REFRAMING THE PROBLEM OF MEDIATION

We now reframe mediation with potential outcomes. Following Holland (1988) and Robins and Greenland (1991), we define several estimands.

We assume treatment assignment Z is binary. We place no restrictions on the mediator M or outcome Y.

We rewrite the potential outcomes $Y_i(0)$ and $Y_i(1)$ as $Y_i(0, M_i(0))$ and $Y_i(1, M_i(1))$, respectively, where $M_i(0)$ is the value unit i would take if not assigned to treatment, and $M_i(1)$ the values if i is assigned to treatment.

We define unit effects of Z on M and Y as $M_i(1) - M_i(0)$ and $Y_i(1, M_i(1)) - Y_i(0, M_i(0))$, respectively.

REFRAMING THE PROBLEM OF MEDIATION

The unit effect of Z on Y is called a total unit effect. It is the outcome when unit i is treated and responds to treatment with the intermediate outcome value $M_i(1)$ versus the outcome when i is not treated and responds with intermediate outcome $M_i(0)$.

The unit effect operates through two channels: 1) “indirectly”: treatment may affect M (when $M_i(0) \neq M_i(1)$) and this may affect outcome Y, and 2) “directly”: even if $M_i(0) = M_i(1)$, $Y_i(0, M_i(0)) = Y_i(1), M_i(1))$. Then Z affects Y through channels other than M.

In order to separate these channels, we consider a hypothetical study with potential outcomes $Y_i(z, m)$. $z = 0, 1$ and m can take all possible values in the range of the random variable M.

REFRAMING THE PROBLEM OF MEDIATION

We can now define unit effects $Y_i(z, m) - Y_i(z^*, m^*)$.

"Controlled direct effects" of Z on Y, $Y_i(z, m) - Y_i(z^*, m)$, are commonly considered. However, controlled direct effects of M on Y, $Y_i(z, m) - Y_i(z, m^*)$, are not commonly considered.

Averaging the unit effects leads to average controlled effects (CDE) or average controlled effects conditional on $\mathbf{X}$:

$$CDE(\mathbf{X}_i) = E(Y_i(z, m) - Y_i(z^*, m^*) \mid \mathbf{X}_i).$$

REFRAMING THE PROBLEM OF MEDIATION

We can also decompose the total effect into "direct" and "indirect" components:

$$
\begin{aligned}
Y_i(1, M_i(1)) - Y_i(0, M_i(0)) &= \{Y_i(1, M_i(1)) - Y_i(1, M_i(0))\} + \{Y_i(1, M_i(0)) - Y_i(0, \\
&= \{Y_i(1, M_i(1)) - Y_i(0, M_i(1))\} + \{Y_i(0, M_i(1)) - Y_i(0,
\end{aligned}
$$

In the first equality, the total unit effect is decomposed into a unit "total indirect effect" with treatment $Z = 1$ and a unit "pure direct effect" of treatment at the mediator value $M_i(0)$ when $Z_i = 0$.

In the second equality, the total unit effect is decomposed into a "total direct effect" when the mediator takes value $M_i(1)$ when $Z_i = 0$ and a "pure indirect effect" when $Z_i = 0$. Notation follows Robins and Greenland (1992). Sometimes called "natural" direct and indirect effects following Pearl (2001).

When it is not necessary to distinguish between pure and total direct and indirect effects, we refer to direct and indirect effects.

REFRAMING THE PROBLEM OF MEDIATION

Two decompositions generally give different values for the direct and indirect components (Holland 1988, Robins and Greenland 1992) unless additivity (no interaction) holds:

$$Y_i(1, m) - Y_i(0, m) = Y_i(1, m^*) - Y_i(0, m^*)$$

for all (m, m^*).

Additivity is unlikely in most applications (Robins, 2003).

Additivity does hold when the unit direct effect is 0. Then Z is an "instrumental variable" as it affects the outcome only through M.

REFRAMING THE PROBLEM OF MEDIATION

From the decomposition:

$$
\begin{aligned}
Y_i(1, M_i(1)) - Y_i(0, M_i(0)) &= \{Y_i(1, M_i(1)) - Y_i(1, M_i(0))\} + \{Y_i(1, M_i(0)) - Y_i(0, \\
&= \{Y_i(1, M_i(1)) - Y_i(0, M_i(1))\} + \{Y_i(0, M_i(1)) - Y_i(0,
\end{aligned}
$$

average direct and indirect effects are readily obtained $E(Y_i(1, M_i(1)) - Y_i(1, M_i(0)))$ and $E(Y_i(1, M_i(0)) - Y_i(0, M_i(0)))$ or

$E(Y_i(1, M_i(1)) - Y_i(1, M_i(0)) \mid \mathbf{X}_i)$ and $E(Y_i(1, M_i(0)) - Y(0, M_i(0)) \mid \mathbf{X}_i)$.

REFRAMING THE PROBLEM OF MEDIATION

n.b. these estimands only make sense if the potential outcomes $Y_i(z, m)$ are well defined. There are contexts in which some outcomes are not reasonable.

e.g. in the encouragement study in the previous lesson, subject i is either assigned to receive or not receive encouragement. Then he/she chooses how much time to study ($M_i(0)$ or $M_i(1)$).

To consider outcomes $Y_i(z, m)$, the investigator must imagine a situation in which each subject might have studied m hours. This does not mean that the investigator must be able to design and/or implement such a study for every combination (z, m).

An investigator could argue that as the experiment did not manipulate the amount of time studied, and as it is not possible to force students to study m hours, consideration of outcomes $Y_i(z, m)$ is irrelevant.

REFRAMING THE PROBLEM OF MEDIATION

e.g. suppose an investigator wants to study the effect of education on earnings, as mediated through occupational choice. Let $Z_i = 0$ for respondent i, who at age 30, has less than a college education, and $Z_i = 1$ otherwise. Let M_i denote occupation at age 31, with m^* denoting "physician", and let Y_i denote earnings at age 32.

Clearly the outcome $Y_i(0, m^*)$, the earnings of a physician with less than a college degree, is not possible.

REFRAMING THE PROBLEM OF MEDIATION

Average direct and indirect effects have received more attention than average controlled direct effects.

Average direct and indirect effects indicate how the treatment operates in the population. Large values indicate the mediator is a significant channel through which Z operates. Small values suggest that either Z does not substantially impact the mediator and/or the mediator does not substantially impact the outcome.

In the former case, an investigator might look for a treatment that more effectively targets the mediator. If the effect on the mediator is substantively significant, it suggests the mediator does not impact the outcome substantially. The investigator might rethink the experiment and try a different treatment aimed at a different channel (n.b. Vanderweele, 2015).

REFRAMING THE PROBLEM OF MEDIATION

Some argue the average direct and indirect effects are of more fundamental scientific significance because they indicate how the treatment operates through the mediator, in contrast to controlled direct effects which fix the value of the mediator for all subjects.

While true of controlled direct effects of treatment, this is not the case for controlled direct effects of the mediator.

Controlled direct effects are the building blocks of the average direct and indirect effects. In theoretical contexts, it may be of more fundamental interest to know the effect of a mediator when its values are controlled than to know the effect of a treatment through mixtures of the mediating values.

REFRAMING THE PROBLEM OF MEDIATION

If it were possible to manipulate both z and m, one may want to know how the controlled direct effects vary with z and m. e.g. the value of $E(Y_i(z, m))$ that would yield maximum overall benefit.

Economists often consider variables Z which are hypothesized to affect Y only indirectly through a mediator of interest, M. While the mediator, M, does not behave as if randomly assigned, Z may have been randomly assigned or behave as if randomly assigned.

The controlled direct effect is of interest, but the indirect effect itself is not. Z is just a tool to ascertain the effect of M when the relationship between M and Y is confounded.