

Some results on confounding

Tyler VanderWeele *

tvanderw@hsph.harvard.edu

The talk will consist of two parts. First, we propose a new criterion for confounder selection when the underlying causal structure is unknown and only limited knowledge is available. We assume all covariates being considered are pretreatment variables and that for each covariate it is known (i) whether the covariate is a cause of treatment, and (ii) whether the covariate is a cause of the outcome. The causal relationships the covariates have with one another is assumed unknown. We propose that control be made for any covariate that is either a cause of treatment or of the outcome or both. We show that irrespective of the actual underlying causal structure, if any subset of the observed covariates suffices to control for confounding then the set of covariates chosen by our criterion will also suffice. We show that other, commonly used, criteria for confounding control do not have this property. Second, we consider a series of formal definitions for a “confounder” (rather than simply “confounding”). The causal inference literature has not produced a clear formal definition of a confounder, as it has given priority to the concept of confounding over that of a confounder. We consider a number of candidate definitions arising from various more informal statements made in the literature. We consider the properties satisfied by each candidate definition, principally focusing on (i) whether under the candidate definition control for all “confounders” suffices to control for “confounding” and (ii) whether each confounder in some context helps eliminate or reduce confounding bias. Several of the candidate definitions do not have these two properties. Only one candidate definition of those considered satisfies both properties. We therefore propose that a “confounder” be defined as a pre-exposure covariate C such that there exists a set of other covariates X such that effect of the exposure on the outcome is unconfounded conditional on (X, C) but such that for no proper subset of (X, C) is the effect of the exposure on the outcome unconfounded given that subset. A variable that helps reduce bias but not eliminate bias we propose referring to as a “surrogate confounder.”

*Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue, Kresge Building, Boston, MA 02115, USA.