eAppendix for "Bounding the infectiousness effect in vaccine trials"

1. Proof that the Crude Estimator is Conservative under Selection Bias Due to Pathogen Virulence

For those with $Y_{i1}(1,0)$ and $Y_{i1}(0,0) = 1$ respectively we let $S_i(1)$ and $S_i(0)$ denote respectively the virulence of the pathogen causing the infection for individual 1 when vaccinated or unvaccinated. Suppose that assumption 1 holds along with a modification of assumption 2 stated below and an assumption on monotonicity for S_i, namely,

Assumption 2*. $E[Y_{i2}(0,0)|A_{i1} = 0, Y_{i1} = 1, S_i = 0] \leq E[Y_{i2}(0,0)|A_{i1} = 1, Y_{i1} = 1, S_i = 0]$.

Assumption 3. For all i, $S_i(0) \leq S_i(1)$.

Assumption 2* states that the average infection rate for individual 2 if both individuals 1 and 2 were unvaccinated would be lower in the subgroup of household for which individual 1 would be infected unvaccinated with low pathogen strengh ($S_i = 0$) than in the subgroup of households for which individual 1 would be infected vaccinated with low pathogen virulence ($S_i = 0$). Assumption 3 essentially states that for individuals who would be infected irrespective of vaccination status, the virulence of the pathogen causing infection when the individual is vaccinated is at least as virulent as the pathogen causing the infection when the individual is unvaccinated.

Result 3. Under assumptions 1, 2* and 3,

$$\begin{aligned} & E[Y_{i2}(1,0) - Y_{i2}(0,0)|Y_{i1}(1,0) = Y_{i1}(0,0) = 1, S_i(0) = S_i(1) = 0] \\ \leq \quad & E[Y_{i2}|A_{i1} = 1, Y_{i1} = 1, S_i = 0] - E[Y_{i2}|A_{i1} = 0, Y_{i1} = 1, S_i = 0]. \end{aligned}$$

Proof of Result 3. Under assumptions 1 and 3 we have that

$$\begin{aligned} & E[Y_{i2}(1,0)|Y_{i1}(1,0) = Y_{i1}(0,0) = 1, S_i(0) = S_i(1) = 0] \\ = \quad & E[Y_{i2}(1,0)|A_{i1} = 1, Y_{i1}(1,0) = Y_{i1}(0,0) = 1, S_i(0) = S_i(1) = 0] \\ = \quad & E[Y_{i2}(1,0)|A_{i1} = 1, Y_{i1}(1,0) = 1, S_i(1) = 0] \\ = \quad & E[Y_{i2}|A_{i1} = 1, Y_{i1} = 1, S_i = 0] \end{aligned}$$

where the first equality follows by randomization of A_{i1}, the second by Assumption 1 and Assumption 3 and the third by consistency. Also under assumptions 1 and 3 we have

$$\begin{aligned}
& E[Y_{i2}(0,0)|Y_{i1}(1,0) = Y_{i1}(0,0) = 1, S_i(0) = S_i(1) = 0] \\
= \;& E[Y_{i2}(0,0)|A_{i1} = 1, Y_{i1}(1,0) = Y_{i1}(0,0) = 1, S_i(0) = S_i(1) = 0] \\
= \;& E[Y_{i2}(0,0)|A_{i1} = 1, Y_{i1}(1,0) = 1, S_i(1) = 0] \\
= \;& E[Y_{i2}(0,0)|A_{i1} = 1, Y_{i1} = 1, S_i = 0] + \{E[Y_{i2}(0,0)|A_{i1} = 0, Y_{i1} = 1, S_i = 0] - E[Y_{i2}(0,0)|A_{i1} = 0, Y_{i1} = 1, S_i = 0]\} \\
= \;& E[Y_{i2}(0,0)|A_{i1} = 1, Y_{i1} = 1, S_i = 0] + \{E[Y_{i2}|A_{i1} = 0, Y_{i1} = 1, S_i = 0] - E[Y_{i2}(0,0)|A_{i1} = 0, Y_{i1} = 1, S_i = 0]\} \\
= \;& E[Y_{i2}|A_{i1} = 0, Y_{i1} = 1, S_i = 0] + \{E[Y_{i2}(0,0)|A_{i1} = 1, Y_{i1} = 1, S_i = 0] - E[Y_{i2}(0,0)|A_{i1} = 0, Y_{i1} = 1, S_i = 0]\}
\end{aligned}$$

where the first equality follows by randomization of A_{i1}, the second by Assumption 1 and Assumption 3 and third and fourth by consistency. Under Assumption 2* we then have

$$\begin{aligned}
& E[Y_{i2}(1,0) - Y_{i2}(0,0)|Y_{i1}(1,0) = Y_{i1}(0,0) = 1, S_i(0) = S_i(1) = 0] \\
= \;& E[Y_{i2}|A_{i1} = 1, Y_{i1} = 1, S_i = 0] - E[Y_{i2}|A_{i1} = 0, Y_{i1} = 1, S_i = 0] \\
& \quad + \{E[Y_{i2}(0,0)|A_{i1} = 0, Y_{i1} = 1, S_i = 0] - E[Y_{i2}(0,0)|A_{i1} = 1, Y_{i1} = 1, S_i = 0]\} \\
\leq \;& E[Y_{i2}|A_{i1} = 1, Y_{i1} = 1, S_i = 0] - E[Y_{i2}|A_{i1} = 0, Y_{i1} = 1, S_i = 0].
\end{aligned}$$

This completes the proof.

To show a similar result for $E[Y_{i2}(1,0) - Y_{i2}(1,0)|Y_{i1}(1,0) = Y_{i1}(0,0) = 1, S_i(0) = S_i(1) = 1]$ when conditioning on a more virulent pathogen causing the infection for individual 1 irrespective of vaccination status ($S_i(0) = S_i(1) = 1$), we need another variant on Assumption 2 and one further assumption is also needed:

Assumption 2**. $E[Y_{i2}(0,0)|A_{i1} = 0, Y_{i1} = 1, S_i(0) = 1] \leq E[Y_{i2}(0,0)|A_{i1} = 1, Y_{i1} = 1, S_i(0) = 1]$.

Assumption 4. $E[Y_{i2}(1,0)|A_{i1} = 1, Y_{i1}(1,0) = 1, S_i(0) = 1] \leq E[Y_{i2}(1,0)|A_{i1} = 1, Y_{i1}(1,0) = 1, S_i(1) = 1]$.

Assumption 4 compares the expectation of $Y_{i2}(1,0)$ for two subpopulations, those with $A_{i1} = 1, Y_{i1}(1,0) = 1, S_i(0) = 1$ and $A_{i1} = 1, Y_{i1}(1,0) = 1, S_i(1) = 1$. A special case in which assumption 4 will hold are when the two expectations are equal. Note that if $S_i(0) = 1$ then under Assumption 3 $S_i(1) = 1$ also; thus for the counterfactual $Y_{i2}(1,0)$, if the virulence of the pathogen when individual 1 is unvaccinated and infected is irrelevant for the outcome of individual 2 when individual 1 is in fact vaccinated then equality will hold and Assumption 4 would apply.

Result 4. Under assumptions 1, 2**, 3 and 4,

$$\begin{aligned}
& E[Y_{i2}(1,0) - Y_{i2}(0,0)|Y_{i1}(1,0) = Y_{i1}(0,0) = 1, S_i(0) = S_i(1) = 1] \\
\leq \;& E[Y_{i2}|A_{i1} = 1, Y_{i1} = 1, S_i = 1] - E[Y_{i2}|A_{i1} = 0, Y_{i1} = 1, S_i = 1].
\end{aligned}$$

Proof of Result 4.

$$\begin{aligned}
& E[Y_{i2}(1,0)|Y_{i1}(1,0) = Y_{i1}(0,0) = 1, S_i(0) = S_i(1) = 1] \\
= \;& E[Y_{i2}(1,0)|A_{i1} = 1, Y_{i1}(1,0) = Y_{i1}(0,0) = 1, S_i(0) = S_i(1) = 1] \\
= \;& E[Y_{i2}(1,0)|A_{i1} = 1, Y_{i1}(1,0) = 1, S_i(0) = 1] \\
\leq \;& E[Y_{i2}(1,0)|A_{i1} = 1, Y_{i1}(1,0) = 1, S_i(1) = 1] \\
= \;& E[Y_{i2}|A_{i1} = 1, Y_{i1} = 1, S_i = 1]
\end{aligned}$$

where the first equality follows by randomization of A_{i1}, the second by Assumption 1 and Assumption 3, the third by Assumption 4 and the fourth by consistency. Under Assumptions 1 and 3 we have,

$$\begin{aligned}
& E[Y_{i2}(0,0)|Y_{i1}(1,0) = Y_{i1}(0,0) = 1, S_i(0) = S_i(1) = 1] \\
= \;& E[Y_{i2}(0,0)|A_{i1} = 1, Y_{i1}(1,0) = Y_{i1}(0,0) = 1, S_i(0) = S_i(1) = 1] \\
= \;& E[Y_{i2}(0,0)|A_{i1} = 1, Y_{i1}(1,0) = 1, S_i(0) = 1] \\
= \;& E[Y_{i2}(0,0)|A_{i1} = 1, Y_{i1}(1,0) = 1, S_i(0) = 1] + \{E[Y_{i2}|A_{i1} = 0, Y_{i1} = 1, S_i = 1] - E[Y_{i2}|A_{i1} = 0, Y_{i1} = 1, S_i = 1]\} \\
= \;& E[Y_{i2}(0,0)|A_{i1} = 1, Y_{i1} = 1, S_i(0) = 1] + \{E[Y_{i2}|A_{i1} = 0, Y_{i1} = 1, S_i = 1] - E[Y_{i2}(0,0)|A_{i1} = 0, Y_{i1} = 1, S_i(0) = 1]\} \\
= \;& E[Y_{i2}|A_{i1} = 0, Y_{i1} = 1, S_i = 1] + \{E[Y_{i2}(0,0)|A_{i1} = 1, Y_{i1} = 1, S_i(0) = 1] - E[Y_{i2}(0,0)|A_{i1} = 0, Y_{i1} = 1, S_i(0) = 1]\}
\end{aligned}$$

where the first equality follows by randomization of A_{i1}, the second by Assumption 1 and Assumption 3 and the fourth by consistency. Under Assumption 2** we then have

$$\begin{aligned}
& E[Y_{i2}(1,0) - Y_{i2}(0,0)|Y_{i1}(1,0) = Y_{i1}(0,0) = 1, S_i(0) = S_i(1) = 1] \\
\leq \;& E[Y_{i2}|A_{i1} = 1, Y_{i1} = 1, S_i = 1] - E[Y_{i2}|A_{i1} = 0, Y_{i1} = 1, S_i = 1] \\
& \quad + \{E[Y_{i2}(0,0)|A_{i1} = 0, Y_{i1} = 1, S_i(0) = 1] - E[Y_{i2}(0,0)|A_{i1} = 1, Y_{i1} = 1, S_i(0) = 1]\} \\
\leq \;& E[Y_{i2}|A_{i1} = 1, Y_{i1} = 1, S_i = 1] - E[Y_{i2}|A_{i1} = 0, Y_{i1} = 1, S_i = 1].
\end{aligned}$$

This completes the proof.

2. Proofs for the Causal Infectiousness Effect for Designs in Which Both Individuals are Randomized to Receive the Vaccine

Under asssumption $1^{\dagger\dagger}$,

$$\begin{aligned}
& E[Y_{i1}(1,1)|Y_{i2}(1,1)=Y_{i2}(1,0)=1] \\
= \quad & E[Y_{i1}(1,1)|A_{i1}=1, A_{i2}=1, Y_{i2}(1,1)=1] \\
= \quad & E[Y_{i1}|A_{i1}=1, A_{i2}=1, Y_{i2}=1]
\end{aligned}$$

where the first equality holds by randomization of A_{i1} and A_{i2} and Assumption $1^{\dagger\dagger}$ and the by consistency. Also under Assumption $1^{\dagger\dagger}$,

$$\begin{aligned}
& E[Y_{i1}(1,0)|Y_{i2}(1,1)=Y_{i2}(1,0)=1] \\
= \quad & E[Y_{i1}(1,0)|A_{i1}=1, A_{i2}=1, Y_{i2}(1,1)=1] \\
= \quad & E[Y_{i1}(1,0)|A_{i1}=1, A_{i2}=1, Y_{i2}(1,1)=1] \\
& \quad + \{E[Y_{i1}(1,0)|A_{i1}=1, A_{i2}=0, Y_{i2}=1] - E[Y_{i1}(1,0)|A_{i1}=1, A_{i2}=0, Y_{i2}=1]\} \\
= \quad & E[Y_{i1}(1,0)|A_{i1}=1, A_{i2}=1, Y_{i2}=1] + \{E[Y_{i1}|A_{i1}=1, A_{i2}=0, Y_{i2}=1] - E[Y_{i1}(1,0)|A_{i1}=1, A_{i2}=0, Y_{i2}=1]\} \\
= \quad & E[Y_{i1}|A_{i1}=1, A_{i2}=0, Y_{i2}=1] + \{E[Y_{i1}(1,0)|A_{i1}=1, A_{i2}=1, Y_{i2}=1] - E[Y_{i1}(1,0)|A_{i1}=1, A_{i2}=0, Y_{i2}=1]\}
\end{aligned}$$

where the first equality holds by randomization of A_{i1} and A_{i2} and Assumption $1^{\dagger\dagger}$ and the third by consistency. We thus have

$$\begin{aligned}
& E[Y_{i1}(1,1) - Y_{i1}(1,0)|Y_{i2}(1,1)=Y_{i2}(1,0)=1] \\
= \quad & E[Y_{i1}|A_{i1}=1, A_{i2}=1, Y_{i2}=1] - E[Y_{i1}|A_{i1}=1, A_{i2}=0, Y_{i2}=1] \\
& \quad + \{E[Y_{i1}(1,0)|A_{i1}=1, A_{i2}=0, Y_{i2}=1] - E[Y_{i1}(1,0)|A_{i1}=1, A_{i2}=1, Y_{i2}=1]\}
\end{aligned}$$

and thus under Assumption $2^{\dagger\dagger}$,

$$\begin{aligned}
& E[Y_{i1}(1,1) - Y_{i1}(1,0)|Y_{i2}(1,1)=Y_{i2}(1,0)=1] \\
\leq \quad & E[Y_{i1}|A_{i1}=1, A_{i2}=1, Y_{i2}=1] - E[Y_{i1}|A_{i1}=1, A_{i2}=0, Y_{i2}=1].
\end{aligned}$$