Supplemental Material

S1 Proof of Chiba et al.'s (2011) Sensitivity Analysis Formula (7)

Note that individuals with observed A = 1 and S = s must have $S_1 = s$. Let $\pi_u = Pr(U = u)$, where $u = ss, s\bar{s}, \bar{s}s, \bar{s}\bar{s}$ denote the proportion of individuals in each principal stratum. We can express $E(Y_1|A = 1, S = s)$ as the weighted sum of $E(Y_1|U = ss)$ and $E(Y_1|U = s\bar{s})$:

$$E(Y_1|A = 1, S = s) = \frac{\pi_{s\bar{s}}E(Y_1|U = s\bar{s}) + \pi_{ss}E(Y_1|U = ss)}{\pi_{s\bar{s}} + \pi_{ss}}, \tag{S1.1}$$

where $\pi_{s\bar{s}}+\pi_{ss} = Pr(S_1 = s, S_0 = \bar{s})+Pr(S_1 = s, S_0 = s) = Pr(S_1 = s) = Pr(S_1 = s|A = 1) = Pr(S = s|A = 1) = p_1$ because S_a (a = 1 or 0) is independent from A due to randomization. Likewise, because individuals with the observed value of A = 0 and S = s are limited to those with $S_0 = s$, $E(Y_0|A = 0, S = s)$ can be expressed by the weighted sum of $E(Y_0|U = ss)$ and $E(Y_0|U = \bar{s}s)$:

$$E(Y_0|A = 0, S = s) = \frac{\pi_{\bar{s}s}E(Y_0|U = \bar{s}s) + \pi_{ss}E(Y_0|U = ss)}{\pi_{\bar{s}s} + \pi_{ss}}, \tag{S1.2}$$

where $\pi_{\bar{s}s} + \pi_{ss} = Pr(S_1 = \bar{s}, S_0 = s) + Pr(S_1 = s, S_0 = s) = Pr(S_0 = s) = Pr(S_0 = s|A = 0) = Pr(S = s|A = 0) = p_0$.

Let $\beta_1 = E(Y_1|U = s\bar{s}) - E(Y_1|U = ss)$ denote the difference in average potential outcomes under TEST between the stratum "PP with TEST only" and the stratum "always PP." Substituting $E(Y_1|U = s\bar{s}) = \beta_1 + E(Y_1|U = ss)$ into equation (S1.1) yields

$$\begin{aligned}
E(Y_1|A = 1, S = s) &= \frac{\pi_{s\bar{s}}(\beta_1 + E(Y_1|U = ss)) + \pi_{ss}E(Y_1|U = ss)}{\pi_{s\bar{s}} + \pi_{ss}} \\
&= \frac{(\pi_{s\bar{s}} + \pi_{ss})E(Y_1|U = ss)) + \pi_{s\bar{s}}\beta_1}{\pi_{s\bar{s}} + \pi_{ss}} \\
&= E(Y_1|U = ss) + \frac{\pi_{s\bar{s}}}{p_1}\beta_1 \\
&= E(Y_1|U = ss) + \frac{p_1 - p_0 + \pi_{\bar{s}s}}{p_1}\beta_1, \qquad \text{(S1.3)}
\end{aligned}$$

where $\pi_{s\bar{s}} = p_1 - \pi_{ss} = p_1 - (p_0 - \pi_{\bar{s}s}) = p_1 - p_0 + \pi_{\bar{s}s}$.

Similarly, Let $\beta_0 = E(Y_0|U = \bar{s}s) - E(Y_0|U = ss)$ denote the difference in average potential outcomes under RLD between the stratum "PP with RLD only" and the stratum "always PP." Substituting $E(Y_0|U = \bar{s}s) = \beta_0 + E(Y_0|U = ss)$ into (S1.2) yields

$$E(Y_0|A = 0, S = s) = E(Y_0|U = ss) + \frac{\pi_{\bar{s}s}}{p_0}\beta_0. \tag{S1.4}$$

In addition, $E(Y_a|A = a, S = s) = E(Y|A = a, S = s)$ because of consistency assumption (a persons potential outcome under a hypothetical condition is precisely the outcome experienced by that person (Robins et al., 2000)), equation (7) is therefore proved.

S2 Proof of the boundaries of $\pi_{\bar{s}s}$ (8)

As previously explained, $\pi_{s\bar{s}} + \pi_{ss} = Pr(S_1 = s, S_0 = \bar{s}) + Pr(S_1 = s, S_0 = s) = Pr(S_1 = s) = Pr(S_1 = s|A = 1) = Pr(S = s|A = 1) = p_1$; Similarly $\pi_{\bar{s}s} + \pi_{ss} = Pr(S_1 = \bar{s}, S_0 = s) + Pr(S_1 = s, S_0 = s) = Pr(S_0 = s) = Pr(S_0 = s|A = 0) = Pr(S = s|A = 0) = p_0$. Therefore, we have following three equations:

$$\begin{cases} \pi_{ss} + \pi_{s\bar{s}} = p_1, \\ \pi_{ss} + \pi_{\bar{s}s} = p_0, \\ \pi_{ss} + \pi_{s\bar{s}} + \pi_{\bar{s}s} + \pi_{\bar{s}\bar{s}} = 1; \end{cases} \tag{S2.1}$$

which implies that

$$\begin{cases} \pi_{ss} = p_0 - \pi_{\bar{s}s}, \\ \pi_{s\bar{s}} = p_1 - p_0 + \pi_{\bar{s}s}, \\ \pi_{\bar{s}\bar{s}} = 1 - p_1 - \pi_{\bar{s}s}. \end{cases} \tag{S2.2}$$

In addition, because $\pi_u = Pr(U = u), u = ss, s\bar{s}, \bar{s}s, \bar{s}\bar{s}$ are bounded probabilities, i.e.,

$$\begin{cases} 0 \le \pi_{\bar{s}s} \le p_0, \\ 0 \le \pi_{ss} \le min(p_0, p_1) \\ 0 \le \pi_{s\bar{s}} \le p_1, \\ 0 \le \pi_{\bar{s}\bar{s}} \le min(1 - p_0, 1 - p_1). \end{cases} \tag{S2.3}$$

Substituting in (S2.2) into (S2.3), we have

$$\begin{cases} 0 \le \pi_{\bar{s}s} \le p_0, \\ 0 \le p_0 - \pi_{\bar{s}s} \le min(p_0, p_1) \\ 0 \le p_1 - p_0 + \pi_{\bar{s}s} \le p_1, \\ 0 \le 1 - p_1 - \pi_{\bar{s}s} \le min(1 - p_0, 1 - p_1). \end{cases} \tag{S2.4}$$

The four inequalities imply that $p_0 - p_1 \leq \pi_{\bar{s}s} \leq p_0$

$$\begin{cases} 0 \leq \pi_{\bar{s}s} \leq p_0, \\ max[0, p_0 - p_1] \leq \pi_{\bar{s}s} \leq p_0, \\ p_0 - p_1 \leq \pi_{\bar{s}s} \leq p_0, \\ max[0, p_0 - p_1] \leq \pi_{\bar{s}s} \leq 1 - p_1. \end{cases} \tag{S2.5}$$

Therefore, $max[0, p_0 - p_1] \leq \pi_{\bar{s}s} \leq min[p_0, 1 - p_1]$, and (8) is proved.