

In: **Computation, Causation, and Discovery**. Eds. P Glymour and G. Cooper.
Menlo Park, CA, Cambridge, MA: AAAI Press / The MIT Press. 1999. pp. 343-345.

CHAPTER ELEVEN

Response to Rejoinder

Clark Glymour, Peter Spirtes, Thomas Richardson

Robins and Wasserman agree with us that under a plausible set of assumptions, Tetrad II is an asymptotically correct procedure for inferring causal structure from observational data. We agree with Robins and Wasserman that on realistic sample sizes and data sets, if a zero-one loss function is used, no causal inference procedure that fails to employ very strong domain specific background knowledge is likely to be both correct and useful if variables of interest have many observed causes, or if influences are weak. Other loss functions, which emphasize the difference between the probabilities predicted from interventions and the actual probabilities upon interventions, are arguably more useful. We agree that the use of a sequence of statistical tests of conditional independence raise significant worries about the low power of the tests, particularly when there are a large number of covariates (as indicated by the remarks in causation, prediction, and search that it would be desirable to output a set of PAGs, rather than a single pattern, because the data might not be able to strongly discriminate among different PAGs). This seems to us to leave several interesting questions to be resolved: Assuming a non zero-one loss function, can the worries about the low power of the tests and large numbers of covariates be addressed by modifications to the Tetrad II procedures (or some other procedure entirely)? And what sorts of priors are needed in order to make the success of such procedures probable on realistic sample sizes? Robins and Wasserman note that the conclusions that epidemiologists draw from observational data rest on the assumption that "there are no strong unmeasured confounders" of the putative cause and its effect. The first point to notice is the somewhat ambiguous phrasing of "no strong unmeasured confounders." Given the kinds of priors that Robins and Wasserman describe, it is quite possible to have "no strong unmeasured confounders" while still having strong confounding (from the combination of many small confounders). Epidemiologists seem actually to make the stronger assumption that there is no strong unmeasured confounding.

What sort of background knowledge could one have which would reasonably support the conclusion that there is no strong unmeasured confounding? In physics, one might believe as much about the motions of the planets because one is very familiar with all the kinds of forces typically operating in such a situation and how those forces are generated. This is in short contrast to the situation in epidemiology where a myriad of different factors, varying from disease to disease, are the operative causes, and the ability to predict who will contract a given disease is quite poor. If the conclusions of epidemiology rest on what Robins and Wasserman consider to be an untestable substantive assumption, then this certainly puts all causal conclusions from observational studies on extremely shaky ground. So any reliable method that can make weaker assumptions would be valuable. One way of addressing the worries about low power and large numbers of covariates, as well as the problem of sequential testing of conditional independence relations and the possible low power of such tests, is to assign a score (such as the Bayes Information Criterion [BIC]) to a (*d*-separation) equivalence class of causal models (see Spirtes and Richardson [1997]). In this way, instead of dealing with a sequence of tests, there is a sense in which all of the conditional independence relations are simultaneously tested. Under the procedures we have described, if model *M1* were created by removing an edge from model *M2* because of a statistical test of conditional independence, the search would output model *M1*, with nothing in the output to indicate whether the test had high or low power. In the modified procedures we have devised, a BIC score would be calculated for both *M1* and *M2*, and if the difference in the scores was small (as would be the case if the test of conditional independence which caused the edge to be removed from *M2* had low power) both of the models and their respective scores would be output. On the other hand, if the test of conditional independence had high power, then the difference in BIC score between the two models would be large, and only *M1* would be output. In addition, because there is a whole family of BIClike scores which are consistent, but assign different penalties to the complexity of a model, it would be possible to have the user put a kind of quasi-prior down on simplicity by adjusting the penalty term.

Even with the suggested modifications to Tetrad, it is still the case that the sorts of priors that would support the conclusions the program draws would incorporate acceptance of a kind of faithfulness assumption. Is this assumption reasonable? It will help to distinguish between several different versions of the causal faithfulness assumption.

We and Robins and Wasserman agree on a principle that we will call the population causal faithfulness assumption: If *X* and *Y* are independent conditional on *Z*, then *X* and *Y* are not directly causally connected except through *Z* (i.e., *X* does not cause *Y* directly, *Y* does not cause *X* directly, and no third variable outside of *Z* causes both *X* and *Y* directly). If one wishes to draw ex-

act causal inferences from finite samples, a different assumption is needed. Call it the *sample causal faithfulness assumption*: in a large sample, if X and Y are almost independent conditional on Z , that is evidence that X and Y are not directly causally connected except through Z . Robins and Wasserman have presented a prior in which the sample causal faithfulness assumption is false. We have already given our comments on their prior.

The sample causal faithfulness assumption is certainly used in our search procedures; however, in many practical situations a still weaker principle will suffice for causal inference. For that reason, we will turn to an assumption that is relevant when the difference between no direct causal relation and a very weak direct causal relation is not important. For those situations, the relevant inference principle may be called the *weak sample causal faithfulness assumption*: in a large sample, if X and Y are almost independent conditional on Z , that is evidence that X and Y are at most weakly directly causally connected except through Z .

Should the weak sample causal faithfulness assumption be accepted? If one is a Bayesian, the answer obviously depends upon what prior one holds. Is the weak sample causal faithfulness assumption accepted by epidemiologists? Note that both the Robins-Wasserman prior, and their claim that epidemiologists place a very high prior on every population correlation being non-zero, while incompatible with the sample causal faithfulness assumption, are perfectly compatible with the weak sample causal faithfulness assumption. Anyone who accepts that an almost zero correlation between X and Y in a large sample (even if the correlation is statistically significant) is at least prima facie evidence for a weak direct causal connection between X and Y thereby accepts an instance of the weak sample causal faithfulness assumption. In addition, anyone who accepts that a small (but perhaps statistically significant) coefficient for a variable X in a regression (linear, logistic, or other kind) is at least prima facie evidence for a weak direct causal connection between X and the outcome variable, also uses an instance of the weak sample causal faithfulness assumption.

Acknowledgement

Research for this paper was supported by grants DMS-9704573 and BES-940239.