SUPPLEMENTARY MATERIAL: IDENTIFICATION OF VACCINE EFFECTS WHEN EXPOSURE STATUS IS UNKNOWN

Appendix A. Proofs

Proof of Proposition 1. For any $a, a' \in \{0, 1\}$, use laws of probability and conditions (5) and (7)-(10) to express

$$\begin{aligned}\mathbb{E}(Y \mid A = a) =& \mathbb{E}(Y \mid E = 1, A = a) \underbrace{P(E = 1 \mid A = a)}_{= P(E=1|A=a')} + \underbrace{\mathbb{E}(Y \mid A = a, E = 0) P(E = 0 \mid A = a)}_{=0} \\ =& \mathbb{E}(Y \mid E = 1, A = a) P(E = 1 \mid A = a').\end{aligned}$$

Thus,

$$\begin{aligned}\frac{\mathbb{E}(Y \mid A = 1)}{\mathbb{E}(Y \mid A = 0)} &= \frac{\mathbb{E}(Y \mid E = 1, A = 1) P(E = 1 \mid A = 1)}{\mathbb{E}(Y \mid E = 1, A = 0) P(E = 1 \mid A = 0)} \\ &= \frac{\mathbb{E}(Y \mid E = 1, A = 1)}{\mathbb{E}(Y \mid E = 1, A = 0)} \\ &= \frac{\mathbb{E}(Y^{a=1} \mid E = 1)}{\mathbb{E}(Y^{a=0} \mid E = 1)},\end{aligned}$$

where the second and third line again follow due to assumption (5) and (7)-(10). □

Proof of Proposition 2. We first derive an upper bound.

$$\begin{aligned}&\mathbb{E}(Y^{a=0} \mid E = 1) - \mathbb{E}(Y^{a=1} \mid E = 1) \\ =& \mathbb{E}(Y \mid E = 1, A = 0) - \mathbb{E}(Y \mid E = 1, A = 1) \quad \text{due to (7) − (9)} \\ =& \frac{\mathbb{E}(Y \mid A = 0)}{P(E = 1 \mid A = 0)} - \frac{\mathbb{E}(Y \mid A = 1)}{P(E = 1 \mid A = 1)} \quad \text{due to (10) and laws of prob.} \\ =& \frac{\mathbb{E}(Y \mid A = 0)}{P(E = 1 \mid A = 0)} - \frac{\mathbb{E}(Y \mid A = 1)}{P(E = 1 \mid A = 0)} \quad \text{due to (5)} \\ \leq& \frac{\mathbb{E}(Y \mid A = 0)}{\mathbb{E}(Y \mid A = 0)} - \frac{\mathbb{E}(Y \mid A = 1)}{\mathbb{E}(Y \mid A = 0)} = 1 - \frac{\mathbb{E}(Y \mid A = 1)}{\mathbb{E}(Y \mid A = 0)}. \qquad \text{(S1)}\end{aligned}$$

The last line is an equality when $(E = 1 \iff Y = 1) \mid A = 0$.

A lower bound on the absolute CECE is given by

$$\begin{aligned}
&\mathbb{E}(Y^{a=0} \mid E = 1) - \mathbb{E}(Y^{a=1} \mid E = 1) \\
=&\frac{\mathbb{E}(Y \mid A = 0)}{P(E = 1 \mid A = 0)} - \frac{\mathbb{E}(Y \mid A = 1)}{P(E = 1 \mid A = 0)} \\
\geq&\mathbb{E}(Y \mid A = 0) - \mathbb{E}(Y \mid A = 1). \qquad \text{(S2)}
\end{aligned}$$

The last line in (S2) is an equality when $P(E = 1) = 1$. □

Proof of Proposition S1.

$$\begin{aligned}
&\frac{\mathbb{E}(Y^{a=1,e=1} \mid L)}{\mathbb{E}(Y^{a=0,e=1} \mid L)} = \frac{\mathbb{E}(Y \mid E = 1, A = 1, L)}{\mathbb{E}(Y \mid E = 1, A = 0, L)} \quad \text{due to (S10) – (S12)} \\
=&\frac{\mathbb{E}(Y \mid A = 1, L)}{\mathbb{E}(Y \mid A = 0, L)},
\end{aligned}$$

where the last equality follows from

$$\mathbb{E}(Y \mid E = 1, A = a, L) = \frac{\mathbb{E}(Y \mid A = a, L)}{P(E = 1 \mid A = a, L)} = \frac{\mathbb{E}(Y \mid A = a, L)}{P(E = 1 \mid A = a', L)},$$

using (10) and (5), similarly to the proof of Proposition (1). □

Proof of Proposition S2. The result follows from including L in the conditioning set in all the derivations of Proposition 1, which then gives the same identification result as in Proposition S1. □

Appendix B. Time-to-events and censoring

We re-introduce the terminology from Section *CECE in Time-to-Event Settings*. Let Y_k and E_k be time-to-event variables indicating whether an individual has experienced the event by time k ($Y_k = 1$) and being exposed by time k ($E_k = 1$), respectively. Let C_k denote loss to follow-up (censoring) by interval $k > 0$, and we define the temporal (and topological) order (C_k, E_k, Y_k) in each interval $k > 0$.

Suppose we are interested in outcomes in time intervals $k = 0, \ldots, K$. We adopt the convention that random variables with a negative subscript are equal to 0 (e.g., $Y_{-1} \equiv 0$).

Let the history of a random variable be denoted by a check symbol, e.g. $\check{Y}_k = (Y_0, Y_1, ..., Y_k)$ is the history of the event of interest through interval k. Further, let the future of a random variable through K be denoted by an underline, e.g. $\underline{Y}_k = (Y_k, Y_{k+1}, ..., Y_K)$.

Consider now classical identifiability conditions for causal effects in time-to-event settings, which are just extensions of (7)-(9).

Assumption (Treatment exchangeability)**.**

$$\text{(S3)} \qquad \check{Y}_K^{a,c=0}, \check{E}_K^{a,c=0} \perp\!\!\!\perp A,$$

$$\text{(S4)} \qquad \underline{Y}_k^{a,c=0} \perp\!\!\!\perp C_k \mid Y_{k-1} = C_{k-1} = 0, A = a.$$

Condition (S3) holds when A is randomly assigned. Condition (S4) requires that losses to follow-up are independent of future counterfactual events, given the measured past; this assumption, which corresponds to classical independent censoring assumptions, does not hold by design in a randomized trial, as losses to follow-up are not randomly assigned in practice. The treatment exchangeability conditions are satisfied in the SWIG in Figure 2.

Assumption (Positivity)**.**

$$\text{(S5)} \qquad P(A = a) > 0 \quad \forall a \in \{0, 1\}$$

$$P(Y_k = 0, C_k = 0, A = a) > 0 \implies$$

$$\text{(S6)} \qquad P(C_{k+1} = 0 \mid Y_k = 0, C_k = 0, A = a) > 0 \ ,$$

for all $a \in \{0, 1\}$ *and* $k < K$*.*

The positivity conditions require that for any possible history of treatment assignment and covariates among those who are event-free and uncensored at k, some subjects will remain uncensored at the next time $k+1$.

Assumption (Consistency)**.**

$$\begin{aligned} &\textit{if } A = a \textit{ and } C_k = 0, \\ &\textit{then } \check{Y}_k = \check{Y}_k^{a,c=0}, \check{E}_k = \check{E}_k^{a,c=0} \end{aligned} \tag{S7}$$

for all $a \in \{0,1\}$ *and* $k \leq K$*.*

Consistency holds if any individual who has data history consistent with the intervention under a counterfactual scenario, would have observed outcomes that are equal to the counterfactual outcomes.

Besides the classical identifiability conditions, we introduce the following conditions, which generalize exposure necessity (10) and the no effect on exposure assumption (5) from the main text.

Assumption (Time-varying exposure necessity)**.**

$$E_k^{a,c=0} = 0 \implies Y_k^{a,c=0} = 0, \tag{S8}$$

for all $a \in \{0,1\}$ *and* $k \leq K$*.*

Like (10), the assumption of time-varying exposure necessity states that the outcome can only happen in individuals who have been exposed to the virus. By definition, $E_{k-1}^{a,c=0} = 1 \implies E_k^{a,c=0} = 1$, so the time-varying nature of exposure and the outcome should not make this assumption less justifiable.

Assumption (No effect on exposure)**.**

$$E_k^{a=0,c=0} = E_k^{a=1,c=0}, \tag{S9}$$

for all $k \leq K$.

This assumption says that the risk of exposure by any time k is the same among treated and untreated. Consider a situation in which a vaccine A prevents or delays the outcome Y. Under blinding, condition (S9) would still hold because prior infection would be the only thing preventing future exposure, but under (S8), anyone with the outcome would have already been exposed. However, we must assume that blinding continues to be successful; that is, this assumption would be violated if over time individuals notice that they are not getting infected after the same level of exposure as people around them, and therefore conclude that they have been vaccinated and change behavior.

Under these conditions we sketch a proof for Proposition 4.

Sketch of proof of Proposition 4. We can invoke (S8)-(S9) to find that

$$\begin{aligned}
&\frac{\mathbb{E}(Y_k^{a=1,c=0})}{\mathbb{E}(Y_k^{a=0,c=0})} = \frac{\mathbb{E}(E_k^{a=1,c=0}Y_k^{a=1,c=0})}{\mathbb{E}(E_k^{a=0,c=0}Y_k^{a=0,c=0})} \\
=&\frac{\mathbb{E}(Y_k^{a=1,c=0} \mid E_k^{a=1,c=0}=1)\mathbb{E}(E_k^{a=1,c=0})}{\mathbb{E}(Y_k^{a=0,c=0} \mid E_k^{a=0,c=0}=1)\mathbb{E}(E_k^{a=0,c=0})} = \frac{\mathbb{E}(Y_k^{a=1,c=0} \mid E_k^{a=1,c=0}=1)}{\mathbb{E}(Y_k^{a=0,c=0} \mid E_k^{a=0,c=0}=1)},
\end{aligned}$$

where we used exposure necessity in the first equality, laws of probability in the second equality and the last equality follows because $\mathbb{E}(E_k^{a=0,c=0}) = \mathbb{E}(E_k^{a=1,c=0})$ under (S9).

Then, using treatment exchangeability, consistency and positivity, it follows that $\mathbb{E}(Y_k^{a,c=0})$ can be expressed in terms of the cumulative incidence function at k, $\mu_k(a)$.

The proof for the additive CECE follows the same structure as the proof of Proposition 2. □

Appendix C. Parallel to risk ratio under perfect specificity

A well-known result in epidemiology is the fact that under so-called non-differential misclassification of the outcome with perfect specificity, the exposure-outcome risk

ratio is unbiased, although the risk difference is not. For example, in the setting of possibly incomplete disease ascertainment in exposed and unexposed cohorts, Lawrence and Greenwald described how a screening program could be implemented to remove false positive cases, resulting in an unbiased risk ratio [**?**]. The requirement of perfect specificity parallels our exposure necessity assumption, and that of non-differential misclassification parallels our assumption of no effect on exposure. We demonstrate these parallels with the DAGs in Figure S1. Each has one partially deterministic arrow and one independence assumption, though the causal structures differ. The partially deterministic arrow and the independence assumption allow in each case for an unbiased ratio measure, as we demonstrate in the following derivation. Take Y to be a binary outcome and A any exposure of interest (also binary for simplicity). We denote a misclassified version of the outcome with Y^*. Then we have for the misclassification setting that

$$
\begin{aligned}
&\frac{P(Y^* = 1 \mid A = 1)}{P(Y^* = 1 \mid A = 0)} \\
&= \frac{P(Y^* = 1 \mid A = 1, Y = 1)P(Y = 1 \mid A = 1) + P(Y^* = 1 \mid A = 1, Y = 0)P(Y = 0 \mid A = 1)}{P(Y^* = 1 \mid A = 0, Y = 1)P(Y = 1 \mid A = 0) + P(Y^* = 1 \mid A = 0, Y = 0)P(Y = 0 \mid A = 0)} \\
&= \frac{P(Y^* = 1 \mid A = 1, Y = 1)P(Y = 1 \mid A = 1) + 0 \times P(Y = 0 \mid A = 1)}{P(Y^* = 1 \mid A = 0, Y = 1)P(Y = 1 \mid A = 0) + 0 \times P(Y = 0 \mid A = 0)} \\
&= \frac{P(Y^* = 1 \mid Y = 1)P(Y = 1 \mid A = 1)}{P(Y^* = 1 \mid Y = 1)P(Y = 1 \mid A = 0)} = \frac{P(Y = 1 \mid A = 1)}{P(Y = 1 \mid A = 0)},
\end{aligned}
$$

and a parallel derivation of our Proposition 1, that is,

$$
\begin{aligned}
&\frac{P(Y = 1 \mid A = 1)}{P(Y = 1 \mid A = 0)} \\
&= \frac{P(Y = 1 \mid A = 1, E = 1)P(E = 1 \mid A = 1) + P(Y = 1 \mid A = 1, E = 0)P(E = 0 \mid A = 1)}{P(Y = 1 \mid A = 0, E = 1)P(E = 1 \mid A = 0) + P(Y = 1 \mid A = 0, E = 0)P(E = 0 \mid A = 0)} \\
&= \frac{P(Y = 1 \mid A = 1, E = 1)P(E = 1 \mid A = 1) + 0 \times P(E = 0 \mid A = 1)}{P(Y = 1 \mid A = 0, E = 1)P(E = 1 \mid A = 0) + 0 \times P(E = 0 \mid A = 0)} \\
&= \frac{P(Y = 1 \mid A = 1, E = 1)P(E = 1)}{P(Y = 1 \mid A = 0, E = 1)P(E = 1)} = \frac{P(Y = 1 \mid A = 1, E = 1)}{P(Y = 1 \mid A = 0, E = 1)},
\end{aligned}
$$

where the second equality uses the appropriate partially deterministic arrow assumption and the third equality the appropriate independence assumption.

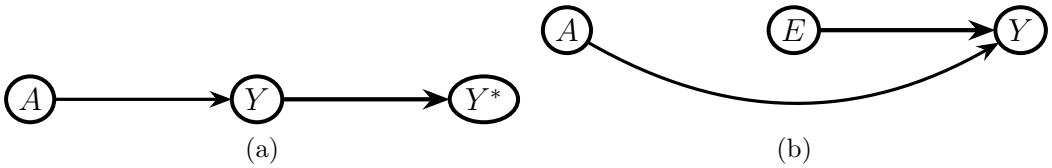

FIGURE S1. Simplified DAGs demonstrating the parallels described in eAppendix C. (a) Non-differential misclassification of the outcome. The assumption that outcome misclassification doesn't depend on exposure results in $A \perp\!\!\!\perp Y^* \mid Y$. The heavier arrow from Y to Y^* represents the perfect specificity assumption: $Y = 0 \implies Y^* = 0$. (b) The setting from the main text (simplified to remove common causes of E and Y). The no effect on exposure assumption results in $A \perp\!\!\!\perp E$. The heavier arrow from E to Y represents the exposure necessity assumption: $E = 0 \implies Y = 0$.

Appendix D. Identification of the CDE (4)

Identification results for the CDE are well-established when the exposure E is measured. To motivate our new result for settings where E is unmeasured, we first present these conventional identifiability conditions.

Assumption (Exposure exchangeability)**.**

$$E^a \perp\!\!\!\perp A \mid L,\ Y^{a,e=1} \perp\!\!\!\perp A \mid L \quad \textit{and} \quad Y^{a,e=1} \perp\!\!\!\perp E^a \mid L, A = a. \tag{S10}$$

Conditions (S10) are classical exchangeability condition, analogous to assumptions that are typically implemented to identify per protocol effects in trials, causal effects from observational data, and mediation effects. This assumption is stronger than exchangeability assumption (7). In particular, (S10) does not hold unless we measure common causes of Y and E, as illustrated in the SWIG in Figure 1d.

Assumption (Exposure positivity)**.**

(S11) $$P(A = a, E = 1 \mid L) > 0 \qquad \textit{for all } a \in \{0, 1\} \textit{ w.p.1.}$$

Assumption (Exposure consistency)**.**

(S12) $$\textit{If } A = a \textit{ then } E = E^a. \textit{ If } A = a \textit{ and } E = 1 \textit{ then } Y = Y^{a,e=1},$$

for all $a \in \{0, 1\}$.

When we impose conditions (S10)-(S12), the CDE can be expressed as a function of factual variables,

$$\mathbb{E}(Y^{a,e=1}) = \mathbb{E}\{\mathbb{E}(Y \mid E = 1, A = a, L)\}.$$

However, because we do not measure E, it is not possible to identify the CDE from our observed data; we cannot identify the term $\mathbb{E}(Y \mid E = 1, A = a, L)$ without measuring E. In particular, the absolute CDE cannot be identified unless we measure E. Nevertheless, our next proposition shows that the relative CDE conditional on L is point identified.

Proposition S1 (CDE conditional on L)**.** *Under conditions* (5)*,* (10) *and* (S10)-(S12)*, the relative CDE conditional on the baseline covariate* L *is*

$$\frac{\mathbb{E}(Y^{a=1,e=1} \mid L)}{\mathbb{E}(Y^{a=0,e=1} \mid L)} = \frac{\mathbb{E}(Y \mid A = 1, L)}{\mathbb{E}(Y \mid A = 0, L)}.$$

The proof is given in Appendix Appendix A. The following proposition relates the CECE within a subpopulation defined by L and the CDE.

Proposition S2 (CECE and CDE conditional on L)**.** *Under conditions* (5)*,* (7)-(10) *and* (S10)-(S12)*, the relative CECE given $L = l$ and the relative CDE conditional on the baseline covariate $L = l$ are equal, that is,*

$$\frac{\mathbb{E}(Y^{a=1} \mid E^{a=1} = 1, L = l)}{\mathbb{E}(Y^{a=0} \mid E^{a=0} = 1, L = l)} = \frac{\mathbb{E}(Y^{a=1,e=1} \mid L = l)}{\mathbb{E}(Y^{a=0,e=1} \mid L = l)} = \frac{\mathbb{E}(Y \mid A = 1, L = l)}{\mathbb{E}(Y \mid A = 0, L = l)}.$$

It is crucial that the covariate vector L in Proposition S1 and Proposition S2 is sufficient to adjust for confounding, i.e. to ensure that exposure exchangeability (S10) holds. Thus, identification of the conditional CDEs requires stronger assumptions compared to identification of the CECE. Although the conditional CECE can be defined and estimated within any set of baseline covariates, it is only interpretable as a conditional CDE when that set of covariates consists of those sufficient to adjust for confounding.

The marginal CDE is not point identified. Whereas the conditional relative CDE can be point identified under (5), (10) and (S10)-(S12), the marginal relative CDE is not identified without additional assumptions. To see this, consider a binary outcome $Y \in \{0, 1\}$. The marginal relative CDE can be expressed as a weighted average of conditional relative CDEs,[1]

$$\begin{aligned}\frac{\mathbb{E}(Y^{a=1,e=1})}{\mathbb{E}(Y^{a=0,e=1})} &= \sum_l \frac{\mathbb{E}(Y^{a=1,e=1} \mid L = l)}{\mathbb{E}(Y^{a=0,e=1} \mid L = l)} P(L = l \mid Y^{a=0,e=1} = 1) \\ &= \sum_l \frac{\mathbb{E}(Y \mid A = 1, L = l)}{\mathbb{E}(Y \mid A = 0, L = l)} P(L = l \mid Y^{a=0,e=1} = 1). \qquad \text{(S13)}\end{aligned}$$

[1]Following the collapsibility results for relative risks in e.g. Huitfeldt et al [**?**].

Using laws of probability and (S10)-(S12), $P(L = l \mid Y^{a=0,e=1} = 1)$ can be written as

$$P(L = l \mid Y^{a=0,e=1} = 1) = \frac{P(Y^{a=0,e=1} = 1 \mid L = l)P(L = l)}{P(Y^{a=0,e=1} = 1)}$$
$$= \frac{P(Y = 1 \mid E = 1, A = 0, L = l)P(L = l)}{\sum_l P(Y = 1 \mid E = 1, A = 0, L = l)P(L = l)},$$

which depends on probabilities conditional on $E = 1$ that are not estimable from observed data. However, we can point-identify the marginal CDE under the additional strong assumption that $\mathbb{E}(Y^{a=0,e=1}) = 1$, that is, the exposure deterministically causes the outcome if untreated. Then, $P(L = l \mid Y^{a=0,e=1} = 1) = P(L = l)$, and thus the marginal relative CDE is point identified by

$$\frac{\mathbb{E}(Y^{a=1,e=1})}{\mathbb{E}(Y^{a=0,e=1})} = \sum_l \frac{\mathbb{E}(Y \mid A = 1, L = l)}{\mathbb{E}(Y \mid A = 0, L = l)} P(L = l).$$

The marginal absolute PPE is point identified as

$$\mathbb{E}(Y^{a=1,e=0}) - \mathbb{E}(Y^{a=1,e=1}) = 1 - \frac{\mathbb{E}(Y^{a=1,e=1})}{\mathbb{E}(Y^{a=0,e=1})} = 1 - \sum_l \frac{\mathbb{E}(Y \mid A = 1, L = l)}{\mathbb{E}(Y \mid A = 0, L = l)} P(L = l).$$

Appendix E. Sensitivity analyses

*Proof of Proposition 3 from Section "*External Data and Sensitivity Analysis*".* We follow the same strategy as for the lower bound (S1) in the main text.

$$
\begin{aligned}
&\mathbb{E}(Y^{a=1} \mid E^{a=1}=1) - \mathbb{E}(Y^{a=0} \mid E^{a=0}=1) \\
=&\mathbb{E}(Y \mid E=1, A=0) - \mathbb{E}(Y \mid E=1, A=1) \quad \text{due to (7)} - \text{(9)} \\
=&\frac{\mathbb{E}(EY \mid A=0)}{P(E=1 \mid A=0)} - \frac{\mathbb{E}(EY \mid A=1)}{P(E=1 \mid A=1)} \quad \text{(laws of prob.)} \\
=&\frac{\mathbb{E}(Y \mid A=0)}{P(E=1 \mid A=0)} - \frac{\mathbb{E}(Y \mid A=1)}{P(E=1 \mid A=1)} \quad \text{due to (10)} \\
=&\frac{\mathbb{E}(Y \mid A=0)}{P(E=1 \mid A=0)} - \frac{\mathbb{E}(Y \mid A=1)}{P(E=1 \mid A=0)} \quad \text{due to (5)} \\
=&\frac{\mathbb{E}(Y \mid A=0)\mathbb{E}(Y \mid E=1, A=0)}{\mathbb{E}(Y \mid A=0)} - \frac{\mathbb{E}(Y \mid A=1)\mathbb{E}(Y \mid E=1, A=0)}{\mathbb{E}(Y \mid A=0)} \\
=&\mathbb{E}(Y \mid E=1, A=0)\left(1 - \frac{\mathbb{E}(Y \mid A=1)}{\mathbb{E}(Y \mid A=0)}\right),
\end{aligned}
$$

where we used exposure necessity (10) in the 5th equality, which implies that $P(Y = 1 \mid E = 1, A = 0)P(E = 1 \mid A = 0) = P(Y = 1 \mid A = 0)$ and that $\mathbb{E}(Y \mid A = 0) \geq \mathbb{E}(Y \mid A = 1)$. □

Proposition 3 motivates a sensitivity analysis and/or use of data from external sources; the investigator can include their background knowledge on $P(Y = 1 \mid E = 1, A = 0)$ – the probability of experiencing the outcome given exposure in the unvaccinated – along with the observed data on $\mathbb{E}(Y \mid A = a)$ to point identify the absolute CECE.

The 4th line of the proof of Proposition 3 motivates an alternative sensitivity analysis: the investigator can specify the marginal risk of being exposed to the infectious agent given no treatment, that is, $P(E = 1 \mid A = 0)$, and then point identify the risk difference.

Appendix F. R code to compute bounds for the aCECE and create Figure 3

`\footnotesize`

```
# install.packages(c("ggplot2", "ggrepel"))
library(ggplot2)

## uncomment these lines and dev.off() at the bottom to recreate the
## figure using tikz
## install.packages("tikzDevice")
# library(grid)
# library(tikzDevice)
#
# tikz(file = "sensitivity_params.tex",
#      standAlone = FALSE,
#      width = 5,
#      height = 5
# )

## Observed values
# P(Y = 1 | A = 1)
pYA1 <- 0.009
# P(Y = 1 | A = 0)
pYA0 <- 0.031

# function to compute the bounds of the aCECE
calc_aCECE <- function(pYA1, pYA0, pE = 1, pYE1A0 = 1) {
  lower_bound <- (pYA0 - pYA1) / pE
  upper_bound <- (1 - pYA1 / pYA0) * pYE1A0
  c(lower_bound = lower_bound, upper_bound = upper_bound)
}

# calculate the bounds using the observed values set above
calc_aCECE(pYA1 = pYA1, pYA0 = pYA0)

# given a certain Pr(Y = 1 | E = 1, A = 0), we can compute Pr(E = 1 | A = 0)
# from the observed P(Y = 1 | A = 0) and P(Y = 1 | A = 1) and vice versa
convert_params <- function(pYE1A0 = NULL, pE = NULL, pYA1, pYA0) {
  if (!is.null(pYE1A0) & !is.null(pE)) stop("Please specify only one of pYE1A0
  and pE")
  if (is.null(pYE1A0) & is.null(pE)) stop("Either pYE1A0 or pE must be specified")
  if (!is.null(pYE1A0)) {
    pE <- -((pYA1 - pYA0) / (1 - (pYA1 / pYA0))) / pYE1A0
    return(c(pE = pE))
  } else {
    pYE1A0 <- -((pYA1 - pYA0) / (1 - (pYA1 / pYA0))) / pE
    return(c(pYE1A0 = pYE1A0))
  }
}

# confirm that we can go back and forth
convert_params(pYE1A0 = 0.05, pYA1 = pYA1, pYA0 = pYA0)
```

```
convert_params(pE = 0.62, pYA1 = pYA1, pYA0 = pYA0)

# as soon as we specify one or the other of Pr(E = 1 | A = 0) or
# Pr(Y = 1 | E = 1, A = 0), the bounds are the same
# and the aCECE is point-identified
# here we are specifying Pr(Y = 1 | E = 1, A = 0) = 0.85 and using that to
# compute Pr(E = 1 | A = 0)
calc_aCECE(pYA1 = pYA1, pYA0 = pYA0,
           pE = convert_params(pYE1A0 = 0.85, pYA1 = pYA1, pYA0 = pYA0),
           pYE1A0 = 0.85)

# specify values to compute aCECE for
pE <- c(0.9,
        0.6,
        convert_params(pYE1A0 = 0.85, pYA1 = pYA1, pYA0 = pYA0))
pYE1A0 <- c(convert_params(pE = 0.9, pYA1 = pYA1, pYA0 = pYA0),
            convert_params(pE = 0.6, pYA1 = pYA1, pYA0 = pYA0),
            0.85)

# apply the function to each pair of values
aCECE <- mapply(FUN = calc_aCECE,
                pE = pE, pYE1A0 = pYE1A0,
                # the other parameters stay the same
                MoreArgs = list(pYA1 = pYA1, pYA0 = pYA0))

# the lower and upper bounds are the same
aCECE

# create dataframe of points to graph
points_to_highlight <- data.frame(x = pE, y = pYE1A0, aCECE = aCECE[1,],
                                  lbl = paste0("aCECE = ",
                                  scales::number(aCECE[1,])))

# plot the range of possible pairs of
# Pr(E = 1 | A = 0) and Pr(Y = 1 | E = 1, A = 0)
# with those values annotated with the corresponding aCECE
fig <- ggplot(data = points_to_highlight, aes(x, y, label = lbl)) +
  xlim(pYA0, 1) +
  geom_function(fun = convert_params, args = list(pYA1 = pYA1, pYA0 = pYA0)) +
  labs(x = "$\\Pr(E = 1 \\mid A= 0)$", y = "$\\Pr(Y = 1 \\mid E = 1, A = 0)$") +
  geom_point() +
  ggrepel::geom_text_repel() +
  theme_minimal() +
  theme(axis.line.y.left = element_line(color = "#565656"),
        axis.line.x.bottom = element_line(color = "#565656"))

print(fig)
```

```
# dev.off()
```