Supplementary Materials

Soojin Park[1]

Steffi Pohl[2]

Peter M. Steiner[3]

[1]School of Education, University of California, Riverside

[2]Department of Education and Psychology, Freie University

[3]Department of Human Development and Quantitative Methodology, University of Maryland

Supplementary Materials

This document provides the online supplementary materials of the article "Identifying Causes of Treatment Effect Heterogeneity Across Replication Studies: Definitions, Identification, and Estimation of Remaining Heterogeneity".

A. Variables Used in the Case Study

Table S1 provides A complete list of person characteristics and setting characteristics used in the case study.

Table S1
Baseline covariates (person characteristics) and mediators (setting characteristics)

Person characteristics	**Setting Characteristics**
Gender [female]	Location [home vs. not at home]
Age	Device [mobile vs. no mobile]
German skills [native vs. non-native]	Presence others [yes vs. no]
Profession [academic vs. non-academic]	Noise
Subjective social status	Outside temperature
Prior knowledge about psychology	
Prior knowledge about social psychology	
Prior knowledge about contact hypothesis	
Prior contact	
Extended contact	
Intergroup anxiety	
Trait empathy	
Authoritarian attitudes	
Social dominance orientation	
Political orientation	
Vividness of imagination	
Motivation to participate	
Interest in incentive	
Interest in topic	

B. A proof of the ATE after fixing $W' = w'$

We can express $E[Y(1, w') - Y(0, w')|R = r, Z = z]$ for any $r \in \{0, 1\}$ as

$$
\begin{aligned}
&= \sum_w E[Y(1, w') - Y(0, w')|R = r, Z = z, W = w)P(w|R = r, Z = z) \\
&= \sum_w E[Y(1, w') - Y(0, w')|R = r, Z = z, W = w, W' = w')P(w|R = r, Z = z) \\
&= \sum_w E[Y(1, w')|R = r, X = 1, Z = z, W = w, W' = w')P(w|R = r, Z = z) \\
&\quad - \sum_w E[Y(0, w')|R = r, X = 0, Z = z, W = w, W' = w')P(w|R = r, Z = z) \\
&= \sum_w E[Y|R = r, X = 1, Z = z, W = w, W' = w']P(w|R = r, Z = z) \\
&\quad - \sum_w E[Y|R = r, X = 0, Z = z, W = w, W' = w']P(w|R = r, Z = z)
\end{aligned}
\tag{1}
$$

The first equality is due to the law of total probability over W. The second equality is due to B1 (no unmeasured W'-specific effect-moderators). The third equality is due to A1 (randomization of treatment). The last equality is due to B3 (consistency). This completes the proof.

C. A proof of the ATE after equalizing W'

We can express $E[Y(1, G_{W'|R=0,X=1}) - Y(0, G_{W'|R=0,X=0})|R = 1, Z = z]$ as

$$
\begin{aligned}
&= \sum_{w'} E[Y(1, G_{W'|R=0} = w') - Y(0, G_{W'|R=0} = w')|R = 1, Z = z, G_{W'|R=0} = w'] \\
&\quad P(G_{W'|R=0} = w'|R = 1, Z = z) \\
&= \sum_{w'} E[Y(1, w') - Y(0, w')|R = 1, Z = z]P(w'|R = 0) \\
&= \sum_{w,w'} E[Y(1, w') - Y(0, w')|R = 1, Z = z, W = w)P(w|R = 1, Z = z]P(w'|R = 0) \\
&= \sum_{w,w'} E[Y(1, w') - Y(0, w')|R = 1, Z = z, W = w, W' = w')P(w|R = 1, Z = z]P(w'|R = 0) \\
&= \sum_{w,w'} E[Y(1, w')|R = 1, X = 1, Z = z, W = w, W' = w']P(w|R = 1, Z = z]P(w'|R = 0) \\
&\quad - \sum_{w,w'} E[Y(0, w')|R = 1, X = 0, Z = z, W = w, W' = w']P(w|R = 1, Z = z]P(w'|R = 0) \\
&= \sum_{w,w'} E[Y|R = 1, X = 1, Z = z, W = w, W' = w']P(w|R = 1, Z = z)P(w'|R = 0) \\
&\quad - \sum_{w,w'} E[Y|R = 1, X = 0, Z = z, W = w, W' = w']P(w|R = 1, Z = z)P(w'|R = 0)
\end{aligned}
\tag{2}
$$

The first equality is due to the law of total probability over W'. The second equality is due to the definition of $G_{W'|R=0}$, which is a random draw from the distribution $P(w'|R = 0)$. As it is a random draw, it is independent of any observed values. The third equality is due to the law of total probability over W. The fourth equality is due to B1 (no unmeasured W'-specific effect-moderators). The fifth equality is due to A1 (randomization of treatment). The last equality is due to B3 (consistency). This completes the proof.

D. A proof of the parametric identification

To estimate the overall treatment effect heterogeneity, we have

$$\begin{aligned} & E[Y(1) - Y(0)|R = 1] \\ =& E[Y|R = 1, X = 1] - E[Y|R = 1, X = 0] \\ =& \tau + \tau_r + \tau_x + \tau_{rx} - \{\tau + \tau_r\} \\ =& \tau_x + \tau_{rx}. \end{aligned} \tag{3}$$

The second equality is by applying equation (12) of the main text. Similarly, $E[Y(1) - Y(0)|R = 0] = \tau_x$. Therefore, the raw treatment effect heterogeneity is $E[Y(1) - Y(0)|R = 1] - E[Y(1) - Y(0)|R = 0] = \tau_{rx}$.

To estimate remaining heterogeneity after adjusting for Z, we have

$$\begin{aligned} & \sum_Z E[Y(1) - Y(0)|R = 1, Z]P(Z) \\ =& \sum_Z E[Y|R = 1, X = 1, Z] - E[Y|R = 1, X = 0, Z]P(Z) \\ =& \phi + \phi_r + \phi_x + \phi_{rx} + \phi_z E[Z] + \phi_{xz} E[Z] + \phi_{rz} E[Z] + \phi_{rxz} E[Z] - \{\phi + \phi_r + \phi_z E[Z] + \phi_{rz} E[Z]\} \\ =& \phi_x + \phi_{rx} + (\phi_{xz} + \phi_{rxz}) E[Z]. \end{aligned} \tag{4}$$

The second equality is by applying equation (12) of the main text. Similarly, $E[Y(1) - Y(0)|R = 0, Z] = \phi_x + \phi_{xz} E[Z]$. Therefore, the conditional heterogeneity is $\sum_Z \{E[Y(1) - Y(0)|R = 1, Z] - E[Y(1) - Y(0)|R = 0, Z]\} = \phi_{rx} + \phi_{rxz} E[Z]$.

To estimate the remaining heterogeneity after applying static intervention, we have

$$
\begin{aligned}
&\sum_z E[Y(1,w) - Y(0,w)|R=1, Z=z]P(Z=z) \\
=&\sum_{z,w}\{E[Y|R=1,X=1,Z=z,W=w,W'=w'] - E[Y|R=1,X=0,Z=z,W=w,W'=w']\} \\
&\cdot P(W=w|R=1,Z=z)P(Z=z) \\
=&\gamma + \gamma_r + \gamma_x + \gamma_w E[W|R=1,Z=z] + \gamma_{w'}w' + \gamma_{xw'}w' + \gamma_{rx} + \gamma_{rxw'}w' + \{\gamma_z + \gamma_{xz} + \gamma_{rz} + \gamma_{rxz}\}E[Z] \\
&-\{\gamma + \gamma_r + \gamma_w E[W|R=1,Z=z] + \gamma_{w'}w' + (\gamma_z + \gamma_{rz})E[Z]\} \\
=&\gamma_x + \gamma_{rx} + \{\gamma_{xw'} + \gamma_{rxw'}\}w' + \{\gamma_{xz} + \gamma_{rxz}\}E[Z].
\end{aligned}
\tag{5}
$$

The second equality is by applying equation (13) of the main text. Similarly, $\sum_z E[Y(1,w) - Y(0,w)|R=0,Z=z]P(Z=z) = \gamma_x + \gamma_{xw'}w' + \gamma_{xz}E[Z]$. Therefore, the remaining heterogeneity at $W'=w'$ is $\sum_z\{E[Y(1,w) - Y(0,w)|R=1,Z=z] - E[Y(1,w) - Y(0,w)|R=0,Z=z]\}P(Z=z) = \gamma_{rx} + \gamma_{rxw'}w' + \gamma_{rxz}E[Z]$.

To estimate the remaining heterogeneity after applying stochastic intervention, we have

$$
\begin{aligned}
&\sum_z E[Y(1) - Y(0)|R=0,Z=z]P(Z=z) \\
=&\sum_z\{E[Y|R=0,X=1,Z=z] - E[Y|R=0,X=0,Z=z]\}P(Z=z) \\
=&\sum_{z,w,w'}\{E[Y|R=0,X=1,z,w,w'] - E[Y|R=0,X=0,z,w,w']\}P(w|R=0,z)P(w'|R=0,z,w)P(z) \\
=&\gamma + \gamma_x + \gamma_w E[W|R=0,Z] + (\gamma_{w'} + \gamma_{xw'})E[W'|R=0,z] + \{\gamma_z + \gamma_{xz}\}E[Z] \\
&-\{\gamma + \gamma_w E[W|R=0,Z] + \gamma_{w'}E[W'|R=0,Z] + \gamma_z E[Z]\} \\
=&\gamma_x + (\gamma_{xw'})E[W'|R=0,Z] + \gamma_{xz}E[Z]
\end{aligned}
\tag{6}
$$

The second equality is due to the law of total probability. The third equality is by applying

equation (13) of the main text. Similarly, $\sum_z E[Y(1) - Y(0)|R = 1, Z = z]P(Z = z) = \gamma_x + (\gamma_{xw'} + \gamma_{rxw'})E[W'|R = 1, Z] + \gamma_{rx} + \gamma_{xz}E[Z] + \gamma_{rxz}E[Z]$.

likewise, we have

$$
\begin{aligned}
&\sum_z E[Y(1, G_{W'|R=0,Z=z}) - Y(0, G_{W'|R=0,Z=z})|R = 1, Z = z]P(z) \\
&= \sum_{z,w,w'} \{E[Y|R = 1, X = 1, Z = z, W = w, W' = w'] - E[Y|R = 1, X = 0, Z = z, W = w, W' = w']\} \\
&\cdot P(W = w|R = 1, Z = z)P(W' = w'|R = 0, Z = z)P(z) \\
&= \gamma + \gamma_r + \gamma_x + \gamma_w E[W|R = 0, Z] + (\gamma_{w'} + \gamma_{xw'} + \gamma_{rxw'})E[W'|R = 0, Z] + \gamma_{rx} \\
&+ \{\gamma_z + \gamma_{xz} + \gamma_{rxz}\}E[Z] - \{\gamma + \gamma_r + \gamma_w E[W|R = 0, Z] + \gamma_{w'}E[W'|R = 1, z] + \gamma_z E[Z]\} \\
&= \gamma_x + \gamma_{rx} + (\gamma_{xw'} + \gamma_{rxw'})E[W'|R = 0, Z] + \{\gamma_{xz} + \gamma_{rxz}\}E[Z]
\end{aligned}
\tag{7}
$$

Using the equations (6) and (7), we can obtain $\zeta_{G_{W'}}$ as

$$
\begin{aligned}
&\sum_z \{E[Y(1, G_{W_k|R=0,X=1}) - E[Y(0, G_{W_k|R=0,X=0})|R = 1, Z = z] - E[Y(1) - Y(0)|R = 0, Z = z]\}P(z) \\
&= \gamma_x + \gamma_{rx} + (\gamma_{xw'} + \gamma_{rxw'})E[W'|R = 0, Z] + \{\gamma_{xz} + \gamma_{rxz}\}E[Z] \\
&- \{\gamma_x + (\gamma_{xw'})E[W'|R = 0, Z] + \gamma_{xz}E[Z]\}. \\
&= \gamma_{rx} + \gamma_{rxw'}E[W'|R = 0, Z] + \gamma_{rxz}E[Z]
\end{aligned}
\tag{8}
$$

Likewise, we can obtain $\delta_{G_{W'}}$ as

$$\begin{aligned}
&\sum_z \{E[Y(1)-Y(0)|R=1,Z=z] - E[Y(1,G_{W_k|R=0,X=1}) - E[Y(0,G_{W_k|R=0,X=0})|R=1,Z=z]\}P(z) \\
=&\gamma_x + \gamma_{rx} + (\gamma_{xw'} + \gamma_{rxw'})E[W'|R=1,Z] + (\gamma_{xz} + \gamma_{rxz})E[Z] \\
&-\{\gamma_x + \gamma_{rx} + (\gamma_{xw'} + \gamma_{rxw'})E[W'|R=0,Z] + (\gamma_{xz} + \gamma_{rxz})E[Z]\} \\
=&(\gamma_{xw'} + \gamma_{rxw'}) \times \{E[W'|R=1,Z] - E[W'|R=0,Z]\}. \\
=&(\gamma_{xw'} + \gamma_{rxw'}) \times (\beta_r + \beta_{rz}E[Z]).
\end{aligned} \tag{9}$$

This completes the proof.

E. Implementation via `R`

To implement the proposed methods, the source code is available at https://github.com/soojinpark33/replication. Once downloaded, the functions can be loaded into the `R` session as follows:

```
# Load the core functions
source("replication_functions.R")
```

Before running the analysis, ensure that the data has all the required inputs of the analysis. Across all three estimators (i.e., coefficient-based, imputation, and Double Lasso) the core inputs are the same:

- `B`: Number of bootstraps.
- `df`: The analytical dataset.
- `Y`, `R`, `X`: Variable names for the outcome, replication status, and treatment.
- `Zvars`, `Wvars`: Vectors for baseline covariates and mediators.

- `Wprime`: The target mediator.
- `w_fit`: The value at which to fix W'

Then each function can be called depending on the desired approach. For the Double Lasso analysis, the `lambda_rule` parameter must be specified as either "min" (when minimizing average validation error) and "1se" (when choosing the simplest model within one standard error from the minimum error). See the example codes below:

```
# 1. Coefficient-based approach (relies on Eq. (12) and (13))
poc_results <- bootstrap_poc(B=1000, df=df, Y="Y", R="R", X="X",
                             Zvars=c("age", "gender"),
                             Wvars=c("noise", "loc"),
                             Wprime="device", w_fix=0)
# 2. Imputation estimator (relies on Eq. (12) and (13))
imp_results <- bootstrap_imputation(B=1000, df=df, Y="Y", R="R", X="X",
                                    Zvars=c("age", "gender"),
                                    Wvars=c("noise", "loc"),
                                    Wprime="device", w_fix=0)
# 3. Double Lasso imputation (requires 'lambda_rule')
lasso_results <- bootstrap_lasso(B=1000, df=df, Y="Y", R="R", X="X",
                                 Zvars=c("age", "gender"),
                                 Wvars=c("noise", "loc"),
                                 Wprime="device", w_fix=0,
                                 lambda_rule="min")
```