Can counterfactual theory be a complete theory of causality as we practice it in epidemiology?

Jan P Vandenbroucke

Leiden University Medical Center

Background idea (provocatively stated)

- Many books, papers, courses, talks in epidemiology about “causal inference” based on counterfactual theory and directed acyclic graphs (DAGs)
- Misnomer: they are **not** about causal inference, as practical assessment and acceptance of causality in epidemiology is based on integration of very diverse types of knowledge
- Precepts for ‘causal inference’ based on counterfactual and DAG thinking should be called ‘recipes for proper thinking’ (a thinking hygiene) to optimize design and analysis of individual studies

Outline

- Two examples of causality assessment in practice
  - Smoking, lung cancer and the constitutional hypothesis
  - Third generation oral contraceptives, venous thrombosis and the role of biochemistry
- Cross-word analogy by Susan Haack
- Elements of counterfactual reasoning
  - online book Hernán-Robins
- Juxtaposition and conclusion

Smoking, lung cancer & Fisher's constitutional hypothesis

- Objection to causal interpretation of association between smoking and lung cancer in cohort and case-control studies
- Hypothesis: a tendency to smoke would be 'constitutionally' (genetically) linked to the tendency to develop lung cancer
- Impossible to solve by epidemiologic studies
- To counter it by data, only solvable by:
  - RCT: impossible
  - Monozygotic twins discordant for smoking: was found, but only decades later (Carmelli, IJE 1996)

Fisher's constitutional hypothesis; Answer, Cornfield *et al.* JNCI 1959

- Incidence lung cancer raised over a few decades
- New environmental factor that causes lung cancer must have been introduced; otherwise, incidence should remain stable
- Counterargument: rise might be 'apparent'
  - diagnostic improvements, being more alert etc.
- Counter-counterargument: screening clinic for TBC in Denmark, since 1941, with unchanged procedures (invitation, radiology etc); increasingly more lung cancer diagnosed
- Time trend data: weak argument for smoking, can be any environmental factor
- Epidemiologic studies 'saved' from the constitutional hypothesis by time trend data

Reasoning, Cornfield *et al.,* JNCI 1959

- Reviewed all counterarguments, one by one, looking at their strength and weaknesses
- “Interlocking” of the arguments: next to time trends, also pathology (carcinoma in situ and epithelial dysfunction in lungs of smokers), animal experiments (high doses of tobacco-tar on skin), human observations of tar as a carcinogen (chimney sweepers)
- Conclusion: sufficient reason for action, even if still loose ends.
- *“Cornfield et al. remind us forcibly that deep conclusions often require synthesis of evidence of different kinds.”* (Cox DR, IJE 2009)

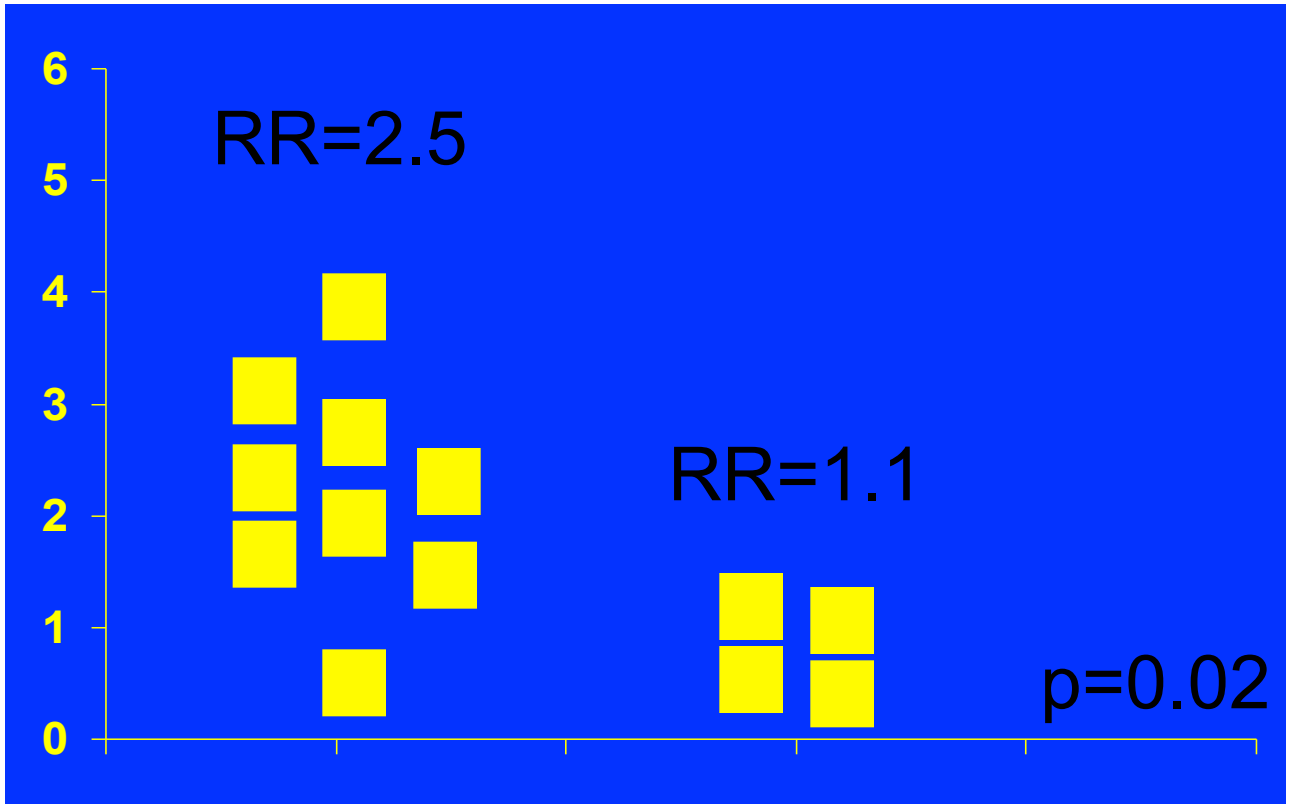
DAGs do not help

- DAGs do not help to decide whether all these data & arguments interlock: time trends, pathology (carcinoma in situ and epithelial dysfunction in lungs of smokers), animal experiments (high doses of tobacco-tar on skin), human observations of tar as a carcinogen (chimney sweepers) and analytic epidemiologic studies
- *Once that decision is made*, a DAG can be made about assumed underlying processes; otherwise these remain disjointed pieces of data
- DAGs help within the context of one study, within one assumed causal structure

Third generation oral contraceptives (OCs) and venous thrombosis

- 1980s, new OCs on the market “3rd generation”:
  - Same low dose estrogen
  - Different progestins; hope for more favorable lipid profile
- Unexpected finding in data from international WHO case-control study on cardiovascular diseases and OCs (1995): in UK and German study centers higher rates of venous thrombosis in users of 3rd generation Ocs, relative to older 2nd generation OCs
- 2nd generation OCs: 4-fold increase in risk relative to non-use; 3rd generation: 2 times higher= 8-fold

Relative Risk for Venous Thrombosis, 3rd generation vs. 2nd generation Oral Contraceptives



3rd generation OC controversy

- My original position: 80% certain
  - WHO, well conducted case-control study, many sites
  - Signal the same in Germany and UK
  - Selection bias or diagnostic bias unlikely: harm unknown
  - Studies not showing association needed ‘extra analyses’

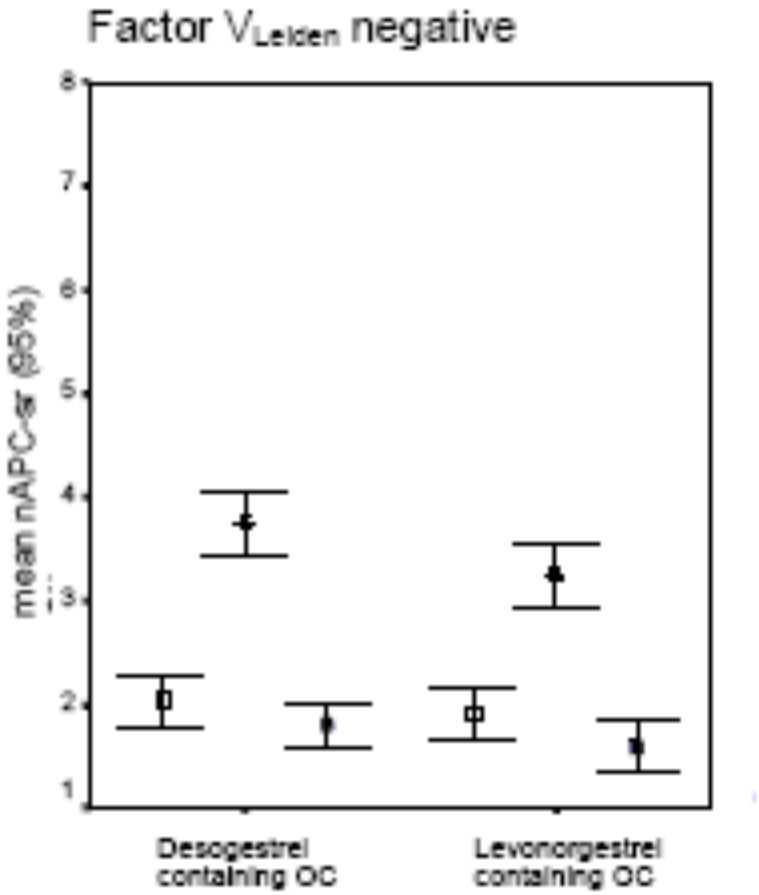
- Shifted to 98% percent certainty because of new biochemical data on effect on coagulation parameter(s), validated in cross-over trial

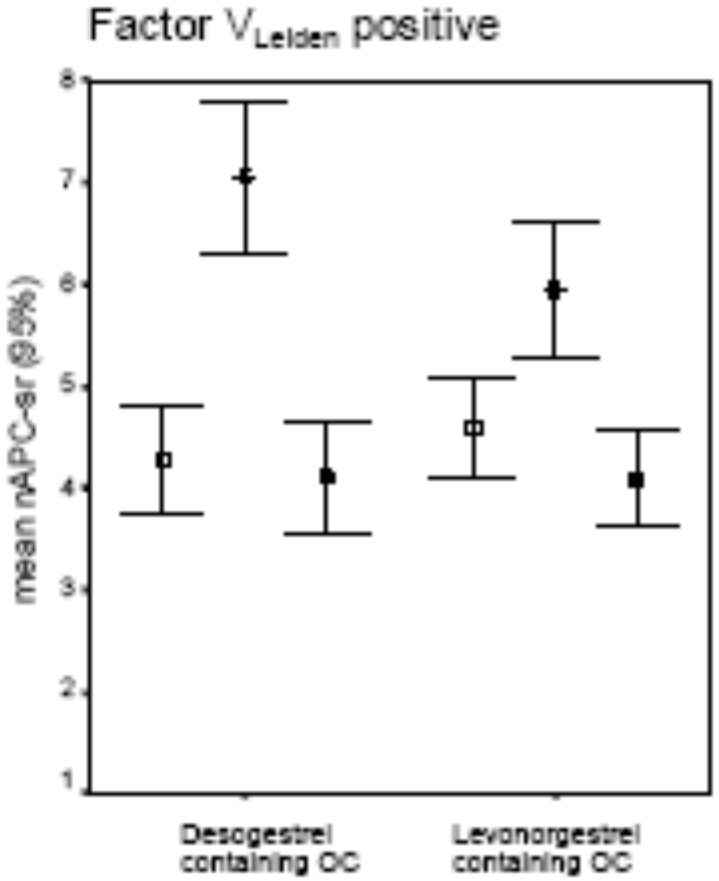
Discovery of Factor V Leiden mutation and APC resistance

- Factor V Leiden mutation: 5% Caucasians: increases risk of DVT 5-8 fold & also increases thrombotic effect of oral contraceptives (1994)
- Mechanism: increased blood clotting due to ‘Activated Protein C resistance’
- Two tests for Activated Protein C resistance:
  - Dahlbäck-test (1993): completely specific, one-to-one correspondence with mutation
  - Rosing-test (1997); global activity clotting system; test positive for persons with Factor V Leiden; by accident found to be sensitive to use of OCs
- Rosing test: discriminated between 2nd and 3rd generation OCs
- Controversy about test: cross-over trial

Cross-over trial of coagulation activation: 3rd vs. 2nd generation OC, women without and with FVL

(Kemmeren, Blood 2003)





Why convincing?

- Mechanism
  - not just imagined
  - shown in cross-over trial
- Effect OCs and in particular 3rd generation OCs on coagulation, similar to FVL
- For FVL, clotting activation due to APC resistance is a credible mechanism

Sanity warning for Epidemiologists!

- Laboratory is **not** 'the stronger causal science"
- Imagine: no increased incidence of venous thrombosis with oral contraceptives:
  - Rosing test: would be discarded - poor test, positive in many persons without FVL; no inquiry into OC use
  - Cross-over study would never have been thought of (madness!).
- Rosing test: meaning derives from epidemiology of OCs
- "Convincingness" derives from the interlocking of the arguments: both needed, reinforce each other
- A DAG will not show this, unless you decide to accept this causal chain: at first, there are separate DAGs for the biochemistry and for the epidemiology; when interlocking is accepted a DAG can be made with both present

Mutual fit of complementary evidence: the cross-word analogy

The clues [of the crossword] are the analogue of experiential evidence, already-completed entries the analogue of background information. How reasonable an entry in a crossword is depends upon how well it is supported by the clue and any other already intersecting entries; how reasonable, independently of the entry in question, those other entries are; and how much of the crossword has been completed.

Haack S. Manifesto of a passionate moderate. Chicago Univ Press 1998;95

Today's "causal inference": counterfactuals - history

- History of causal thinking: philosophers never gave satisfactory answer;
  - Hume (1739), impossible to derive causality from observations
  - Russel (1913), let's do away with the notion of causality; later revoked (1948)
- Pragmatic definition of cause: if you take it away, the incidence of disease will decrease *(Lilienfeld, Pub Health Rev 1957)*

Today's "causal inference": counterfactuals

- Today's hubris: assess causality from observations
- Imagine a *counterfact,* i.e., treat an individual one way, rewind the universe back in time, treat same person differently: proves causality in that individual
- Presupposes discrete well-defined intervention: no causal inference without manipulation (Holland 1986)

Counterfactuals

- Counterfactuals may not exist for gender (debate!), for socio-economic class, race...
- Ideal: medical interventions, educational interventions - the same person once treated and once not
- Request for “consistency”: interventions should be precisely defined and always exactly the same in different persons
- Non trivial: e.g. obesity *(Hernán, Int J Obesity 2008):* does a counterfactual exist for obesity?

What about rewinding the universe?

- We settle for group comparisons
- “Group causality”: more problematic: confounding, play of chance, loss to follow-up, uncertainty about which is causal mechanism investigated
- For group comparisons of interventions to be causally credible: complete ‘exchangeability’ (infinite numbers in an RCT?), no loss to follow-up, complete treatment adherence

Is anything in reality approaching this ideal?

- ‘Real life’ RCTs? Not really.
- Observational studies?
  - In theory: even Mendelian randomization falls short of randomization: e.g., HLA-matched siblings for bone marrow transplantation (Gray 1991); adjustment needed for family size (GD Smith 2006) – informative studies about effect in daily practice
  - For many problems observational evidence is better in practice: e.g., adverse effects close to unbiased allocation if adverse effect unpredictable; little extra adjustment needed; large numbers and reflects prescribing in daily practice

Ideas from another “Cornfield 1959”

- Dismisses notion that non-experimental studies never lead to accepting causality – *“with the implication that experiments can”*
- *“…the validation of experimental findings often requires their repetition under a variety of different circumstances”*
- There are *“important differences in degree”* between the possibility of spurious effects of randomized trials and observational studies, but… *“there is no difference in kind”.*
- *“If important alternative hypotheses are compatible with available evidence, then the question is unsettled, even if the evidence is experimental. But, if only one hypothesis can explain all the evidence, then the question is settled, even if the evidence is observational.”*

Cornfield, Principles of Research; Am J Ment Def 1959
(republished with commentaries in Stat Med, 2012, in press)

Juxtaposition

- Causality assessment in practice consists of integrating diverse types of knowledge; remains a judgment; Hume still right: ‘causal agents’ cannot be witnessed
- Almost never acceptance of causality based a single study; always multiple and preferably diverse studies from diverse branches of science
- Precepts about sharply defined and consistent interventions, exchangeability and use of DAGs:
  - very valuable for thinking about designing, analyzing, and interpreting single studies within a single accepted framework of thinking
  - make each study strive for the impossible: as if a causal judgment would be possible on that one study alone; this leads to better science

Theoretical epidemiology makes progress, and improves our practice

- Thinking about consistent well defined interventions & use of DAGs:
  - Makes it clear why nutritional epidemiology is stronger challenge than pharmacoepidemiology or genetics;
  - Makes it clear how difficult social epidemiology is, and what is needed (RCTs or IV, or change in exposure-analyses)
  - Helped to solve Hormone Replacement controversy about myocardial infarctions (Prentice & Hernán)
  - DAGs: clearer definition of necessary variables in analysis
- Analytic problems about repeated measurements of interventions solved (HIV treatment in cohorts)

Progress, but a mismatch remains

"None of this is to dismiss the utility of formal causal models. We need these models to give precise causal meaning to the associations we offer as estimated effects. That need becomes particularly acute with complex treatments, longitudinal interventions, and mediation analysis. **But the primary challenge in producing an accurate causal inference or effect estimate is most often that of integrating a diverse collection of ragged evidence into predictions to an as-yet unobserved target. This process does not fit into formal causal-inference methodologies currently in use,** … (….). Thus, while theory and methods for causal modeling have come a long way in past 40 years, they still have a very long way to go before they approach a complete system for causal inference."

Greenland, IN Berzuini *et al.* Causal Inference (2012), in press

Conclusion = Background idea (provocatively stated)

- Many books, papers, courses, talks in epidemiology about “causal inference” based on counterfactual theory and directed acyclic graphs (DAGs)
- Misnomer: they are **not** about causal inference, as practical assessment and acceptance of causality in epidemiology is based on integration of very diverse types of knowledge
- Precepts for ‘causal inference’ based on counterfactual and DAG thinking should be called ‘recipes for proper thinking’ (a thinking hygiene) to optimize design and analysis of individual studies

Discussion slides

Can the counterfactual exist?

Example *(Goldthorpe, 2001)*:

- "She did well on the exam because she is a woman" (*)
- "She did well on the exam because she studied for it"
- "She did well on the exam because she was coached by her teacher"

(*) Discussion of gender and causality: VanderWeele & Hernán. IN Berzuini *et al.* Causal Inference (2012) in press.