THE 3RD "NEGATIVE CONTROL": NC POPULATION

Causal inference group meeting

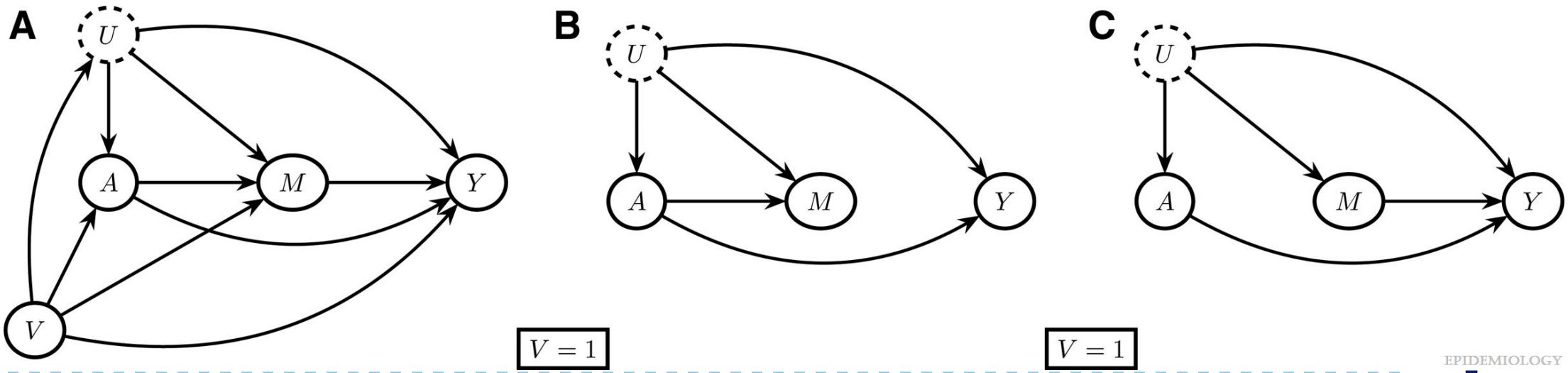
Dept. of Epidemiology,
Erasmus MC

30 April 2025

Erasmus MC
Universitair Medisch Centrum Rotterdam

Using Negative Control Populations to Assess Unmeasured Confounding and Direct Effects

Marco Piccininni[a,b,c] and Mats Julius Stensrud[d]



Piccininni M, Stensrud MJ. Using Negative Control Populations to Assess Unmeasured Confounding and Direct Effects. Epidemiology 35(3):313-9.

SUMMARY

- *What's known:*
  NC population is a (sub-)population that is not affected by the treatment,
  e.g. $\mathbb{E}\left[Y^a|V=1\right] = \mathbb{E}\left[Y^{a'}|V=1\right] \forall a, a' \in \mathcal{A}$
- *What's new:*
  Formal definition and causal diagrams of NC population is provided;
  Usage of NC population for (1) checking unmeasured confounding and (2) checking exclusion restriction (the presence of direct effects) is provided.
- *What's useful:*
  Eligible NC population could be a nice method to rule out unmeasured confounding and the presence of unknown causal pathways

MOTIVATING EXAMPLES OF USE

- Example 1. In an *ex vivo* experiment, I am worried about that some pre-experiment (e.g. assay agent production) issues are actually the cause of my positive outcome.

- Example 2. In an instrumental variable analysis, I am worried about the violation of exclusion restriction assumption for some reasons.

- Example 3. In an RCT, I am worried that placebo effects account for the majority of the observed effects of my painkiller; the RCT is however open-label and cannot be blinded.

INFORMAL DEFINITION OF NC POPULATION

- NC population: a “control” in which the subjects are not affected by a certain treatment, and similar to “the treated and the control groups subjected to randomization”
- Informal requirement of this (sub-)population:
  - The NC population is similar enough to the population of interest;
  - The NC population would have had the same outcome with or without the treatment;
  - The NC population is expected to have a null observed effect after the treatment.
- Therefore, a positive observed effect over the NC population is used to falsify a certain causal conclusion over the target population.

FORMAL DEFINITION OF NC POPULATION

Assumption set 1: for ruling out unmeasured confounding

- An NC-population is a population denoted by pretreatment variable $V = 1$ (either a sub-population of the target, or an external population that is *similar enough**) in which there is:

  [1] No response to the treatment, **on average**:

  $$E[Y^{a=1}|V = 1] = E[Y^{a=0}|V = 1]$$

  [2] No perfect cancellation:

  $$\text{if } Y \perp A \mid (V = 1), \text{ then } Y^a \perp A \mid V = 1$$

  [3] similar confounding structure: (?*)

  $$\text{If } Y^a \perp A \mid (V = 1), \text{ then } Y^a \perp A \mid V = 0$$

FORMAL DEFINITION OF NC POPULATION

Assumption set 1: for ruling out unmeasured confounding

- An NC-population is a population denoted by pretreatment variable $V = 1$ (either a sub-population of the target, or an external population that is *similar enough**) in which there is:

  [1] No response to the treatment, **sharply**:

  $$Y_i^{a=1} = Y_i^{a=0} \; \forall \; i \text{ such that } V_i = 1$$

  ~~[2] No perfect cancellation:~~

  ~~if $Y \perp A \mid V = 1$, then $Y^a \perp A \mid V = 1$~~

  [3] similar confounding structure: (?*)

  If $Y^a \perp A \mid V = 1$, then $Y^a \perp A \mid V = 0$

ESTIMATION OF ATE

Usage 1: Ruling out unmeasured confounding in estimating ATE

- Empirical testing of the presence of unmeasured confounding:

$$E[Y|A=1,V=1] \neq E[Y|A=0,V=1] \Rightarrow \text{unmeasured confounding}$$

- Estimating ATE **<u>when [2] and [3] are satisfied</u>**:

$$E[Y|A=1,V=1] = E[Y|A=0,V=1] \Rightarrow$$

$$ATE = \Pr[V=0]\,(E[Y|A=1,V=0] - E[Y|A=0,V=0]);$$

$$CATE = \tau(V=0) = E[Y|A=1,V=0] - E[Y|A=0,V=0]$$

FORMAL DEFINITION OF NC POPULATION

Assumption set 2: for ruling out both unmeasured confounding and direct effect (or effect outside known mechanisms)

- An NC-population is a population denoted by ***pretreatment and pre-confounder*** variable $V = 1$ (either a sub-population of the target, or an external population that is *similar enough**) in which there is:
- [4] No indirect effect in NC-population:
  - 4a: $\Pr[M^{a=1} = m|U, V = 1] = \Pr[M^{a=0} = m|U, V = 1]$ (no A-M pathway), *or*
  - 4b: $Y \perp M \mid (V = 1, U, A)$ (no M-Y pathway)

  [5] No perfect cancellation of direct effects:
  - $Y \perp A \mid (V = 1) \Rightarrow E[Y^{a=1}|M^{a=1}, U, V = 1] = E[Y^{a=0}|M^{a=0}, U, V = 1]$

  [6] Similar direct effect structure ("no context-specific direct effect"):
  - no direct effect in $V = 1 \Rightarrow$ no direct effect in $V = 0$

Usage 2: for ruling out both unmeasured confounding and direct effect

- Empirical testing of the presence of unmeasured confounding *and* direct effect (or other causal pathway via unmeasured mediators):

$$E[Y|A = 1, V = 1] \neq E[Y|A = 0, V = 1] \Rightarrow$$

unmeasured confounding *or* presence of direct effect

- (In IV analysis case this indicates the violation of exclusion restriction assumption, if there are other arguments ruling out unmeasured instrument-outcome confounding)

SOME INTERESTING QUESTIONS TO DISCUSS

- 1. Can individuals with immune ($Y^{a=1} = Y^{a=0} = 0$) or doomed ($Y^{a=1} = Y^{a=0} = 1$) response types be used as NC-population? – and relationship between response type and NC-population.
- 2. Is there any better ways to define “similar confounding structure”?
- 3. (in case you read eText2 in supplementary material): how different is it between “a population in which alcohol consumption is absent” and “individuals who do not drink alcohol”: is it just a matter of language?

A FUN FACT...

PubMed®

"negative control"[All Fields] AND ("generation R"[All Fields] OR "Rotterdam

Search

Advanced Create alert Create RSS User Guide

Save Email Send to Sort by: Most recent Display options

MY CUSTOM FILTERS

All (4)
RCT (Cochrane) (1)
Review (0)
humans (MB) (4)
Children (MB) (4)
– show fewer

4 results

Page 1 of 1

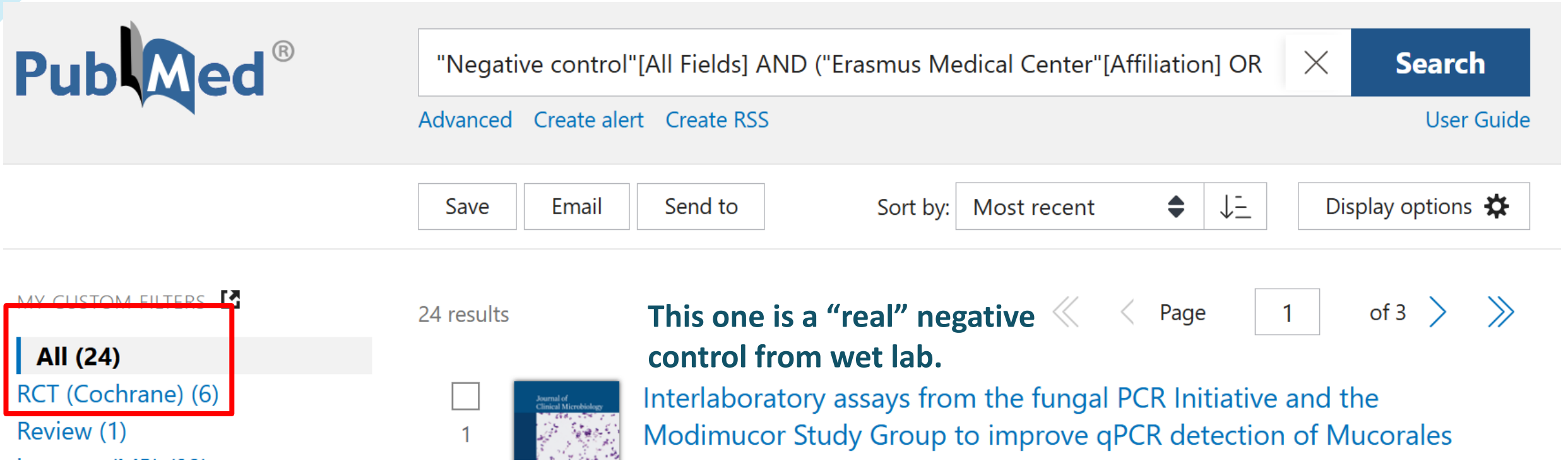
1
Cite
Share

Integrating multiple lines of evidence to assess the effects of maternal BMI on pregnancy and perinatal outcomes.

Borges MC, Clayton GL, Freathy RM, Felix JF, Fernández-Sanlés A, Soares AG, Kilpi F, Yang Q, McEachan RRC, Richmond RC, Liu X, Skotte L, Irizar A, Hattersley AT, Bodinier B, Scholtens DM, Nohr EA, Bond TA, Hayes MG, West J, Tyrrell J, Wright J, Bouchard L, Murcia M,

Erasmus MC
Universitair Medisch Centrum Rotterdam

TWO FUN FACTS…



- (fyi: only 9 of 24 mention epi/clin- negative controls)

QUESTIONS?

And suggestions?