Izamna Fuentes
(Universidad Nacional Autónoma de México)

Inferring Causation in Epidemiology:
Causal Discovery and Complexity

Epidemiology is best understood as the study of distribution, frequency and determinants of disease and other health related events in a population (Bonita et al, 2006). At the same time, disease can be seen as an example of the product of a complex system, where the interactions of different events and factors give rise to the emergence disease. The biological components of disease has been the main focus of health sciences and biomedical research, however, the psychosocial determinants of health are drawing the attention of many researchers, giving rise to social epidemiology. The Brazilian epidemiologist, proponents of the theory of critical epidemiology claim that disease can't be understood by seeking explanations only at the biological level. Instead they argue that we must look into the political, cultural and economic levels to understand disease causation and distribution.

Based on the ideas presented above we´ll discuss the problem of complexity and causal discovery in epidemiology. We will argue that epidemiology´s main goal that is to explain disease distribution in a population should be understood as the explanation of a complex system whose components are the social, economical, cultural and biological factors while the outcome of that system is disease and disease distribution. To do so we´ll look to the multifactorial model of disease defended in Rothmand & Greenland (2005). We will see how the multifactorial concept of causality implies disease causation at different levels and allows for factors that range from the microlevel (genetic and biochemical factors) to the macrolevel (social and political factors). We´ll discuss the main ideas behind the multifactorial model of disease and how it deals with complexity in epidemiology. We will also discuss the critic made in Broadbent (2009) about the problems that plague what he calls the *bare multifactorial* model of disease, mainly, that admitting that disease is caused by different factors it´s not enough to give a proper concept of disease causation. Then we will look at what he calls the *contentful multifactorial* model of disease and see if this model can deal with the complexity of disease causation of epidemiology.

We will then at a probabilistic notion of causality as presented in Pearl (2000) and Parascandola (2001). On a probabilistic interpretation of causality a risk factor X is said to cause disease Y if the presence of X makes the presence of Y more likely, that is P(Y/X) > P(Y). Parascandola argues that the probabilistic concept of cause it´s more fitted for epidemiology since both the monocausal and the multifactorial models of disease can be expressed in a probabilistic language. We´ll argue that, while this is essentially true, epidemiology needs more than saying that there are multiple factors that are determinant on the emergence of disease and that we can use a probabilistic language to express this, we also need a counterfactual notion of disease in order to discriminate the proximate causes from those that farther away in the causal chain.

With that in mind, we'll discuss the reverse counterfactual presented by Broadbent in Broadbent (2008). His proposal it´s a change from the counterfactual analysis of causation given by Lewis, where C is said to cause E if ˜C→˜E. Broadbent proposes that reversing the counterfactual so that C is said to cause E if, all things remain similar, given the absence of the effect we assume the absence of the cause. According to Broadbent, this strategy makes it easier for us to distinguish the cause from a condition. We´ll argue that this strategy can help us to work out the problem of different factors causing disease, by discriminating factors from different domains with a counterfactual analysis.

This will take us to the next part of our discussion, where we will see how to deal with causal complexity and evidence in epidemiology, and how to make causal inferences in epidemiology. We will discuss the notion of explicative coherence presented in Thagard (1989) and the use of bayesian networks as those presented in Pearl (2000) and others.

We will see of this two theories, one based on coherence and the other based on probabilistic dependencies, deal with the problem of multiple evidence in epidemiology coming from different fields. It will be our proposal that a pluralistic methodology it´s the best way to deal with complexity and causality, and that any model of causality must take in count the mechanisms that give rise to a disease and the evidence coming from all fields.

References:

Beaglehole, R., Bonita, R., Kjellström, T. (2006). *Basic epidemiology.* China: World Health Organization.

Broadbent, A. (2008). The Difference Between Cause and Condition. *Proceedings of the Aristotelian Society*, 108, 355-364

Broadbent, A., (2009). Causation and models of disease. *Studies in the history and philosophy of the biological and biomedical sciences*, 40, 302-311

Greenland, S., Rothman K., (2005). Causation and causal inference in epidemiology. *American Journal of Public Health*, 95, 144-150

Parascandola, M., Weed, L., (2001) Causation in epidemiology. *Journal of Epidemiology and Community Health*. 55, 905-912

Pearl, J., (2000). *Causality, models, reasoning, and inference*. Estados Unidos. Cambridge University Press.

Thagard, P. (1989). Explanatory Coherence. Behavioral and brain sciences. 12, 435-502