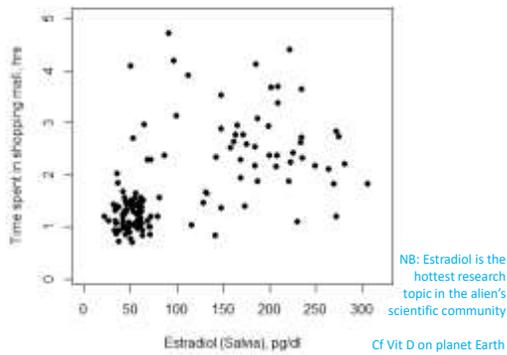
Confounding is bias of the estimated effect of an exposure on an outcome due to the presence of a common cause of the exposure and the outcome.

Confounding can be reduced by proper adjustment. Exploring data is not sufficient to identify whether a variable is a confounder, and such evaluation of confounding may lead to bias. Other evidence like pathophysiological and clinical knowledge and external data is needed. Directed Acyclic Graphs (DAGs) are useful tools when considering confounding variables.

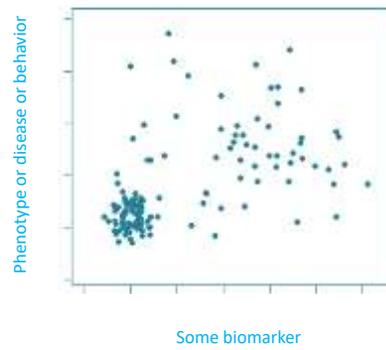
See e.g. Vandenbroucke et al: Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Explanation and Elaboration. Epidemiology; 2007:18:6



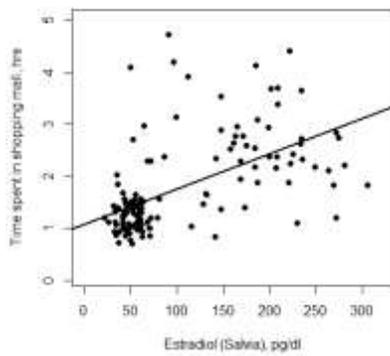
Ex: Shopping time vs estradiol level. Simulated data.



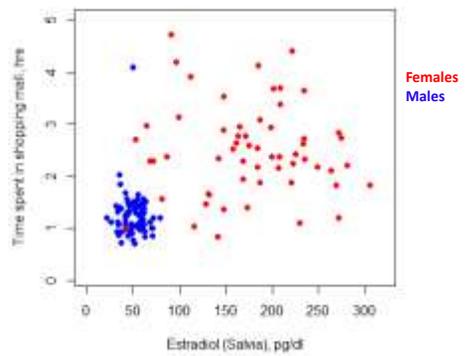
How common is the alien's approach in real research?



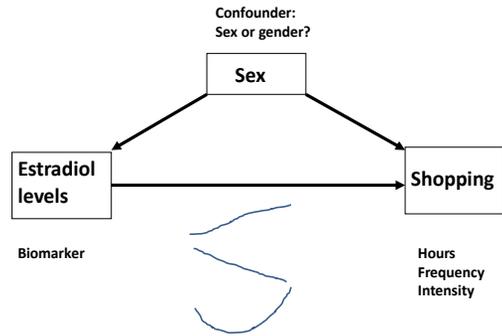
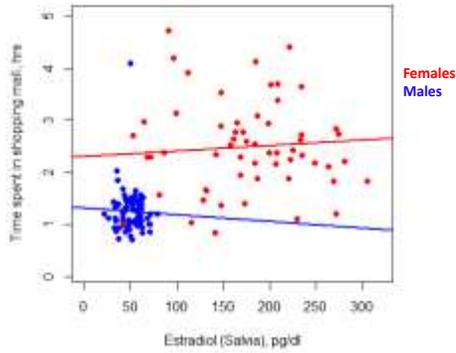
Ex: Shopping time vs estradiol level. Simulated data.



Ex: Shopping time vs estradiol level. Simulated data.



Ex: Shopping time vs estradiol level. Simulated data.

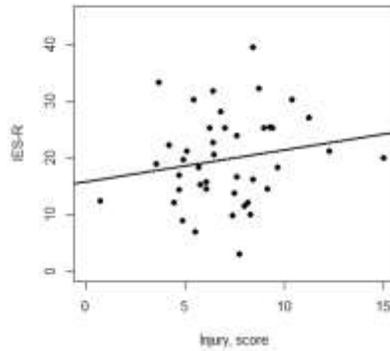


COLLIDER

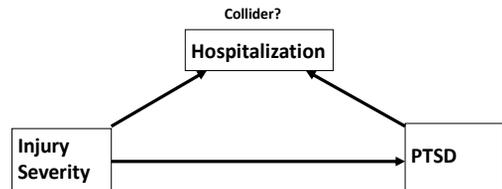
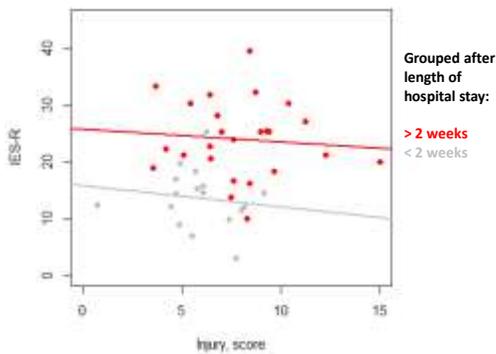
A variable directly affected by two or more other variables in the causal diagram.

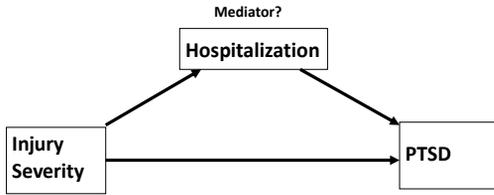
Porta M: A Dictionary of Epidemiology, Fifth Edition, 2008

Ex: Post-traumatic stress after terror attack. Simulated data.

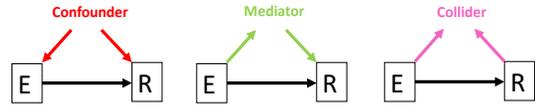


Ex: Post-traumatic stress after terror attack. Simulated data.

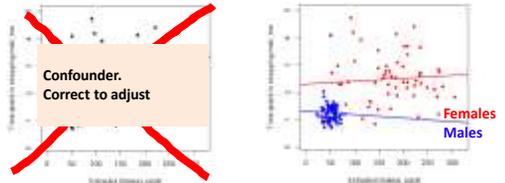
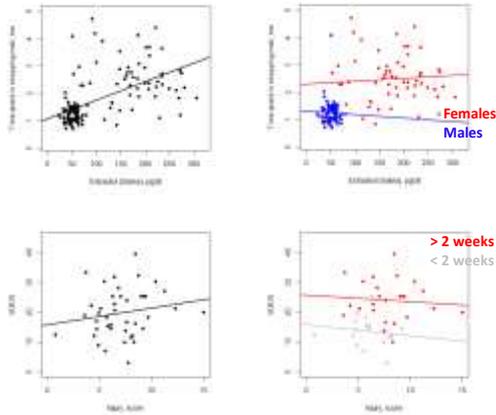




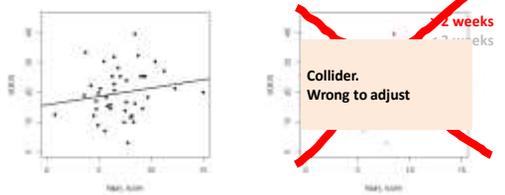
Confounder, mediator, collider



Note: The presence of either of these may affect the association of interest. Hence, including either of these in regression analyses may change effect estimates.



Only expert knowledge can tell us what to do



DAG, Confounder, Mediator, Collider

Hernán MA, Hernández-Díaz S, Werler MM, Mitchell AA. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. Am J Epidemiol. 2002;155:176-84.

Janszky I, Ahlbom A, Svensson AC: The Janus face of statistical adjustment: confounder versus collider. Eur J Epidemiol 2010,25:361-363