Causality through the stochastic system approach

Daniel Commenges

INSERM, Centre de Recherche Epidémiologie et Biostatistique, Equipe Biostatistique, Bordeaux
http://sites.google.com/site/danielcommenges/

June 6, 2011

Organization of the talk

1. An epistemological theory of causality in epidemiology;
   1.1 Levels
   1.2 Systems and physical laws
2. A theory of influence between components of a stochastic process;
3. Examples: Good modeling and the WHI study.

An epistemological theory of causality in epidemiology

- We shall use the concept of cause with an extended meaning: not deterministic and plurality of causes
- We shall not use the counterfactual point of view
- We shall discuss the concept of intervention: intervention plays a central role but does not enter in the definition of causality

I will begin with the story of the volcano.

Eruption of Vesuvius

Death of the inhabitants of Pompeii

The story of the volcano

Causal link

In 79 AC there was an eruption of Vesuvius and shortly after, it happened that all the inhabitants of Pompeii died. Most people would agree that there is a causal link between the two events.

Although no intervention is possible

- It is impossible to prevent the eruption of a volcano
- Even if an intervention was possible it has not been implemented in Pompeii in 79

The story of the volcano

Causal links through physical laws.

- Vulcanology tells us that an eruption can produce a great quantity of ashes, lava and a heat wave
- We know from biology that living beings cannot survive at a temperature higher than 200° C.

The story of the volcano

Counterfactual point of view

- "There was no eruption of Vesuvius in 79 AC" is counterfactual
- Under this counterfactual assumption the inhabitants of Pompeii do not die in 79
- Thus there is a causal link between eruption in 79 and death of the inhabitants in 79

For me the counterfactual point of view, although stimulating for the imagination, does not add anything to the physical law point of view.

The riddle of the rooster

Post-hoc fallacy and the scientific approach

- Rooster crow is repeatedly followed by sun rise.
- Does the rooster crow cause the sun to rise ?
- Causal effect can be dismissed on general scientific ground.

The stochastic system approach to causality

Systems and physical laws

- Physical laws applied to a system give the correct law of the components of the system.
- If the evolution of a component k between t and $t + \Delta t$ depends on the state of component j at time t then component j has a direct effect on component k:

$$X_j \longrightarrow_X X_k$$

- If the evolution of a component k after t depends on the state of component j at time t then component j has a direct or indirect effect on component k.

$$X_j \rightarrow\rightarrow_X X_k$$

Problems and questions

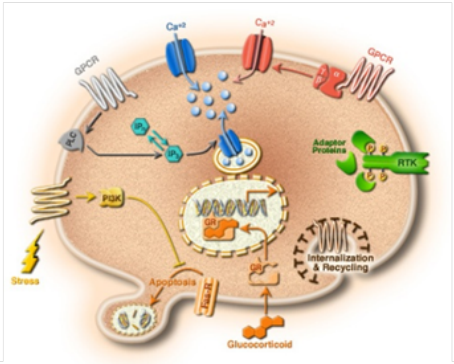
- What is a physical law ?
- What is a system ? What is a good system ?
- Does physical laws apply to biology, epidemiology, sociology ?
- How does this point of view helps for distinguishing causal links from spurious associations ?
- What can we do when we don't know the physical laws?
- What can we do when we don't know what is the good system ?
- Or neither physical laws not the good system ?
- How can we formalize influence between components ?
- How can we compute the effects ?

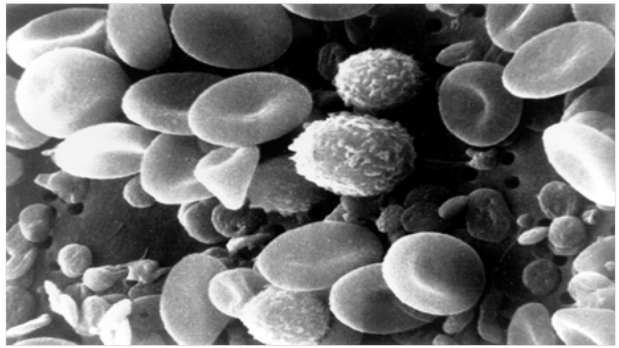
Levels

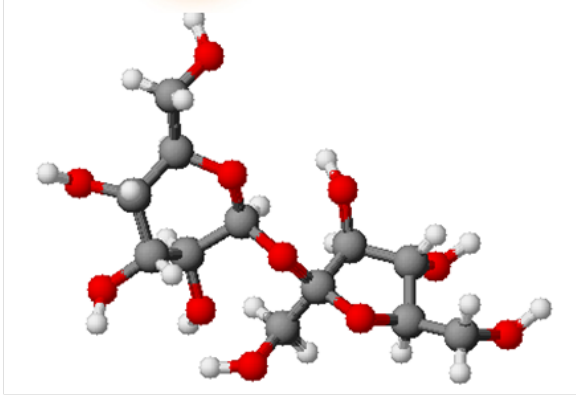
Systems and physical laws are defined for a given level. Levels are specified in essentially two dimensions:

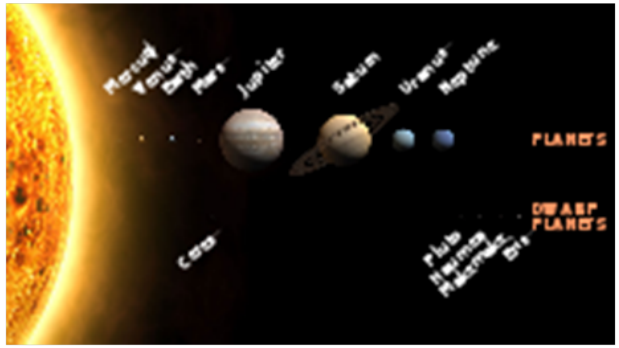
- Complexity: quarks, atoms, molecules, cells, organisms, societies
- scale: space-time size, number of entities

Complexity









Scale

Physical laws and Systems

Physical laws

Rules that when applied to good ("'perfect"') systems allow to compute the law of the observed events.
Intuition: the events are "*produced*" according to these laws.

Example:

Movement of the Earth conditioned on the position of the Sun, Mars and Venus. We must know both the physical laws (Newton laws) and the system to which to apply them.

Distinguishing features: Wide applicability and reductionism

- Newton's laws can be applied to planets in the solar system and in other stellar systems, to satellites of planets; the law of mass attraction can be applied to atoms...See Pearl (2000), 1.3.2 about "stability".
- Basic laws and laws at upper levels: laws at upper levels ca be explained by laws at lower level (Boltzman).

Levels, Physical laws and Systems

- Some laws are relevant at different levels: gravitation law (however negligible at particle level, main force at the level of celestial bodies)
- Generally different laws apply to different levels.
- Reductionism: Laws at one level can be grounded on laws of lower level.
  Example: law of ideal gases which can be explained by the (stochastic) kinetic laws of the molecules (Boltzman).
- Principle of emergence: Laws at one level are not mere application of the laws at a lower level.
- Generally at each level there is one or several particular sciences; a science can operate at different levels but there is often a privileged level.

Levels of epidemiology

Epidemiology is mainly in between:

- individual level: risk factors of diseases
- population (human beings) level: impact of a disease (prevalence, incidence), infectious diseases

At the margin of epidemiology and system biology one is interested in two levels:

- population of cells: interaction between virus and immune system
- individual level: occurrence of diseases in relation to number of CD4+ T cells

The scientific approach and epidemiology

Causal effects in epidemiology

- antiretroviral treatment decreases the risk of AIDS;
- smoking increases the risk of lung cancer;
- excessive calory intake leads to obesity;
- obesity is a risk factor of ischemic heart disease.

Causal effects rely on laws at lower level

- the antiretroviral may prevent new virions to be constructed after transcription of the integrated ADN of the virus; this in turn can be explained by biochemistry properties of the drug.
- ...

Example of a simple system

Learning effects when the system is known

- Process of interest: D (a disease).
- We know which entities or states have a direct effect on D.
- Suppose that we find two entities G and F. Can we learn the physical law quantitatively, that is learn the effect of F on D by observation ?
- "'yes"': In any system containing both F and G the observed effect will be the causal effect.
- We learn by observing what happens to different systems, that is the subjects included in the sample of the study.

Example of a simple system

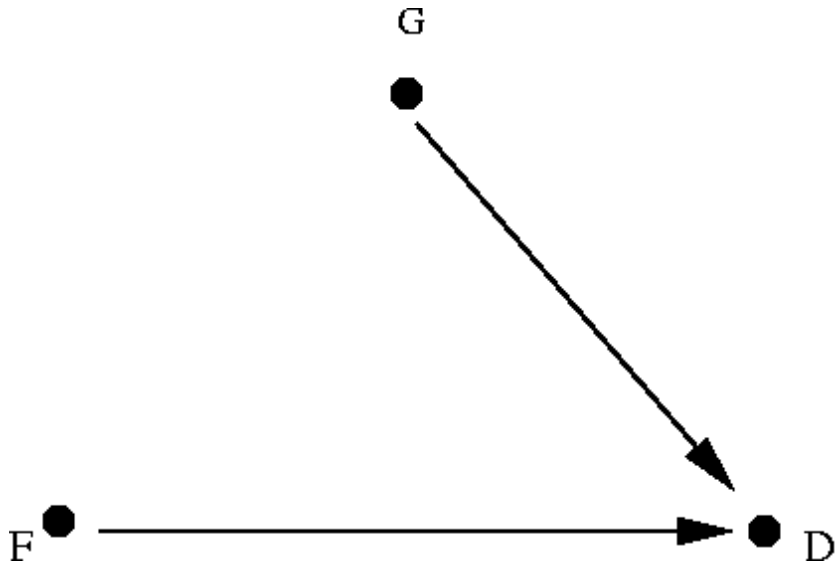


Figure: Causal parents of *D*

The problem of epidemiology

Both system and laws unknown

- Which factors modify the risk of a disease ? Learning the system;
- What is the dynamics of this development ? Learning the laws.

How to learn ?

- Advantage: we can have many replica of systems (subjects)
- Drawback: we have generally few longitudinal data. The situation is improving. Typically case-control studies do not have longitudinal observations, but cohorts have.
- Most models are inadequate in that they do not grasp the dynamics of the system. Dynamical models: Cox model, and multivariate: multistate models or joint models.

The problem of epidemiology

Both system and laws unknown

- Which factors modify the risk of a disease ? Learning the system;
- What is the dynamics of this development ? Learning the laws.

How to learn ?

- Advantage: we can have many replica of systems (subjects)
- Drawback: we have generally few longitudinal data. The situation is improving. Typically case-control studies do not have longitudinal observations, but cohorts have.
- Most models are inadequate in that they do not grasp the dynamics of the system. Dynamical models: Cox model, and multivariate: multistate models or joint models.

A theory of influence between components of a stochastic process

Direct influence

Doob-Meyer decomposition

Let $\mathcal{F}_t = \mathcal{F}_0 \vee \mathcal{X}_t$; $\mathcal{F}_0$ contains the initial value and possibly other information. We shall consider the class of special semi-martingales, that is the class of processes which admit a unique Doob-Meyer decomposition in the $(\mathcal{F}_t)$ filtration, under probability P:

$$\boldsymbol{X}_t = \Lambda_t + M_t, t \geq 0, \tag{1}$$

where M_t is a martingale and Λ_t is a previsible process with bounded variation.

We have defined a class $\mathcal{D}'$ including in particular counting processes and diffusion processes.

Direct influence

Doob-Meyer decomposition

Let $\mathcal{F}_t = \mathcal{F}_0 \vee \mathcal{X}_t$; $\mathcal{F}_0$ contains the initial value and possibly other information. We shall consider the class of special semi-martingales, that is the class of processes which admit a unique Doob-Meyer decomposition in the $(\mathcal{F}_t)$ filtration, under probability P:

$$\boldsymbol{X}_t = \Lambda_t + M_t, t \geq 0, \tag{1}$$

where M_t is a martingale and Λ_t is a previsible process with bounded variation.

We have defined a class $\mathcal{D}'$ including in particular counting processes and diffusion processes.

Weak conditional local independence (WCLI)

X_k is WCLI of X_j in $\boldsymbol{X}$ on $[0, \tau]$ if and only if

Λ_k **is** $(\mathcal{F}_{-jt})$**-previsible on** $[0, \tau]$,

where $\mathcal{F}_{-jt} = \mathcal{F}_0 \vee \mathcal{X}_{-jt}$ and $\mathcal{X}_{-jt} = \vee_{l \neq j} \mathcal{X}_{-lt}$.

Equivalently we can say in that case that X_k has the same Doob-Meyer decomposition in $(\mathcal{F}_t)$ and in $(\mathcal{F}_{-jt})$.

We will note in that case $X_j \not\longrightarrow_{\boldsymbol{X}} X_k$.

The differential equation representation

Doob-Meyer decomposition; Λ_t is differentiable

$$d\boldsymbol{X}_t = \lambda_t dt + dM_t, \tag{2}$$

with $\Lambda_t = \int_0^t \lambda_u du$.

Example:
$dX_{1t} = a\, dt + dW_{1t}$; $dX_{2t} = X_{1t} dt + dW_{2t}$

we have $X_1 \longrightarrow_{\boldsymbol{X}} X_2$ and $X_2 \not\longrightarrow_{\boldsymbol{X}} X_1$.

The differential equation representation

Doob-Meyer decomposition; Λ_t is differentiable

$$d\boldsymbol{X}_t = \lambda_t dt + dM_t, \tag{2}$$

with $\Lambda_t = \int_0^t \lambda_u du$.

Example:
$dX_{1t} = a\, dt + dW_{1t}$; $dX_{2t} = X_{1t} dt + dW_{2t}$

we have $X_1 \longrightarrow_{\boldsymbol{X}} X_2$ and $X_2 \not\longrightarrow_{\boldsymbol{X}} X_1$.

Joint model of HIV, immune system and disease

The HIV-immune system model

$$\begin{aligned} dT_t &= [\lambda - \gamma T_t V_t - \mu_T T_t]dt \\ dT_t^* &= [\gamma T_t V_t - \mu_{T^*} T_t^*]dt \\ dV_t &= [(1 - \eta I_t^A)\pi T_t^* - \mu_V V_t]dt \end{aligned}$$

The disease-immune system model

$$dD_t = I_{\{D_{t-}=0\}}\alpha_0 \exp(\beta_1 T_t + \beta_2 Z)dt + dM_t,$$

Guedj, Commenges, Thiébaut, *Biometrics*, 2011.

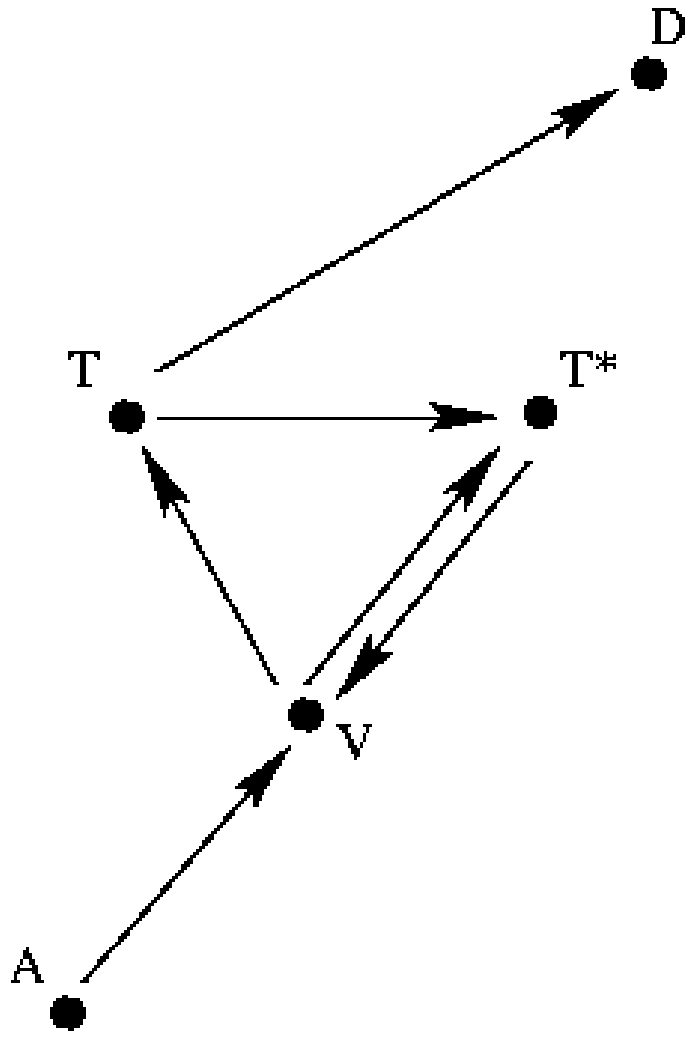


Figure: Graph for HIV: A: anti-protease; T: uninfected CD4; T*: infected CD4; D: clinical events.

Intervention systems

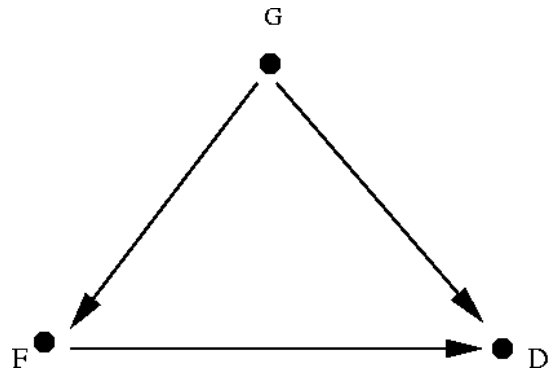


Figure: Observation system

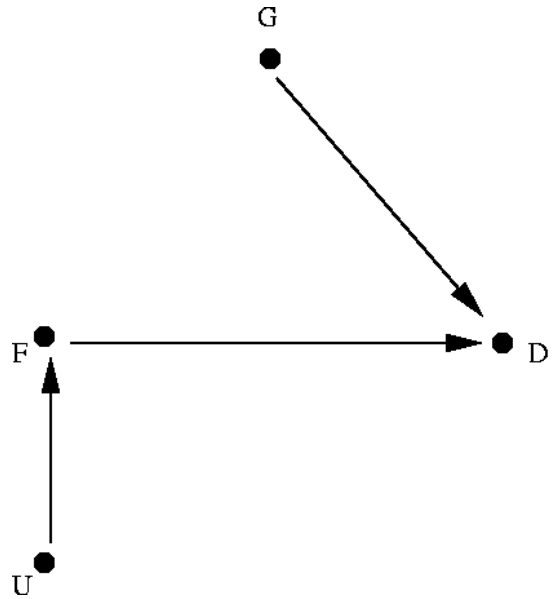


Figure: Intervention system

Marginal causal effect

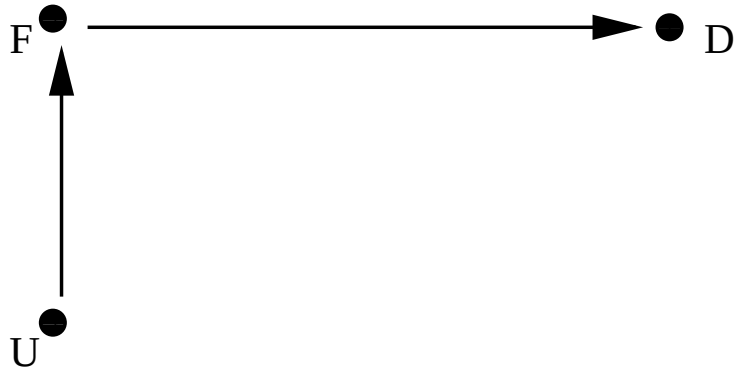


Figure: Imperfect intervention system

Marginal causal effect: the effect of F on D in this (imperfect) intervention system: the way the law of D changes as one changes the value f taken by F.

Marginal causal effect

Effect can be specified by: $P(D_t = 1|F = f) = S(t|f)$.
Generally, for G a time-constant process taking values 0 or 1 for all t :

$$S(t|f) = P(G = 1|F = f)S(t|f, G = 1)+P(G = 0|F = f)S(t|f, G = 0)$$

In the intervention system $P(G = 1|F = f) = P(G = 1)$.
Thus:

$$S^I(t|f) = P(G = 1)S(t|f, G = 1) + P(G = 0)S(t|f, G = 0)$$

Computation of the marginal causal effect from an observational system

From observation of a large number of replica of the good "observation system", we learn $S(t|f, G = 1$ and $S(t|f, G = 0)$, and also the marginal distribution of G. Thus we can compute

$$S'(t|f) = P(G = 1)S(t|f, G = 1) + P(G = 0)S(t|f, G = 0)$$

Imperfect observation system

Rather than using the perfect system:

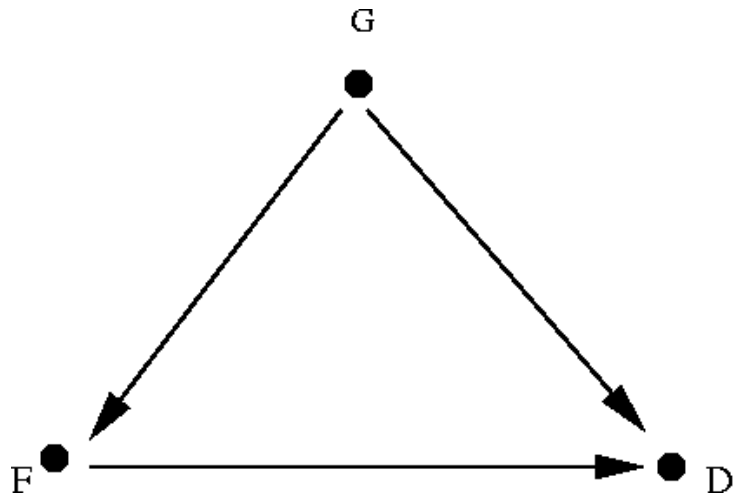


Figure: Perfect observation system

we use the imperfect system:

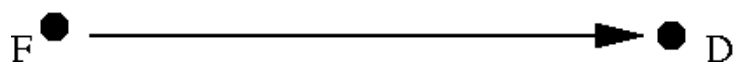


Figure: Imperfect observation system

Observed effect in imperfect observation systems

We learn:

$$S^O(t|f) = P(G=1|F=f)S(t|f,G=1)+P(G=0|F=f)S(t|f,G=0)$$

This is different from $S^I(t|f)$ when G has an influence on F.

Limitations of observation systems

Confounders

We are never sure that we have a perfect system; thus confounding is always possible.

Limitations of the marginal causal effect

- Emphasis on marginal causal effects: the so called back-door criterion for instance (Pearl, 2000: 3.3.1) gives conditions under which it can be computed.
- Dynamics better described in terms of causal effects including all the causal parents.
  $S(t|f, G)$ is more precise that $S(t|f)$
  For instance a treatment may be more efficient (or have less adverse effects) in subjects having a specific genotype or risk factor.
- Example of Statins: Give statins to every body to decrease CHD risk ?

Limitations of intervention systems

- Excessive confidence in the fact that the intervention trial will yield the marginal causal effect
- leads to a lack of modeling
- highly selected
- short term studies
- small number (often 2) of modalities

The example of the Women's Health Initiative (WHI)

Conflict between observation and intervention in the WHI

- Observational research on postmenopausal hormone therapy: $40 - 50\%$ reduction in coronary heart disease incidence
- Clinical trial of the WHI estrogen plus progestin found an elevated incidence.

Comparison between results from the WHI trial and observational study

- WHI observational study: 53,054 women 33% of whom were estrogen-plus-progestin users at baseline
- Trial:16,608 postmenopausal women aged 50-79 years,

Prentice et al. (2005), *American J. Epidemiol.*

Reconciliating observation and intervention in the WHI

Adjusting

Adjustment	Trial: HR	Observation: HR
Crude incidence ratio	1.18	0.50
Age-adjusted	1.21	0.71
Adjusted on risk factors	1.27	0.87

Crossing hazards

Time	Trial	HR	95% CI	Observation	HR	95% CI
<2	80	1.68	1.15, 2.45	5	1.12	0.46, 2.74
2-5	80	1.25	0.87, 1.79	27	1.05	0.70, 1.58
>5	28	0.66	0.36, 1.21	126	0.83	0.67, 1.01

Good modeling and common sense

Good modeling

- Good modeling aims at identifying all the main factor which are causal parents of a process of interest
- Good modeling aims at estimating the effect of the main factors. For this, it is necessary to have a flexible model and to take into account the way observations have been collected: sampling scheme, incomplete observations.

Bradford-Hill criteria

Even if we can go beyond thanks to sophisticated statistics, the Bradford-Hill criteria remain relevant for practical epidemiology.

Bradford Hill Criteria

1. Strength of association (odds ratio)
2. Consistency
3. Specificity
4. Temporal relationship (temporality) - not heuristic; factually necessary for cause to precede consequence
5. Biological gradient (dose-response relationship)
6. Plausibility (biological plausibility)
7. Coherence
8. Experiment (reversibility)
9. Analogy (consideration of alternate explanations)

Austin Bradford Hill (1965), "The Environment and Disease: Association or Causation?," *Proceedings of the Royal Society of Medicine*.

The Dynamic models approach to causality

Granger, C. W. J. 1969 Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**, 424-438.
Aalen, O. (1987) Dynamic modeling and causality. *Scand. Actuarial J.* **1987**, 177-190.
Didelez, V., 2008. Graphical models for marked point processes based on local independence, *JRSS-B*, **70**, 245-264.
Fosen, J., Ferkingstad, E., Borgan, O and Aalen, O.O. (2006) Dynamic path analysis: a new approach to analyzing time-dependent covariates. *Lifetime Data Analysis* **12**, 143-167.
Daniel Commenges and Anne Gégout-Petit (2009) A general dynamical model with causal interpretation. *JRRS-B* **71**, part 4, 1-18.
Anne Gégout-Petit and Daniel Commenges (2010) A general definition of influence between stochastic processes. *Lifetime Data Analysis* **16**, 33-44.