Evaluation of Medication-Mediated Effects in Pharmacoepidemiology"

By EJ Tchetgen Tchetgen and K Phiri

eAPPENDIX 1

Additional discussion of Identification of NDE(a,a*) *and NIE*(a,a*)

It is well known that the average total effect TE(a,a*) of A on Y is identified under assumption (a.1), and is given by the g-formula of Robins.[1] However, assumptions (a.1)-(a.3) do not suffice to identify NDE(a,a$*$) and NIE(a,a*). Recently, Pearl[2, 3] provided an alternative interpretation of the graph in Figure 1 under a nonparametric structural equation model with independent error (NPSEM-IE), which yields in addition to (a.1)-(a.2) the following stronger version of assumption (a.3)

(c.3) M(a) is independent of Y(a',m) given C and A=a, for all a, a'.

While assumptions (a.1)-(a.3) could in principle be enforced under an experimental design, assumption (c.3) could never be made to hold, even under an experimental design because this latter assumption involves potential outcomes for a given person under possibly conflicting values for the mediator .[4] However, whereas neither NDE(a,a$*$) nor NIE(a,a*) is identified under assumptions (a.1)-(a.3), Pearl[2] formally established that under the NPSEM-IE interpretation of the causal graph of Figure 1, i.e. essentially under assumptions (a.1), (a.2) and (c.3), the average natural direct effect is identified by the so-called mediation formula. Therefore the average natural indirect effect is likewise identified since NIE(a,a$*$)=TE(a,a$*$)-NDE(a,a$*$). So far, this discussion has only considered the situation depicted in Figure 1, where one can reasonably assume that there is no exposure-induced confounding of M. This assumption may be

unrealistic in many epidemiologic applications, particularly in pharmacoepidemiology settings where A indicates disease status, and the mediator M is medication taken to slow disease progression. Then, in observational settings, because the decision to initiate treatment is likely based on a patient's current health status, exposure-induced confounding of the mediator-outcome relation can seldom be ruled out. In the causal diagram depicted in Figure 2, N now encodes an exposure-induced confounder of M. Thus, in this graph, N is simultaneously a confounder of the effects of the mediator M on Y, and on the causal pathway from exposure to outcome. While the total effect of A on Y remains identified in Figure 2 by Robins' g-formula, so that the presence of N presents no new difficulty, the situation is quite different for NDE(a,a$*$). Specifically, according to a result by Avin et al,[5] a causal effect along a specific path is identifiable in a fully observable nonparametric structural equations model if and only if there is no so-called "recanting witness," namely there is no variable that mediates the causal pathway of interest from A to Y, while at the same time mediating another causal pathway from A to Y that is not of interest. Note that the direct effects of A on outcome Y with respect to the mediator of interest M, in Figure 2, consists of the two pathways A→Y and A→N→Y. But the variable N also mediates the indirect effect A →N→M→Y, which is not of interest when direct effects are in view, and therefore the exposure-dependent confounder N is a recanting witness for the direct-effect path A →N→Y. This in turn implies that NDE(a,a$*$) is not identified without an additional assumption even under Pearl's NPSEM-IE for the causal diagram in Figure 2. As shown, the recanting witness is an exposure-dependent confounder of the mediator-outcome relation. Thus one may conclude that, without an additional assumption, if such a variable is present in a given application, NDE(a,a$*$) is not identifiable even under a nonparametric

structural equations model. To address this issue, Robins and Richardson,[4] and Tchetgen Tchetgen and VanderWeele[6] considered a variety of assumptions under which an NPSEM-IE corresponding to the graph in Figure 2 leads to identification of NDE(a,a$*$). Tchetgen Tchetgen and Phiri[7] recently took a different approach and provided nonparametric bounds for NDE(a,a*) in the simple case of binary M under the more conventional no-unobserved confounding assumptions encoded in the graph in Figure 2.

In this paper, we have considered more straightforward, arguably less controversial identification conditions under the graph depicted in Figure 2, given by (a.1), (a.2), (b.3) and (b.4). The first three assumptions are standard no unobserved confounding conditions necessary to make certain causal statements about the effects of interventions from observational data, and the last assumption is expected to hold in the setting considered herein. Thus, our results firmly establish our earlier claim that medication-mediated effects are in a sense immune to recent criticism leveled at causal mediation methodology as relying on overly stringent and generally untestable conditions.

Proof of Robustness of medication-mediated effects to unobserved confounding of disease status

Under assumptions (a.1), (a.2), (b.3) and (b.4) a straightforward application of Robins' g-formula gives:

$$\text{NIE}(1,0,c) = E\{Y(1,M(1))|c\} - E\{Y(1,M(0))|c\}$$

$$= \sum_{m,n} \mathrm{E}(\mathrm{Y}|\mathrm{A}=1,\mathrm{m},\mathrm{n},\mathrm{c})\Pr(\mathrm{M}=\mathrm{m},\mathrm{N}=\mathrm{n}|\mathrm{A}=1,\mathrm{c})$$

$$- \sum_{n} \mathrm{E}(\mathrm{Y}|\mathrm{A}=1,\mathrm{M}=0,\mathrm{n},\mathrm{c})\Pr(\mathrm{N}=\mathrm{n}|\mathrm{A}=1,\mathrm{c})$$

However, also note that

$$\sum_{m,n} \mathrm{E}(\mathrm{Y}(\mathrm{m})|\mathrm{A}=1,\mathrm{m},\mathrm{n},\mathrm{c})\Pr(\mathrm{M}=\mathrm{m},\mathrm{N}=\mathrm{n}|\mathrm{A}=1,\mathrm{c})$$

$$-\sum_{n} \mathrm{E}(\mathrm{Y}(\mathrm{m}=0)|\mathrm{A}=1,\mathrm{M}=0,\mathrm{n},\mathrm{c})\Pr(\mathrm{N}=\mathrm{n}|\mathrm{A}=1,\mathrm{c}) \text{ (by consistency)}$$

$$= \mathrm{E}(\mathrm{Y}(\mathrm{M}(1))|\mathrm{A}=1,\mathrm{c}) - \mathrm{E}(\mathrm{Y}(\mathrm{M}(0)|\mathrm{A}=1,\mathrm{c}) \text{(by (b. 3) and (b. 4))}$$

$$= \mathrm{E}(\mathrm{Y}\big(\mathrm{M}(1)\big) - \mathrm{Y}\big(\mathrm{M}(0)\big)|\mathrm{A}=1,\mathrm{c}),$$

Thus proving the result

References:

1. Robins JM. (1997). Causal Inference from Complex Longitudinal Data. Latent Variable Modeling and Applications to Causality. Lecture Notes in Statistics (120), M. Berkane, Editor. NY: Springer Verlag, pp. 69-117.
2. Pearl J. The causal mediation formula–a guide to the assessment of pathways and mechanisms. *Prev Sci*. 2012;13:426–436.
3. Pearl, J. (2001). Direct and indirect effects. In Proceedings of the 17th Annual Conference on Uncertainty in Artificial Intelligence (UAI-01), San Francisco, CA, pp. 411--42. Morgan Kaufmann.
4. Robins JM, Richardson, TS. (2010). Alternative graphical causal models and the identification of direct effects. To appear in Causality and Psychopathology: Finding the Determinants of Disorders and Their Cures. P. Shrout, Editor. Oxford University Press.
5. Avin, C., I. Shpitser, and J. Pearl (2005). Identifiability of path-specific effects. In IJCAI-05, Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence, Edinburgh, Scotland, UK, July 30-August 5, 2005, pp. 357--363.
6. Tchetgen Tchetgen EJ and VanderWeele T. On identification of natural direct effects when a confounder of the mediator is directly affected by exposure. (2013) Epidemiology. Mar;25(2):282-91.
7. Tchetgen Tchetgen EJ, Phiri K. Bounds for pure direct effect. Epidemiology. 2014;25:775–776.

.

eAPPENDIX 2:

Below we provide the SAS code used in the data application section of the manuscript to obtain point estimates and wild bootstrap estimates used to obtain confidence intervals (for the microcephaly outcome).

```
**************************************
Step 1 - Estimate the TOTAL EFFECT
**************************************;

%macro analysis (data=newborn);

data newsamp; set newborn;
bootw=rand('exponential'); *bootstrap weight, should be commented out to
obtain point estimates for the observed sample;
*bootw = 1; *use this to perform analysis in the observed sample in order to
obtain point estimates;
run;

proc logistic data=newsamp outest = total (keep= epilepsy) descending;
model micro = epilepsy momage newrace;
weight bootw;
run;

data total;
set total;
epilepsytotal=epilepsy;
drop epilepsy;
run;

**************************************
Step 2 - Estimate the DIRECT EFFECT
**************************************;

/* 1) First calculate the weights for A = 1 (i.e Women with epilepsy)*/
data epilepsy;
set newsamp;
where epilepsy = 1;
run;

*DENOMINATOR for the weights;
proc genmod data=epilepsy;
model AED = alc2 cig2 sub2 seiz/ dist=bin link=log;
output out=model2 predicted=pAED_1;
weight bootw;
run;

*Unstabilized WEIGHTS for A = 1;
data epilepsy_w;
set model2;
weight = 1/pAED_1;
run;
```

```
/* 2) Second calculate the weights for A = 0 (i.e Women without epilepsy)*/
data noepilepsy; *These are women who have no epilepsy;
set newsamp;
where epilepsy = 0;

*WEIGHTS for E = 0;
weight = 1*bootw;
run;

/* 3) Regress Y (major malformation) on Epilepsy among subset with M=0 (i.e.
in those with AED = 0, either with epilepsy or not) to estimate the DIRECT
EFFECT*/;
data Epilepsy_noAED; *These are women who have epilepsy but not on AEDs;
set epilepsy_w;
where AED = 0;
run;

*Combine datasets noepilepsy (i.e women not on AED because they dont have
epilepsy) and epilepsy_noAED created above;
data noAED_all; *These are all women not on AEDs (with or without epilepsy);
set noepilepsy epilepsy_noAED;
run;

proc logistic data = noAED_all outest = direct(keep = epilepsy) descending;
model micro = epilepsy momage newrace;
weight weight;
run;

data direct;
set direct;
epilepsydirect=epilepsy;
drop epilepsy;
run;

data result;
merge total direct;
run;

%mend;

/* 4) Generate 1000 random samples with replacement to obtain wild bootstrap
95% confidence intervals for all estimated effects*/

%macro loop;
 data final_result; run;
 ods listing close;
     %do sim_count = 1 %to 1000; *use this command line to perform bootstrap
analysis, should be commented out to obtain point estimates for the observed
sample;
       /*%do sim_count = 1 %to 1;*/*use this to perform analysis in the
observed sample in order to obtain point estimates;

      %put running sample &sim_count ;
      title "for sample &sim_count ";
          %analysis(data=newborn);
            %if &sim_count=1 %then %do;
                  data trial1; set result; run;
```

```
            %end;
            %else %do;
          data trial1;set result trial1;
           sample = %eval(&sim_count) ;
           run;
           %end;

      %end;
%mend ;

  %loop;
ods listing;


*****************************************
Step 3 - Estimate the INDIRECT EFFECT
*****************************************;
data trial2 ;
set trial1;
epilepsyindirect=epilepsytotal-epilepsydirect;
run;
```