Supplementary Materials

for

Comparative Causal Mediation and Relaxing the Assumption of No Mediator-Outcome Confounding: An Application to International Law and Audience Costs

Kirk Bansak

Assistant Professor of Political Science, University of California, San Diego
Department of Political Science, La Jolla, CA 92093, USA
Email: kbansak@ucsd.edu

Appendix A: Formal Results

Proof of Proposition 1.

Given assumptions 1 and 2,

$$\kappa_j(t_j) = E[\alpha_{ji}(\beta_i + \gamma_{ji}t_j)] = E[\alpha_{ji}\beta_i] + E[\alpha_{ji}\gamma_{ji}t_j]$$

for $j = 1, 2$.

Given assumption 5,

$$E[\alpha_{ji}\beta_i] + E[\alpha_{ji}\gamma_{ji}t_j] = E[\alpha_{ji}]E[\beta_i] + E[\alpha_{ji}]E[\gamma_{ji}]t_j = \alpha_j(\beta + \gamma_j t_j)$$

for $j = 1, 2$.

Given assumption 4,

$$\alpha_j(\beta + \gamma_j t_j) = \alpha_j \beta$$

for $j = 1, 2$.

Thus,

$$\frac{\kappa_2(t_2)}{\kappa_1(t_1)} = \frac{\alpha_2\beta}{\alpha_1\beta}$$

and

$$\frac{\left(\frac{\kappa_2(t_2)}{\tau_2}\right)}{\left(\frac{\kappa_1(t_1)}{\tau_1}\right)} = \frac{(\frac{\alpha_2\beta}{\tau_2})}{(\frac{\alpha_1\beta}{\tau_1})}$$

Now, given assumption 3,

$$\begin{aligned}
E[\eta_i|T_{1i}, T_{2i}] &= E[\tilde{\pi}_i + \tilde{\alpha}_{1i}T_{1i} + \tilde{\alpha}_{2i}T_{2i}|T_{1i}, T_{2i}] \\
&= E[\tilde{\pi}_i|T_{1i}, T_{2i}] + E[\tilde{\alpha}_{1i}|T_{1i}, T_{2i}]T_{1i} + E[\tilde{\alpha}_{2i}|T_{1i}, T_{2i}]T_{2i} \\
&= E[\tilde{\pi}_i] + E[\tilde{\alpha}_{1i}]T_{1i} + E[\tilde{\alpha}_{2i}]T_{2i} \\
&= E[\pi_i - \pi] + E[\alpha_{1i} - \alpha_1]T_{1i} + E[\alpha_{2i} - \alpha_2]T_{2i} \\
&= 0
\end{aligned}$$

and

$$
\begin{aligned}
E[\rho_i|T_{1i},T_{2i}] &= E[\tilde{\chi}_i+\tilde{\tau}_{1i}T_{1i}+\tilde{\tau}_{2i}T_{2i}|T_{1i},T_{2i}] \\
&= E[\tilde{\chi}_i|T_{1i},T_{2i}]+E[\tilde{\tau}_{1i}|T_{1i},T_{2i}]T_{1i}+E[\tilde{\tau}_{2i}|T_{1i},T_{2i}]T_{2i} \\
&= E[\tilde{\chi}_i]+E[\tilde{\tau}_{1i}]T_{1i}+E[\tilde{\tau}_{2i}]T_{2i} \\
&= E[\chi_i-\chi]+E[\tau_{1i}-\tau_1]T_{1i}+E[\tau_{2i}-\tau_2]T_{2i} \\
&= 0
\end{aligned}
$$

Therefore, under standard regularity conditions for the generalized linear regression model,

$$
\begin{aligned}
\operatorname*{plim}_{N\to\infty}\hat{\alpha}_1^N &= \alpha_1 \\
\operatorname*{plim}_{N\to\infty}\hat{\alpha}_2^N &= \alpha_2 \\
\operatorname*{plim}_{N\to\infty}\hat{\tau}_1^N &= \tau_1 \\
\operatorname*{plim}_{N\to\infty}\hat{\tau}_2^N &= \tau_2
\end{aligned}
$$

Further, by Slutsky's theorem, and given non-zero parameters,

$$
\operatorname*{plim}_{N\to\infty}\left(\frac{\hat{\alpha}_2^N\hat{\beta}^N}{\hat{\alpha}_1^N\hat{\beta}^N}\right)=\operatorname*{plim}_{N\to\infty}\left(\frac{\hat{\alpha}_2^N}{\hat{\alpha}_1^N}\right)=\left(\operatorname*{plim}_{N\to\infty}\hat{\alpha}_2^N\right)\Big/\left(\operatorname*{plim}_{N\to\infty}\hat{\alpha}_1^N\right)=\frac{\alpha_2}{\alpha_1}=\frac{\alpha_2\beta}{\alpha_1\beta}=\frac{\kappa_2(t_2)}{\kappa_1(t_1)}
$$

And by the same argument

$$
\operatorname*{plim}_{N\to\infty}\left(\frac{\left(\frac{\hat{\alpha}_2^N\hat{\beta}^N}{\hat{\tau}_2^N}\right)}{\left(\frac{\hat{\alpha}_1^N\hat{\beta}^N}{\hat{\tau}_1^N}\right)}\right)=\frac{\left(\frac{\alpha_2\beta}{\tau_2}\right)}{\left(\frac{\alpha_1\beta}{\tau_1}\right)}=\frac{\left(\frac{\kappa_2(t_2)}{\tau_2}\right)}{\left(\frac{\kappa_1(t_1)}{\tau_1}\right)}
$$

Proof of Proposition 2.

Given assumptions 1 and 2,

$$\kappa_j(1) = E[\alpha_{ji}(\beta_i + \gamma_{ji})] = E[\alpha_{ji}\omega_{ji}]$$

for $j = 1, 2$.

Given assumption 5,

$$E[\alpha_{ji}\omega_{ji}] = E[\alpha_{ji}]E[\omega_{ji}] = \alpha_j\omega_j$$

for $j = 1, 2$.

Thus,

$$\frac{\kappa_2(1)}{\kappa_1(1)} = \frac{\alpha_2\omega_2}{\alpha_1\omega_1}$$

and

$$\frac{\left(\frac{\kappa_2(1)}{\tau_2}\right)}{\left(\frac{\kappa_1(1)}{\tau_1}\right)} = \frac{\left(\frac{\alpha_2\omega_2}{\tau_2}\right)}{\left(\frac{\alpha_1\omega_1}{\tau_1}\right)}$$

Now, given assumption 3, as in the proof of Proposition 1, under standard regularity conditions,

$$\operatorname*{plim}_{N\to\infty} \hat{\alpha}_1^N = \alpha_1$$

$$\operatorname*{plim}_{N\to\infty} \hat{\alpha}_2^N = \alpha_2$$

$$\operatorname*{plim}_{N\to\infty} \hat{\tau}_1^N = \tau_1$$

$$\operatorname*{plim}_{N\to\infty} \hat{\tau}_2^N = \tau_2$$

It will thus be the case that

$$\operatorname*{plim}_{N\to\infty} \frac{\hat{\alpha}_2^N\hat{\omega}_2^N}{\hat{\alpha}_1^N\hat{\omega}_1^N} < \frac{\alpha_2\omega_2}{\alpha_1\omega_1}$$

and

$$\operatorname*{plim}_{N\to\infty} \frac{\left(\frac{\hat{\alpha}_2^N\hat{\omega}_2^N}{\hat{\tau}_2^N}\right)}{\left(\frac{\hat{\alpha}_1^N\hat{\omega}_1^N}{\hat{\tau}_1^N}\right)} < \frac{\left(\frac{\alpha_2\omega_2}{\tau_2}\right)}{\left(\frac{\alpha_1\omega_1}{\tau_1}\right)}$$

if

$$\operatorname*{plim}_{N\to\infty} \frac{\hat{\omega}_2^N}{\hat{\omega}_1^N} < \frac{\omega_2}{\omega_1}$$

which is met when:

$$\frac{\omega_2 + \xi_2}{\omega_1 + \xi_1} < \frac{\omega_2}{\omega_1}$$

and hence when:

$$\omega_1 \xi_2 < \omega_2 \xi_1$$

□

Appendix B: Finite-Sample Adjustments

Finite-sample adjustments for the CCM estimators can be derived using Taylor series expansion.

Consider the first estimator under the no-interaction assumption, $\frac{\hat{\alpha}_2\hat{\beta}}{\hat{\alpha}_1\hat{\beta}}$, which can (quite apparently) be simplified to $\frac{\hat{\alpha}_2}{\hat{\alpha}_1}$. Similarly, the estimand of interest can be seen simply as:

$$\frac{ACME_2}{ACME_1} = \frac{\alpha_2\beta}{\alpha_1\beta} = \frac{\alpha_2}{\alpha_1} = \frac{E[\hat{\alpha}_2]}{E[\hat{\alpha}_1]}$$

However, a first problem is the following:

$$E\left[\frac{\hat{\alpha}_2}{\hat{\alpha}_1}\right] \neq \frac{E[\hat{\alpha}_2]}{E[\hat{\alpha}_1]} = \frac{\alpha_2}{\alpha_1}$$

A second problem is that $E\left[\frac{\hat{\alpha}_2}{\hat{\alpha}_1}\right]$ may not even exist. To address both of these problems, the estimator $\frac{\hat{\alpha}_2}{\hat{\alpha}_1}$, which will be denoted as $f(\hat{\Theta})$ can be approximated using a (second-order) multivariate Taylor series expansion around the estimand $f(\Theta)$:

$$f(\hat{\Theta}) \approx f(\Theta) + \sum_{\theta\in\Theta}(\hat{\theta}-\theta)f_{\hat{\theta}}(\Theta) + \frac{1}{2}\sum_{\theta\in\Theta}\sum_{\theta'\in\Theta}(\hat{\theta}-\theta)(\hat{\theta}'-\theta')f_{\hat{\theta}\hat{\theta}'}(\Theta)$$

where Θ contains the full set of parameters (denoted individually by θ), $f_{\hat{\theta}}$ refers to the first derivative of f with respect to $\hat{\theta}$, and $f_{\hat{\theta}\hat{\theta}'}$ refers to the second derivative of f with respect to $\hat{\theta}$ and $\hat{\theta}'$.

If we treat the higher-order terms in the Taylor series expansion as negligible, as conventionally done, then we can identify the approximate divergence between the estimator and the estimand, which is a quantity for which we can characterize the moments:

$$E\left[\sum_{\theta\in\Theta}(\hat{\theta}-\theta)f_{\hat{\theta}}(\Theta)+\frac{1}{2}\sum_{\theta\in\Theta}\sum_{\theta'\in\Theta}(\hat{\theta}-\theta)(\hat{\theta}'-\theta')f_{\hat{\theta}\hat{\theta}'}(\Theta)\right]$$

The first-order terms in this expression are zero in expectation (i.e. $E[\hat{\theta}-\theta]=0$), while the leading components of the second-order terms are covariances in expectation (i.e. $E[(\hat{\theta}-\theta)(\hat{\theta}'-\theta')]=Cov(\hat{\theta},\hat{\theta}')$). Thus, the divergence is approximately:

$$\frac{1}{2}\sum_{\theta\in\Theta}\sum_{\theta'\in\Theta}Cov(\hat{\theta},\hat{\theta}')f_{\hat{\theta}\hat{\theta}'}(\Theta)$$

This divergence can thus be estimated—by plugging in $\hat{\Theta}$ for Θ and estimating the covariances—and then subtracted from the simple estimator $f(\hat{\Theta})$ of interest to yield an adjusted estimator that is approximately centered on the estimand of interest. Also evident from the expression is that this divergence term goes to zero as the sample size n grows to infinity. The following applies this process to the actual estimators in question.

A. Adjusted Estimators under the No-Interaction Assumption

A.1. Adjusted Estimator 1

The estimator for the first estimand is $\frac{\hat{\alpha}_2\hat{\beta}}{\hat{\alpha}_1\hat{\beta}}=\frac{\hat{\alpha}_2}{\hat{\alpha}_1}$. In expectation, the second-order Taylor Series expansion of the estimator, $T(\frac{\hat{\alpha}_2}{\hat{\alpha}_1})$, around the estimand is:

$$E\left[T\left(\frac{\hat{\alpha}_2}{\hat{\alpha}_1}\right)\right]\approx\frac{\alpha_2}{\alpha_1}-\frac{Cov(\hat{\alpha}_1,\hat{\alpha}_2)}{\alpha_1^2}+\frac{Var(\hat{\alpha}_1)\alpha_2}{\alpha_1^3}$$

Hence, we can identify the component of the approximation that diverges from the estimand. Because of the exogeneity of T, α_1 and α_2 can both be estimated without bias, allowing for the individual pieces of that component to be estimated by regression. This can then be subtracted from the estimator $\frac{\hat{\alpha}_2}{\hat{\alpha}_1}$ to yield an adjusted estimator approximately centered on the estimand:

$$\frac{\hat{\alpha}_2}{\hat{\alpha}_1} + \frac{\widehat{Cov}(\hat{\alpha}_1, \hat{\alpha}_2)}{\hat{\alpha}_1^2} - \frac{\widehat{Var}(\hat{\alpha}_1)\hat{\alpha}_2}{\hat{\alpha}_1^3}$$

In the special case of balanced control and treatment assignment (i.e. $P(C) = P(T_1) = P(T_2) = \frac{1}{3}$), the adjusted estimator simplifies to:

$$\frac{\hat{\alpha}_2}{\hat{\alpha}_1} + \frac{3\hat{\sigma}_\eta^2}{\hat{\alpha}_1^2 N} - \frac{6\hat{\sigma}_\eta^2 \hat{\alpha}_2}{\hat{\alpha}_1^3 N}$$

where $\hat{\sigma}_\eta^2$ refers to the estimated error variance from equation (5). Clearly, as N grows to infinity, this converges on the simple estimator $\frac{\hat{\alpha}_2}{\hat{\alpha}_1}$.

A.2. Adjusted Estimator 2

The simple estimator for the second estimand is $(\frac{\hat{\alpha}_2\hat{\beta}}{\hat{\tau}_2})/(\frac{\hat{\alpha}_1\hat{\beta}}{\hat{\tau}_1}) = (\frac{\hat{\alpha}_2}{\hat{\tau}_2})/(\frac{\hat{\alpha}_1}{\hat{\tau}_1}) = \frac{\hat{\alpha}_2\hat{\tau}_1}{\hat{\alpha}_1\hat{\tau}_2}$. As above, a second-order Taylor Series expansion can be used to formulate an adjusted estimator that is approximately centered on the estimand in finite samples:

$$\begin{aligned}
&\frac{\hat{\alpha}_2\hat{\tau}_1}{\hat{\alpha}_1\hat{\tau}_2} - \widehat{Var}(\hat{\alpha}_1)\frac{\hat{\alpha}_2\hat{\tau}_1}{\hat{\alpha}_1^3\hat{\tau}_2} - \widehat{Var}(\hat{\tau}_2)\frac{\hat{\alpha}_2\hat{\tau}_1}{\hat{\alpha}_1\hat{\tau}_2^3} + \widehat{Cov}(\hat{\alpha}_2, \hat{\alpha}_1)\frac{\hat{\tau}_1}{\hat{\alpha}_1^2\hat{\tau}_2} + \widehat{Cov}(\hat{\alpha}_2, \hat{\tau}_2)\frac{\hat{\tau}_1}{\hat{\alpha}_1\hat{\tau}_2^2} \\
&-\widehat{Cov}(\hat{\alpha}_2, \hat{\tau}_1)\frac{1}{\hat{\alpha}_1\hat{\tau}_2} - \widehat{Cov}(\hat{\alpha}_1, \hat{\tau}_2)\frac{\hat{\alpha}_2\hat{\tau}_1}{\hat{\alpha}_1^2\hat{\tau}_2^2} + \widehat{Cov}(\hat{\alpha}_1, \hat{\tau}_1)\frac{\hat{\alpha}_2}{\hat{\alpha}_1^2\hat{\tau}_2} + \widehat{Cov}(\hat{\tau}_2, \hat{\tau}_1)\frac{\hat{\alpha}_2}{\hat{\alpha}_1\hat{\tau}_2^2}
\end{aligned}$$

B. Adjusted Estimators when Relaxing the No-Interaction Assumption

Having discarded the no-interaction assumption, the estimator of the first estimand of interest, $\frac{\kappa_2(1)}{\kappa_1(1)}$, is $\frac{\hat{\alpha}_2(\hat{\beta}+\hat{\gamma}_2)}{\hat{\alpha}_1(\hat{\beta}+\hat{\gamma}_1)} = \frac{\hat{\alpha}_2\hat{\omega}_2}{\hat{\alpha}_1\hat{\omega}_1}$.

As shown, $\text{plim}_{N\to\infty}\ \hat{\omega}_j = \omega_j + \xi_j$, because of a confounding bias that does not disappear asymptotically. Let ω_j^* denote the biased and inconsistent version of ω_j (i.e. $\text{plim}_{N\to\infty}\ \hat{\omega}_j = \omega_j^*$). As shown above, under certain reasonable and testable assumptions, $\frac{\alpha_2\omega_2^*}{\alpha_1\omega_1^*}$ is conservative (i.e. attenuated toward 1) for $\frac{\alpha_2\omega_2}{\alpha_1\omega_1}$ and hence the estimator of interest is asymptotically conservative for the estimand of interest. Unfortunately, for two reasons, this does not mean that in small samples the estimator of interest is in expectation also conservative. First, as before, the expectation may not even actually exist. Second, also as before, the ratio form of the estimand leads the estimator to be decentered from the point to which it converges. However, also as in the case with the no-interaction assumption, a second-order Taylor Series expansion can be used to construct an adjusted estimator that in finite samples is approximately centered upon the conservative point for which the estimator is consistent.

Specifically, the adjusted estimator is:

$$\frac{\hat{\alpha}_2\hat{\omega}_2}{\hat{\alpha}_1\hat{\omega}_1} - \widehat{Var}(\hat{\alpha}_1)\frac{\hat{\alpha}_2\hat{\omega}_2}{\hat{\alpha}_1^3\hat{\omega}_1} - \widehat{Var}(\hat{\omega}_1)\frac{\hat{\alpha}_2\hat{\omega}_2}{\hat{\alpha}_1\hat{\omega}_1^3} + \widehat{Cov}(\hat{\alpha}_2,\hat{\alpha}_1)\frac{\hat{\omega}_2}{\hat{\alpha}_1^2\hat{\omega}_1} + \widehat{Cov}(\hat{\alpha}_2,\hat{\omega}_1)\frac{\hat{\omega}_2}{\hat{\alpha}_1\hat{\omega}_1^2}$$
$$-\widehat{Cov}(\hat{\alpha}_2,\hat{\omega}_2)\frac{1}{\hat{\alpha}_1\hat{\omega}_1} - \widehat{Cov}(\hat{\alpha}_1,\hat{\omega}_1)\frac{\hat{\alpha}_2\hat{\omega}_2}{\hat{\alpha}_1^2\hat{\omega}_1^2} + \widehat{Cov}(\hat{\alpha}_1,\hat{\omega}_2)\frac{\hat{\alpha}_2}{\hat{\alpha}_1^2\hat{\omega}_1} + \widehat{Cov}(\hat{\omega}_1,\hat{\omega}_2)\frac{\hat{\alpha}_2}{\hat{\alpha}_1\hat{\omega}_1^2}$$

where $\hat{\omega}_j = \hat{\beta} + \hat{\gamma}_j$ from Equation 6 and covariance terms can be estimated via the bootstrap.

Following the same approach for the second CCM estimand, the adjusted version of the second estimator, $\left(\frac{\hat{\alpha}_2\hat{\omega}_2}{\hat{\tau}_2}\right) / \left(\frac{\hat{\alpha}_1\hat{\omega}_1}{\hat{\tau}_1}\right)$, is:

$$\begin{aligned}
&\frac{(\frac{\hat{\alpha}_2\hat{\omega}_2}{\hat{\tau}_2})}{(\frac{\hat{\alpha}_1\hat{\omega}_1}{\hat{\tau}_1})} - \widehat{Var}(\hat{\alpha}_1)\frac{\hat{\alpha}_2\hat{\omega}_2\hat{\tau}_1}{\hat{\alpha}_1^3\hat{\omega}_1\hat{\tau}_2} - \widehat{Var}(\hat{\omega}_1)\frac{\hat{\alpha}_2\hat{\omega}_2\hat{\tau}_1}{\hat{\alpha}_1\hat{\omega}_1^3\hat{\tau}_2} - \widehat{Var}(\hat{\tau}_2)\frac{\hat{\alpha}_2\hat{\omega}_2\hat{\tau}_1}{\hat{\alpha}_1\hat{\omega}_1\hat{\tau}_2^3}\\
&+\widehat{Cov}(\hat{\alpha}_2,\hat{\alpha}_1)\frac{\hat{\omega}_2\hat{\tau}_1}{\hat{\alpha}_1^2\hat{\omega}_1\hat{\tau}_2} - \widehat{Cov}(\hat{\alpha}_2,\hat{\omega}_2)\frac{\hat{\tau}_1}{\hat{\alpha}_1\hat{\omega}_1\hat{\tau}_2} + \widehat{Cov}(\hat{\alpha}_2,\hat{\omega}_1)\frac{\hat{\omega}_2\hat{\tau}_1}{\hat{\alpha}_1\hat{\omega}_1^2\hat{\tau}_2}\\
&+\widehat{Cov}(\hat{\alpha}_2,\hat{\tau}_2)\frac{\hat{\omega}_2\hat{\tau}_1}{\hat{\alpha}_1\hat{\omega}_1\hat{\tau}_2^2} - \widehat{Cov}(\hat{\alpha}_2,\hat{\tau}_1)\frac{\hat{\omega}_2}{\hat{\alpha}_1\hat{\omega}_1\hat{\tau}_2} + \widehat{Cov}(\hat{\alpha}_1,\hat{\omega}_2)\frac{\hat{\alpha}_2\hat{\tau}_1}{\hat{\alpha}_1^2\hat{\omega}_1\hat{\tau}_2}\\
&-\widehat{Cov}(\hat{\alpha}_1,\hat{\omega}_1)\frac{\hat{\alpha}_2\hat{\omega}_2\hat{\tau}_1}{\hat{\alpha}_1^2\hat{\omega}_1^2\hat{\tau}_2} - \widehat{Cov}(\hat{\alpha}_1,\hat{\tau}_2)\frac{\hat{\alpha}_2\hat{\omega}_2\hat{\tau}_1}{\hat{\alpha}_1^2\hat{\omega}_1\hat{\tau}_2^2} + \widehat{Cov}(\hat{\alpha}_1,\hat{\tau}_1)\frac{\hat{\alpha}_2\hat{\omega}_2}{\hat{\alpha}_1^2\hat{\omega}_1\hat{\tau}_2}\\
&+\widehat{Cov}(\hat{\omega}_2,\hat{\omega}_1)\frac{\hat{\alpha}_2\hat{\tau}_1}{\hat{\alpha}_1\hat{\omega}_1^2\hat{\tau}_2} + \widehat{Cov}(\hat{\omega}_2,\hat{\tau}_2)\frac{\hat{\alpha}_2\hat{\tau}_1}{\hat{\alpha}_1\hat{\omega}_1\hat{\tau}_2^2} - \widehat{Cov}(\hat{\omega}_2,\hat{\tau}_1)\frac{\hat{\alpha}_2}{\hat{\alpha}_1\hat{\omega}_1\hat{\tau}_2}\\
&-\widehat{Cov}(\hat{\omega}_1,\hat{\tau}_2)\frac{\hat{\alpha}_2\hat{\omega}_2\hat{\tau}_1}{\hat{\alpha}_1\hat{\omega}_1^2\hat{\tau}_2^2} + \widehat{Cov}(\hat{\omega}_1,\hat{\tau}_1)\frac{\hat{\alpha}_2\hat{\omega}_2}{\hat{\alpha}_1\hat{\omega}_1^2\hat{\tau}_2} + \widehat{Cov}(\hat{\tau}_2,\hat{\tau}_1)\frac{\hat{\alpha}_2\hat{\omega}_2}{\hat{\alpha}_1\hat{\omega}_1\hat{\tau}_2^2}
\end{aligned}$$

In sum, if the assumption of no interaction between the treatments and the mediator is relaxed, the CCM estimators are no longer consistent, but they are asymptotically conservative provided additional conditions are met. Those additional conditions are both theoretically reasonable and empirically testable. Furthermore, finite-sample adjustments can be added to the estimators such that they are also conservative in smaller samples.

Appendix C: Tests and Sensitivity Analysis for the Conservatism of Estimators with Interactions

As explained in the main text, given the conditions described in Proposition 2, the bias involved in estimating $\frac{\kappa_2(1)}{\kappa_1(1)}$ and $\left(\frac{\kappa_2(1)}{\tau_2}\right) \Big/ \left(\frac{\kappa_1(1)}{\tau_1}\right)$ results in conservative (attenuated toward 1) estimates of these estimands. While assumption 4 (no interaction between the treatments and mediator) was relaxed, Proposition 2 introduces the following additional condition that was not present in Proposition 1: $\omega_2\xi_1 > \omega_1\xi_2$. This appendix shows how this condition can be partially assessed empirically.

Recall the semi-parametric model:

$$M_i = \pi + \alpha_1 T_{1i} + \alpha_2 T_{2i} + \eta_i \tag{5}$$

$$Y_i = \lambda + \delta_1 T_{1i} + \delta_2 T_{2i} + \beta M_i + \gamma_1 T_{1i} M_i + \gamma_2 T_{2i} M_i + \iota_i \tag{6}$$

$$Y_i = \chi + \tau_1 T_{1i} + \tau_2 T_{2i} + \rho_i \tag{7}$$

Now, consider equations 5 and 6 in the model by treatment subsets:

$$(M_i | T_{1i} = 1, T_{2i} = 0) = \pi + \alpha_1 + \eta_i \tag{8}$$

$$(Y_i | T_{1i} = 1, T_{2i} = 0) = \lambda + \delta_1 + \omega_1 M_i + \iota_i \tag{9}$$

$$(M_i | T_{1i} = 0, T_{2i} = 1) = \pi + \alpha_2 + \eta_i \tag{10}$$

$$(Y_i | T_{1i} = 0, T_{2i} = 1) = \lambda + \delta_2 + \omega_2 M_i + \iota_i \tag{11}$$

where $\omega_1 = \beta + \gamma_1$ and $\omega_2 = \beta + \gamma_2$. Given the saturation of the model presented in equations 5 and 6, estimation of the parameters via linear least squares regression would yield identical results if applied to equations 5 and 6 or the subsetted equations.

Consider estimation of ω_1 and ω_2 via linear least squares regression as applied to subsetted equations 9 and 11. For both cases, $j = 1, 2$, this is a bivariate regression, and thus:

$$\begin{aligned}
\operatorname*{plim}_{N\to\infty} \hat{\omega}_j &= \frac{Cov(Y_i, M_i|T_{ij} = 1, T_{ij'} = 0)}{Var(M_i|T_{ij} = 1, T_{ij'} = 0)} = \frac{Cov(\lambda + \delta_j + \omega_j M_i + \iota_i, M_i|T_{ij} = 1, T_{ij'} = 0)}{Var(M_i|T_{ij} = 1, T_{ij'} = 0)} \\
&= \frac{\omega_j Cov(M_i, M_i|T_{ij} = 1, T_{ij'} = 0) + Cov(\iota_i, M_i|T_{ij} = 1, T_{ij'} = 0)}{Var(M_i|T_{ij} = 1, T_{ij'} = 0)} \\
&= \omega_j + \frac{Cov(\iota_i, \eta_i|T_{ij} = 1, T_{ij'} = 0)}{Var(\eta_i|T_{ij} = 1, T_{ij'} = 0)}
\end{aligned}$$

That is,

$$\operatorname*{plim}_{N\to\infty} \hat{\omega}_1 = \omega_1 + \xi_1 = \omega_1 + \frac{Cov(\iota_i, \eta_i|T_{i1} = 1, T_{i2} = 0)}{Var(\eta_i|T_{i1} = 1, T_{i2} = 0)}$$

$$\operatorname*{plim}_{N\to\infty} \hat{\omega}_2 = \omega_2 + \xi_2 = \omega_2 + \frac{Cov(\iota_i, \eta_i|T_{i1} = 0, T_{i2} = 1)}{Var(\eta_i|T_{i1} = 0, T_{i2} = 1)}$$

Now, consider that: $\omega_2\xi_1 > \omega_1\xi_2$ implies that

$$\left(\operatorname*{plim}_{N\to\infty} \hat{\omega}_2 - \xi_2\right)\xi_1 > \left(\operatorname*{plim}_{N\to\infty} \hat{\omega}_1 - \xi_1\right)\xi_2$$

$$\left(\operatorname*{plim}_{N\to\infty} \hat{\omega}_2\right)\xi_1 > \left(\operatorname*{plim}_{N\to\infty} \hat{\omega}_1\right)\xi_2$$

$$\left(\operatorname*{plim}_{N\to\infty} \hat{\omega}_2\right)\frac{Cov(\iota_i, \eta_i|T_{i1} = 1, T_{i2} = 0)}{Var(\eta_i|T_{i1} = 1, T_{i2} = 0)} > \left(\operatorname*{plim}_{N\to\infty} \hat{\omega}_1\right)\frac{Cov(\iota_i, \eta_i|T_{i1} = 0, T_{i2} = 1)}{Var(\eta_i|T_{i1} = 0, T_{i2} = 1)}$$

Unfortunately, the possibility of unobserved confounding given non-randomization of the mediator makes it impossible to reliably estimate or compare $Cov(\iota_i, \eta_i|T_{ij} = 1, T_{ij'} = 0)$ for $i = 1, 2$ without additional assumptions. However, in large samples, $\operatorname{plim}_{N\to\infty} \hat{\omega}_j$ can be approximated by $\hat{\omega}_j$ and $Var(\eta_i|T_{ij} = 1, T_{ij'} = 0)$ can be approximated by $\widehat{Var}(\eta_i|T_{ij} =$

$1, T_{ij'} = 0) = \widehat{Var}(M_i | T_{ij} = 1, T_{ij'} = 0) = \hat{\sigma}^2_{\eta_j}$ using the observed data.

Hence,

$$\left(\operatorname*{plim}_{N\to\infty} \hat{\omega}_2\right) \frac{Cov(\iota_i, \eta_i | T_{i1} = 1, T_{i2} = 0)}{Var(\eta_i | T_{i1} = 1, T_{i2} = 0)} > \left(\operatorname*{plim}_{N\to\infty} \hat{\omega}_1\right) \frac{Cov(\iota_i, \eta_i | T_{i1} = 0, T_{i2} = 1)}{Var(\eta_i | T_{i1} = 0, T_{i2} = 1)}$$

can be partially assessed via:

$$\hat{\omega}_2 \hat{\sigma}^2_{\eta_2} > \hat{\omega}_1 \hat{\sigma}^2_{\eta_1}$$

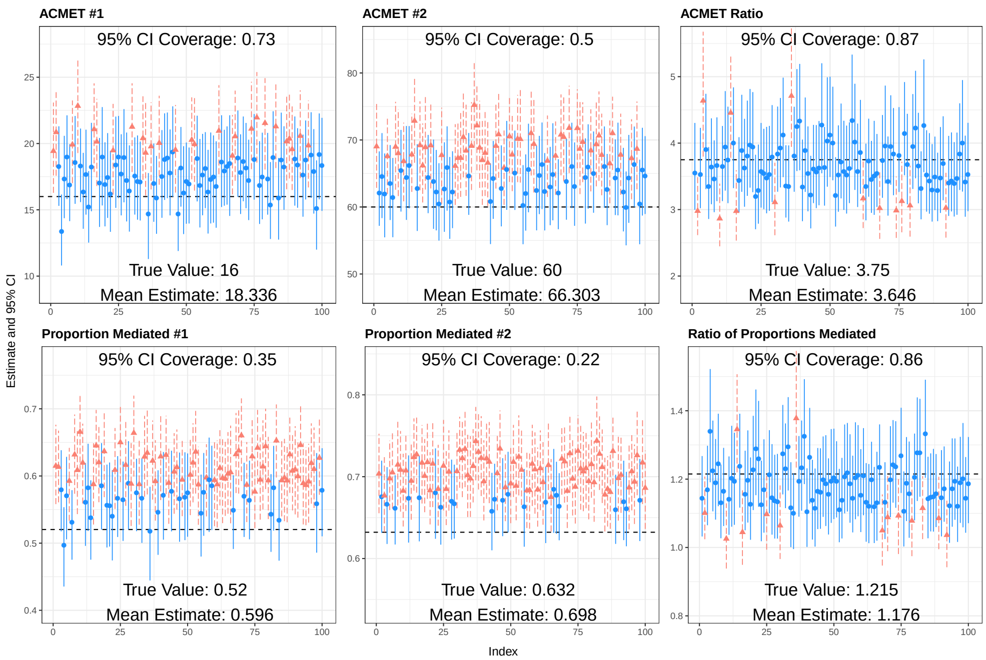
Appendix D: Simulations when Relaxing the No-Interaction Assumption

To illustrate the properties of the CCM estimators once the no-interaction assumption has been relaxed, this section presents the results of a simulation. The data-generating process was similar to that of the simulation presented earlier except, in this case, the effect of the mediator on the outcome involves interactions with both treatments. In addition, the simulated sample size has been increased to 1000 units per treatment condition in order to better illustrate the asymptotic tendencies.[21] As before, positive bias is introduced by construction through the omission in the estimation of a confounder that affects both the outcome and mediator. Also as before, the ACME for the treated for the second treatment is larger than that of the first treatment; further, the interaction between the mediator and the second treatment is also made larger than the interaction between the mediator and the first treatment. Thus, the additional conditions required for conservative estimation of the CCM estimands are met. Figure D1 shows the resulting estimates in the simulation.

As can be seen in the top row of Figure D1, the estimators of the ACMEs for the treated are again biased upward and, as a result, also have bad confidence-interval coverage. In contrast, however, the estimator of the ratio of ACMEs for the treated is much more well-behaved. While no longer consistent, and hence not properly centered in this medium-sized sample, the estimator is conservative (attenuated toward unity), as indicated by the mean estimate being closer to one than the true value. As a result of this conservatism, there is unfortunately confidence-interval under-coverage. However, what makes this problem less

[21] For this reason, the finite-sample adjustments make little difference, and hence the adjusted estimators are not presented here.

Figure D1: Comparative Causal Mediation Simulation, With Interactions



concerning is that the under-coverage is the result of attenuated estimates, as shown by the majority of bad confidence intervals being below the true value, rather than the result of systematically undersized confidence intervals.

The results are similar for the bottom row of Figure D1, which presents the estimates for the proportions mediated, as well as the ratio of the proportions mediated. Again, the traditional estimators are biased upward, while the CCM estimator is conservative.

Appendix E: Application Text

Prologue

Please consider the following hypothetical scenario:

ISIS militants in Iraq were threatening rocket attacks on neighboring countries in the region. In response, the U.S. government considered taking military action. The U.S. ruled out drone strikes and other options because the ISIS militants were hiding in a civilian zone, and the U.S. government wanted to avoid harming civilians. Instead, U.S. commandos were deployed in a covert operation. In order to avoid inflicting permanent harm on nearby civilians, the commandos used a non-lethal "incapacitating" chemical gas to knock out and capture the ISIS militants. However, critics of the operation have pointed out that people have varying levels of sensitivity to the incapacitating gas, and exposure can be fatal for some people. Hence, the operation may have put civilian lives in harm's way.

Treatment

CONTROL (*no additional information provided*)

OR

INFORMAL TREATMENT: Furthermore, the U.S. government has pledged never to use incapacitating chemical gas in previous public statements. Hence, the U.S. government has broken its pledge.

OR

LEGAL TREATMENT: Furthermore, the U.S. government has pledged never to use incapacitating chemical gas under its membership in the Chemical Weapons Convention, the international treaty banning chemical weapons. Hence, the U.S. government has broken international law.

DV 1: Disapproval

In general, do you approve or disapprove of the U.S. government's decision to use the incapacitating gas in the operation?

- *Approve Strongly, Approve, Neither Approve nor Disapprove, Disapprove, Disapprove Strongly*
- Variable is dichotomized for analysis, with 1 indicating "Disapprove" or "Disapprove Strongly," and 0 otherwise.

DV 2: Punishment

Imagine that one of your U.S. Senators voted in favor of using the incapacitating gas. Would this increase or decrease your willingness to vote for that Senator in the next election?

- *Increase Greatly, Increase, Neither Increase nor Decrease, Decrease, Decrease Greatly*
- Variable is dichotomized for analysis, with 1 indicating "Decrease" or "Decrease Greatly," and 0 otherwise.

Mediator: Perceived Immorality

To what extent do you believe that the decision to use the incapacitating gas in the operation was morally right or wrong?

- *Definitely Right, Probably Right, Not Morally Right or Wrong, Probably Wrong, Definitely Wrong*
- Variable is dichotomized for analysis, with 1 indicating "Probably Wrong" or "Definitely Wrong" and 0 otherwise.

Mediator: Expected Harm

To what extent do you agree with the following statement: The decision to use the incapacitating gas will harm U.S. security in the long-run by encouraging our adversaries to acquire and use such weapons in the future.

- *Agree Strongly, Agree, Neither Agree nor Disagree, Disagree, Disagree Strongly*
- Variable is dichotomized for analysis, with 1 indicating "Agree" or "Agree Strongly," and 0 otherwise.

Appendix F: Application Demographics and Balance

Table F1: Overall Sample Demographics

Gender

Female	Male
46.3%	53.7%

Age

18-29	30-44	45-64	65+
38.9%	41.1%	18.5%	1.6%

Education

No High School	High School	Some College	College Graduate
0.9%	11.9%	34.1%	53.1%

Table F2: Sample Demographics by Treatment Condition

	Gender	
	Female	Male
Control	48.3%	51.7%
Informal Treatment	42.2%	57.8%
Legal Treatment	48.2%	51.8%

χ^2 test p-value: 0.074

	Age			
	18-29	30-44	45-64	65+
Control	37.0%	40.2%	20.7%	2.1%
Informal Treatment	40.9%	40.6%	17.8%	0.7%
Legal Treatment	38.7%	42.4%	17.0%	1.9%

χ^2 test p-value: 0.316

	Education			
	No High School	High School	Some College	College Graduate
Control	1.1%	10.9%	31.4%	56.6%
Informal Treatment	0.7%	12.7%	36.3%	50.3%
Legal Treatment	0.7%	12.0%	34.8%	52.5%

χ^2 test p-value: 0.503

Note: The χ^2 tests are contingency table tests of the independence between the treatment assignment and each covariate.

Appendix G: Additional Application Analysis

The main text of this study presents evidence that legalization has the potential to enhance audience costs by affecting voters' normative perceptions of a policy issue, with violations of foreign policy pledges being perceived as more morally objectionable when they have legal status. However, another channel through which legalization could increase audience costs is by affecting voters' consequentialist perceptions of the issue. For instance, voters may be more likely to fear international repercussions in response to a foreign policy commitment violation if that commitment has international legal status. The application presented in this study also tested one such consequentialist mechanism, namely the fear that other countries would follow suit and hence harm U.S. interests. Specifically, respondents were asked to what extent they believed the decision to use the chemical incapacitants would harm U.S. security in the long-run by encouraging adversaries to acquire and use such weapons in the future. This mediator was measured on a five-point scale in the survey (see Appendix E), and it is dichotomized to facilitate interpretation in the analysis presented here. The binary version of the mediator captures whether or not each respondent believed the policy decision would harm U.S. security, called Expected Harm here.

The results of applying the comparative causal mediation analysis to this mediator are displayed in Table G1. Similar to the Perceived Immorality mediator, estimates of the ratio of mediation effects for the Expected Harm mediator are substantively large and statistically distinguishable from 1 for both dependent variables, while the estimates of the ratios of proportions mediated are not statistically distinguishable from 1. These results suggest that the Expected Harm mediator also plays an important role in the enhancement of audience costs by legalization, though does not increase as a proportion of the total audience costs effect given legalization.

In addition, Tables G2, G3, and G4 display all results—average treatment effects and comparative causal mediation estimates for both dependent variables and both mediators—when analyzing the dependent variables and mediators on their raw five-point scale. While on a different scale, the results remain substantively and statistically unchanged.

Table G1: Comparative Causal Mediation via Expected Harm Mechanism, Using Binary Mediator and Dependent Variables

DV: Disapproval

	$\widehat{ACME}_1$ Mediation Effect for Informal Treatment	$\widehat{ACME}_2$ Mediation Effect for Legal Treatment	$\frac{\widehat{ACME}_2}{\widehat{ACME}_1}$ **Ratio of Mediation Effects**	$\left(\frac{\widehat{ACME}_2}{\widehat{ATE}_2}\right) \Big/ \left(\frac{\widehat{ACME}_1}{\widehat{ATE}_1}\right)$ **Ratio of Proportions Mediated**
Estimate	0.058	0.118	**2.041**	**1.243**
95% CI	[0.033, 0.084]	[0.091, 0.148]	**[1.450, 3.364]**	**[0.882, 1.942]**

DV: Punishment

	$\widehat{ACME}_1$ Mediation Effect for Informal Treatment	$\widehat{ACME}_2$ Mediation Effect for Legal Treatment	$\frac{\widehat{ACME}_2}{\widehat{ACME}_1}$ **Ratio of Mediation Effects**	$\left(\frac{\widehat{ACME}_2}{\widehat{ATE}_2}\right) \Big/ \left(\frac{\widehat{ACME}_1}{\widehat{ATE}_1}\right)$ **Ratio of Proportions Mediated**
Estimate	0.050	0.102	**2.041**	**1.322**
95% CI	[0.028, 0.073]	[0.077, 0.128]	**[1.450, 3.364]**	**[0.915, 2.082]**

Table G2: Sample Estimates of ATEs, Using 5-Point Dependent Variables

DV: Disapproval

	$\widehat{ATE}_1$ Informal treatment effect	$\widehat{ATE}_2$ Legal treatment effect	$\widehat{ATE}_2 - \widehat{ATE}_1$ Difference in treatment effects
Estimate	0.477	0.799	0.321
95% CI	[0.338, 0.614]	[0.659, 0.938]	[0.177, 0.466]

DV: Punishment

	$\widehat{ATE}_1$ Informal treatment effect	$\widehat{ATE}_2$ Legal treatment effect	$\widehat{ATE}_2 - \widehat{ATE}_1$ Difference in treatment effects
Estimate	0.301	0.529	0.228
95% CI	[0.192, 0.411]	[0.416, 0.646]	[0.113, 0.343]

Table G3: Comparative Causal Mediation via Perceived Immorality Mechanism, Using 5-Point Mediator and Dependent Variables

	DV: Disapproval			
	$\widehat{ACME}_1$	$\widehat{ACME}_2$	$\frac{\widehat{ACME}_2}{\widehat{ACME}_1}$	$\left(\frac{\widehat{ACME}_2}{\widehat{ATE}_2}\right) \Big/ \left(\frac{\widehat{ACME}_1}{\widehat{ATE}_1}\right)$
	Mediation Effect for Informal Treatment	Mediation Effect for Legal Treatment	**Ratio of Mediation Effects**	**Ratio of Proportions Mediated**
Estimate	0.295	0.501	**1.697**	**1.014**
95% CI	[0.192, 0.397]	[0.397, 0.607]	**[1.286, 2.460]**	**[0.822, 1.279]**
	DV: Punishment			
	$\widehat{ACME}_1$	$\widehat{ACME}_2$	$\frac{\widehat{ACME}_2}{\widehat{ACME}_1}$	$\left(\frac{\widehat{ACME}_2}{\widehat{ATE}_2}\right) \Big/ \left(\frac{\widehat{ACME}_1}{\widehat{ATE}_1}\right)$
	Mediation Effect for Informal Treatment	Mediation Effect for Legal Treatment	**Ratio of Mediation Effects**	**Ratio of Proportions Mediated**
Estimate	0.214	0.364	**1.697**	**0.965**
95% CI	[0.140, 0.289]	[0.288, 0.443]	**[1.286, 2.460]**	**[0.717, 1.266]**

Table G4: Comparative Causal Mediation via Expected Harm Mechanism, Using 5-Point Mediator and Dependent Variables

DV: Disapproval

	$\widehat{ACME}_1$ Mediation Effect for Informal Treatment	$\widehat{ACME}_2$ Mediation Effect for Legal Treatment	$\frac{\widehat{ACME}_2}{\widehat{ACME}_1}$ **Ratio of Mediation Effects**	$\left(\frac{\widehat{ACME}_2}{\widehat{ATE}_2}\right) \Big/ \left(\frac{\widehat{ACME}_1}{\widehat{ATE}_1}\right)$ **Ratio of Proportions Mediated**
Estimate	0.211	0.411	**1.949**	**1.165**
95% CI	[0.135, 0.289]	[0.332, 0.493]	**[1.487, 2.843]**	**[0.880, 1.601]**

DV: Punishment

	$\widehat{ACME}_1$ Mediation Effect for Informal Treatment	$\widehat{ACME}_2$ Mediation Effect for Legal Treatment	$\frac{\widehat{ACME}_2}{\widehat{ACME}_1}$ **Ratio of Mediation Effects**	$\left(\frac{\widehat{ACME}_2}{\widehat{ATE}_2}\right) \Big/ \left(\frac{\widehat{ACME}_1}{\widehat{ATE}_1}\right)$ **Ratio of Proportions Mediated**
Estimate	0.137	0.268	**1.949**	**1.109**
95% CI	[0.087, 0.190]	[0.213, 0.325]	**[1.487, 2.843]**	**[0.770, 1.589]**

Appendix H: Choosing a CCM Estimand

Tables H1 and H2 summarize the general research questions and theoretical implications related to each CCM estimand. Which of the two estimands is of interest will depend upon the empirical and theoretical goals of a particular research project. When the researcher's main goal is to identify which treatment has the strongest absolute effect transmitted via a specific causal channel, the first estimand is likely to be of primary interest. The case of evaluating different job training programs, as presented in the main text, provides an example. From the standpoint of optimal policy implementation, the researcher may choose to focus on one specific causal channel, prioritizing transmission of the causal effect via that channel and discounting transmission via other channels. For instance, if the researcher knows that the training programs under consideration will, in the post-evaluation period, be rolled out in target areas where increasing job-search motivation is unlikely to be an effective method of increasing employment (e.g. in local economies with a low supply of low-skill jobs), then it makes sense for the researcher to prioritize the skill-development causal channel. In other words, the researcher's goal should be to identify which job training program leads to the largest increase in employment specifically via the skill-development channel, regardless of the magnitude of the effect transmitted via the channel of job-search motivation and perhaps even regardless of the relative magnitudes of programs' overall ATEs. In that case, the researcher's goal would be achieved by investigating the first CCM estimand, which would measure how much larger one treatment's skill-development causal channel is than that of the alternative treatment(s).

If, instead, the researcher is interested in better understanding multiple treatments' relative causal anatomies more generally, then both the first and second CCM estimands should be of interest. Considering both estimands could be useful in particular for theoretically motivated researchers who are seeking to test theories involving multiple treatments. Such theories not only predict whether one treatment should be more effective than another but also often dictate (a) the specific causal mechanisms that should grow or shrink when switching from one treatment to another and (b) the specific causal mechanisms that should contribute a larger share of the overall ATE for one treatment versus another. Indeed, for the purposes of theory testing and exploration, the two CCM estimands could be considered in conjunction with the ATEs to form a full picture of the relative causal anatomies of different treatments. To illustrate, Table H2 provides a set of some of the theoretical implications that would follow from testing hypotheses about the CCM estimands in combination with the ATEs.

Table H1: General Research Questions Related to Each CCM Estimand

Estimand 1	Does T_2 exhibit stronger effect transmission via mediator M than T_1 does? Does the second treatment have a larger mediated effect in absolute terms? $H_0 : \frac{ACME_2}{ACME_1} = 1 \quad H_a : \frac{ACME_2}{ACME_1} > 1$
Estimand 2	Does effect transmission via mediator M make up a larger proportion of ATE_2 relative to ATE_1? Is M more important for ATE_2 than ATE_1? $H_0 : \frac{\left(\frac{ACME_2}{ATE_2}\right)}{\left(\frac{ACME_1}{ATE_1}\right)} = 1 \quad H_a : \frac{\left(\frac{ACME_2}{ATE_2}\right)}{\left(\frac{ACME_1}{ATE_1}\right)} > 1$

Table H2: Theoretical Implications of Combined Hypotheses

	$\frac{ACME_2}{ACME_1} > 1$	$\frac{\left(\frac{ACME_2}{ATE_2}\right)}{\left(\frac{ACME_1}{ATE_1}\right)} > 1$	
$ATE_2 > ATE_1$	**yes**	**yes**	**Disproportionate scaling up**: Causal channel via M is larger in both absolute and proportional terms for second treatment. M is disproportionately responsible for enhancement of the effect when switching from first to second treatment.
	no	**no**	**Unrelatedness of mediator**: The larger effect of the second treatment is not due to M.
	yes	**no**	**Proportionate scaling up**: Causal channel via M is larger in absolute but not proportional terms for second treatment. M shares responsibility with other causal channels for enhancement of the effect when switching from first to second treatment.
$ATE_2 = ATE_1$	**yes**	**yes**	**Distinct causal anatomies**: Despite equivalent ATEs, the treatments are comprised of differently sized causal channels, with M constituting a larger channel for the second treatment.
	no	**no**	**Indistinguishable causal anatomies**: Any differences in the treatments' causal anatomies are unrelated to M.

Note: Missing yes/no conditions are not applicable.