Reviewer's report

Title: Combining Directed Acyclic Graphs and the change-in-estimate procedure as a novel approach to adjustment-variable selection in epidemiology

Version: 3 Date: 10 September 2012

Reviewer: Ilya Shpitser

Reviewer's report:

The authors propose a method which performs limited causal model search, followed by an estimation procedure of a causal effect by covariate adjustment. This method proceeds as follows:

First, given a set of causal DAGs that are "mostly," but not "completely" correct, find a minimal adjustment set S, obtain an effect estimate for a causal effect of A on Y via the adjustment formula \( \sum_{s} p(y | a, s) p(s) \).

Second, add (subtract) one additional element to S, and check whether the resulting adjustment formula gives "sufficiently different" effect estimates. If the differences do not follow given any of the prior DAGs (given some sort of faithfulness assumption), exclude inconsistent DAGs and perform a kind of local search for DAGs that are consistent with the pattern of equalities among adjustment functionals.

The authors apply their algorithm to a hypothetical and a real situation.

This is a second revision of a manuscript whose first version I also reviewed. First, I wanted to thank the authors for their hard work -- I think the manuscript has improved tremendously. I do have a few additional comments from a causal graphical models point of view.

First, the ad hoc augmentation algorithm provided in the supplementary materials has unclear wording:

"Write out each path in the DAG from A to Y, connected to A but not to Y, and connected to Y but not to A. Include the path AY. Show measurement error in the DAG as descending variables on paths;"

I don't understand this, needs rephrasing. A path from A to Y is by definition connected to A and to Y. What does the last sentence mean? Genealogy terminology would probably help (measured proxies as children of unobserved variables of interest?)

Second, it is not entirely clear what can be said about the algorithm the authors propose. When is it guaranteed to terminate? What is the correct answer (the set of some/all DAGs consistent with the set of adjustment formulae considered + the effect estimate?) When will it return the correct answer? What can be said about the set of DAGs that are returned? What is its running time (polynomial?)
The algorithm looks to be “limited horizon local search” in that it doesn’t seem like it will work if the initial set of DAGs is “far away” from the truth. But this is just intuition -- I am not really sure what would happen for various sorts of inputs, and there is no analysis of this in the paper.

The authors seem to consider hidden variables in some cases in the search -- e.g. measurement error of covariates, or “instrumental variable” situations, but these situations are ad hoc. Why not a general scheme for handling hidden variables via Verma/Pearl latent projections (or Richardson et al. Acyclic Directed Mixed Graphs ADMGs), which can represent arbitrary hidden state via bidirected arrows <->.

The authors ran the algorithm on two examples. Is there a proof that the procedure is sound in any sense? That is, is it the case that it will never return “the wrong thing” for any type of inputs plausibly encountered in epidemiologic practice? If so, how would we know? Is there a proof that the procedure is “exhaustive”/complete in any sense? That is, is it the case that when it returns _all_ DAGs with a given set of observables which are consistent with the set of adjustment functionals considered?

What exactly is the faithfulness assumption that is needed to get this algorithm to work? It’s not ordinary faithfulness. It could be that the graphs are faithful (that is d-separation in the graph if and only if conditional independence in the corresponding density is true) but it may still happen by coincidence that adjustment functionals may be equal even if one of them uses an invalid adjustment set for a given graph. I think one needs a type of faithfulness that holds after reweighting (that is division by a conditional density p(treatment|past) ).

To summarize, I think my original comments were addressed in the revision. I understand, roughly, what the algorithm the authors propose is supposed to do. However, based on the manuscript in its current state, I have a very limited idea of what exactly is going on with this algorithm. When does it terminate, how quickly does it run, is it sound and/or exhaustive in any sense? I think more evidence than two examples is needed to convince me that this is a correct and useful procedure. At the very least, something about the soundness of the algorithm needs to be said.

Minor notes:

The pdf version of the manuscript I received contained many instances of bolded text (sometimes entire sentences). I am not sure if this was intended typesetting, or an artifact of converting from Word to pdf, or something else. I did find this typesetting somewhat unusual for a journal manuscript, and distracting.

**Level of interest:** An article of limited interest

**Quality of written English:** Acceptable
Statistical review: No, the manuscript does not need to be seen by a statistician.