eAppendix 1

In this appendix, we provide an overview of identifiability conditions and identification results for the five causal effects described in the main text of the paper. To simplify the exposition, as in the main text, we consider a simplified setting with a time-fixed treatment and a binary outcome measured at a single time-point, and no losses to follow-up. Formal results about analogous estimands in more realistic settings (e.g., for failure-time outcomes) can be found elsewhere [1–4].

Data and notation: Let A denote an indicator for random assignment to treatment at baseline (1 for letrozole; 0 for gonadotropin), D an indicator for the competing event (1 for no live birth; 0 for live birth) and Y the indicator for the event of interest (1 for neonatal complications; 0 otherwise) at the end of follow-up (where D occurs before Y). Let Q denote the composite event (1 if no live birth or neonatal complications; 0 otherwise). Let L denote the measured baseline covariates, and U a set of unmeasured covariates that may exert effect on having a live born baby (D) and on neonatal complications (Y). We assume that the data are independent and identically distributed realizations of the random tuple (L, A, D, Y). Throughout, we use italic capital letters for random variables and corresponding lower case letters for their realizations. For example, L denotes baseline covariate random variable that takes values l in the set of possible values, $\mathcal{L}$.

We use superscripts to denote counterfactual variables. In particular, Y^a is the counterfactual indicator of the event of interest if the individual had, possibly contrary to fact, been assigned treatment a; D^a is the counterfactual indicator of the competing event if the individual had, possibly contrary to fact, been assigned treatment a.

Randomization and exchangeability assumptions: We assume that the trial generating the data is marginally randomized, so that

$$(Y^a, D^a, L) \perp\!\!\!\perp A.$$

By the decomposition and weak union properties of conditional independence [5, 6], the independence condition above implies several others. Here, we only list the implications that we use in derivations in this Appendix:

$$(Y^a, D^a, L) \perp\!\!\!\perp A \Longrightarrow \left\{ \begin{array}{c} Y^a \perp\!\!\!\perp A \\ D^a \perp\!\!\!\perp A \\ (Y^a, D^a) \perp\!\!\!\perp A \\ (Y^a, L) \perp\!\!\!\perp A \\ (Y^a, D^a) \perp\!\!\!\perp A | L \end{array} \right\}.$$

Causal Effect 1. The total effect of treatment on the composite event is defined as $\Pr[Q^{a=1} = 1] - \Pr[Q^{a=0} = 1]$. Assume that the following identifiability conditions hold [1]:

1. Exchaneability: $(Y^a, D^a) \perp\!\!\!\perp A$ for $a \in \{0, 1\}$.
2. Positivity: $\Pr[A = a] > 0$ for each $a \in \{0, 1\}$.
3. Consistency: if $A = a$, then $Y^a = Y$ and $D^a = D$ for each $a \in \{0, 1\}$.

The components of the total effect of treatment on the composite event are the risks $\Pr[Q^a = 1]$ of the composite event had everyone in the study population been assigned treatment $a \in \{0, 1\}$.

Under the above identifiability conditions, $\Pr[Q^a = 1]$ is identified as follows:

$$\begin{aligned}
\Pr[Q^a = 1] &= \Pr[Y^a = 1, D^a = 0] + \Pr[Y^a = 0, D^a = 1] \\
&= \Pr[Y^a = 1, D^a = 0|A] + \Pr[Y^a = 0, D^a = 1|A] \text{ (by exchangeability)} \\
&= \Pr[Y = 1, D = 0|A = a] + \Pr[Y = 0, D = 1|A = a] \text{ (by consistency)} \\
&= \Pr[Y = 1|D = 0, A = a] \Pr[D = 0|A = a] + \Pr[D = 1|A = a].
\end{aligned}$$

The last step in the derivation above uses the following fact: for each treatment a, we have

$$\begin{aligned}
\Pr[D = 1|A = a] &= \sum_{y=0}^{1} \Pr[Y = y, D = 1|A = a] \\
&= \Pr[Y = 0, D = 1|A = a] + \Pr[Y = 1, D = 1|A = a] \\
&= \Pr[Y = 0, D = 1|A = a].
\end{aligned}$$

Thus, the total effect of treatment on the composite event is identified as

$$\begin{aligned}
&\Pr[Q^a = 1] - \Pr[Q^{a=0} = 1] = \\
&\qquad \Pr[Y = 1|D = 0, A = 1] \Pr[D = 0|A = 1] + \Pr[D = 1|A = 1] \\
&\qquad\quad - \big\{ \Pr[Y = 1|D = 0, A = 0] \Pr[D = 0|A = 0] + \Pr[D = 1|A = 0] \big\}.
\end{aligned}$$

Causal Effect 2. The total effect of treatment on the event of interest is defined as $\Pr[Y^{a=1} = 1] - \Pr[Y^{a=0} = 1]$. Assume that the following identifiability conditions hold [1]:

1. Exchangeability: $Y^a \perp\!\!\!\perp A$ for $a \in \{0, 1\}$.
2. Positivity: $\Pr[A = a] > 0$ for $a \in \{0, 1\}$.
3. Consistency: if $A = a$, then $Y^a = Y$ for $a \in \{0, 1\}$.

The components of the total effect of treatment on the event of interest are the risks $\Pr[Y^a = 1]$ of the event of interest had everyone in the population been assigned to treatment $a \in \{0, 1\}$.

Under the above identifiability conditions, $\Pr[Y^a = 1]$ can be identified as follows:

$$\begin{aligned}
\Pr[Y^a = 1] &= \Pr[Y^a = 1 | A = a] \text{ (by exchangeability)} \\
&= \Pr[Y = 1 | A = a] \text{ (by consistency)} \\
&= \sum_{d=0}^{1} \Pr[Y = 1, D = d | A = a] \\
&= \Pr[Y = 1, D = 0 | A = a] + \Pr[Y = 1, D = 1 | A = a] \\
&= \Pr[Y = 1, D = 0 | A = a] \\
&= \Pr[Y = 1 | D = 0, A = a] \Pr[D = 0 | A = a].
\end{aligned}$$

Thus, the total effect of treatment on the event of interest is identified as

$$\begin{aligned}
\Pr[Y^{a=1} = 1] - \Pr[Y^{a=0} = 1] = \Pr[Y = 1 | D = 0, A = 1] \Pr[D = 0 | A = 1] \\
- \Pr[Y = 1 | D = 0, A = 0] \Pr[D = 0 | A = 0].
\end{aligned}$$

Causal Effect 3. Controlled direct effect of treatment on the event of interest. $\Pr[Y^{a=1,d=0}=1]-\Pr[Y^{a=0,d=0}=1]$. Compared to the total effect, controlled direct effect requires additional exchangeability conditions. In our setting, the identifiability assumptions [1] can be simplified to:

1. Exchangeability 1: $(Y^{a,d=0},L)\perp\!\!\!\perp A$ for each $a\in\{0,1\}$.
2. Exchangeability 2: $Y^{a,d=0}\perp\!\!\!\perp D\mid(L,A)$ for each $a\in\{0,1\}$.
3. Positivity: for each $a\in\{0,1\}$, if $f_L(l)\neq 0$, then $\Pr[A=a|L=l]>0$; if $f_{L,A}(l,a)\neq 0$, then $\Pr[D=0|L=l,A=a]>0$.
4. Consistency: if $A=a$ and $D=0$, then $Y^{a,d=0}=Y$

The components of the controlled direct effect of treatment on the event of interest are the risks $\Pr[Y^{a,d=0}=1]$ of the event of interest had everyone in the population been assigned to treatment $a\in\{0,1\}$ and had the competing event (no live births) been eliminated.

Under the above identifiability conditions, the risk under treatment a and elimination of the competing event can be identified by:

$$\begin{aligned}\Pr[Y^{a,d=0}=1]&=\sum_l \Pr[Y^{a,d=0}=1|L=l]f_L(l)\\&=\sum_l \Pr[Y^{a,d=0}=1|A=a,L=l]f_L(l)\text{ (by exchangeability 1)}\\&=\sum_l \Pr[Y^{a,d=0}=1|A=a,L=l,D=d]f_L(l)\text{ (by exchangeability 2)}\\&=\sum_l \Pr[Y=1|A=a,L=l,D=0]f_L(l)\text{ (by consistency).}\end{aligned}$$

Thus, risk difference of the direct effect of treatment on the event of interest had competing events been eliminated are:

$$\begin{aligned}\Pr[Y^{a=1,d=0}=1]-\Pr[Y^{a=0,d=0}=1]=&\sum_l \Pr[Y=1|A=1,L=l,D=0]f_L(l)\\&-\sum_l \Pr[Y=1|A=0,L=l,D=0]f_L(l).\end{aligned}$$

Causal Effect 4. The separable direct and indirect effects of treatment on the effect of interest are defined as $\Pr[Y^{a_Y=1,a_D} = 1] - \Pr[Y^{a_Y=0,a_D} = 1]$ and $\Pr[Y^{a_Y,a_D=1} = 1] - \Pr[Y^{a_Y,a_D=0} = 1]$. Identifying these effects requires the assumption that A can be decomposed into two components (A_Y and A_D) that exert effects through different causal pathways: A_Y affects the event of interest Y, and A_D affects competing events. In our setting, the assumptions described in reference [2] can be simplified as follows:

1. Exchangeability: $(Y^a, D^a) \perp\!\!\!\perp A|L$ for each $a \in \{0, 1\}$.

2. Consistency: if $A = a$, then $Y^a = Y$ and $D^a = D$, for each $a \in \{0, 1\}$.

3. Positivity: if $f_L(l) \neq 0$, then $\Pr[A = a|L = l] > 0$ for each $a \in \{0, 1\}$.
   If $f_{L,D}(l, 0) \neq 0$, then $\Pr[A = a|L = l, D = 0] > 0$ for each $a \in \{0, 1\}$.

4. Dismissible component condition 1:
   $\Pr[D^{a_Y=1,a_D} = 1|L = l] = \Pr[D^{a_Y=0,a_D} = 1|L = l]$ for $a_D \in \{0, 1\}$.

5. Dismissible component condition 2:
   $\Pr[Y^{a_Y,a_D=1} = 1|D^{a_Y,a_D=1} = 0, L = l] = \Pr[Y^{a_Y,a_D=0} = 1|D^{a_Y,a_D=0} = 0, L = l]$ for $a_Y \in \{0, 1\}$.

The components of separable direct effect of treatment on the event of interests are risks $\Pr[Y^{a_Y=1,a_D} = 1]$ of the event of interest had everyone in the study population been assigned to $A_Y = 1$ and $A_D = a_D$, where a_D can take values of 0 or 1.

Under the above assumptions, the risk under $a_Y = 1$ and a_D can be identified because

$$\begin{aligned}
\Pr[Y^{a_Y=1,a_D} = 1] &= \sum_l \Pr[Y^{a_Y=1,a_D} = 1|L = l] f_L(l) \\
&= \sum_l \sum_{d=0}^{1} \Pr[Y^{a_Y=1,a_D} = 1|D^{a_Y=1,a_D} = d, L = l] \Pr[D^{a_Y=1,a_D} = d|L = l] f_L(l) \\
&= \sum_l \Pr[Y^{a_Y=1,a_D} = 1|D^{a_Y=1,a_D} = 0, L = l] \Pr[D^{a_Y=1,a_D} = 0|L = l] f_L(l) \\
&= \sum_l \Pr[Y^{a_Y=1,a_D=1} = 1|D^{a_Y=1,a_D=1} = 0, L = l] \Pr[D^{a_Y=a_D,a_D} = 0|L = l] f_L(l) \\
&\quad \text{(by dismissible assumptions)} \\
&= \sum_l \Pr[Y^{a=1} = 1|D^{a=1} = 0, L = l] \Pr[D^{a=a_D} = 0|L = l] f_L(l) \\
&= \sum_l \Pr[Y^{a=1} = 1|D^{a=1} = 0, L = l, A = 1] \Pr[D^{a=a_D} = 0|L = l, A = a_D] f_L(l) \\
&\quad \text{(by exchangeability)} \\
&= \sum_l \Pr[Y = 1|D = 0, L = l, A = 1] \Pr[D = 0|L = l, A = a_D] f_L(l) \\
&\quad \text{(by consistency).}
\end{aligned}$$

The risks $\Pr[Y^{a_Y=0,a_D} = 1]$ of the event of interest had everyone in the study population been assigned to $A_Y = 0$ and $A_D = a_D$, where a_D can take values of 0 or 1, can be identified analogously.

Thus, the separable direct effect of treatment on the event of interest when A_D is set to a_D can be identified as follows:

$$\begin{aligned}
&\Pr[Y^{a_Y=1,a_D} = 1] - \Pr[Y^{a_Y=0,a_D} = 1] = \\
&\quad \sum_l \Pr[Y = 1|A = 1, L = l, D = 0] \Pr[D = 0|A = a_D, L = l] f_L(l) \\
&\qquad - \sum_l \Pr[Y = 1|A = 0, L = l, D = 0] \Pr[D = 0|A = a_D, L = l] f_L(l).
\end{aligned}$$

Similarly, the components of separable indirect effect of treatment on the event of interests are risks $\Pr[Y^{a_Y,a_D=1}=1]$ of the event of interest had everyone in the study population been assigned to $A_Y=a_Y$ and $A_D=1$, where a_Y can take values of 0 or 1.

The risk under a_Y and $a_D=1$ can be identified by:

$$\begin{aligned}
\Pr[Y^{a_Y,a_D=1}=1] &= \sum_l \Pr[Y^{a_Y,a_D=1}=1|L=l]f_L(l) \\
&= \sum_l \Pr[Y^{a_Y,a_D=1}=1|D^{a_Y,a_D=1}=0,L=l]\Pr[D^{a_Y,a_D=1}=0|L=l]f_L(l) \\
&= \sum_l \Pr[Y^{a_Y,a_D=a_Y}=1|D^{a_Y,a_D=a_Y}=0,L=l]\Pr[D^{a_Y=1,a_D=1}=0|L=l]f_L(l) \\
&\quad \text{(by dismissible assumptions)} \\
&= \sum_l \Pr[Y^{a=a_Y}=1|D^{a=a_Y}=0,L=l]\Pr[D^{a=1}=0|L=l]f_L(l) \\
&= \sum_l \Pr[Y^{a=a_Y}=1|D^{a=a_Y}=0,L=l,A=a_Y]\Pr[D^{a=1}=0|L=l,A=1]f_L(l) \\
&\quad \text{(by exchangeability)} \\
&= \sum_l \Pr[Y=1|D=0,L=l,A=a_Y]\Pr[D=0|L=l,A=1]f_L(l) \\
&\quad \text{(by consistency).}
\end{aligned}$$

The risks $\Pr[Y^{a_Y,a_D=0}=1]$ of the event of interest had everyone in the study population been assigned to $A_Y=a_Y$ and $A_D=0$, where a_Y can take values of 0 or 1, can be identified analogously.

Thus, the separable indirect effects when A_Y is set to a_Y can be identified by

$$\begin{aligned}
&\Pr[Y^{a_Y,a_D=1}=1]-\Pr[Y^{a_Y,a_D=0}=1]= \\
&\quad \sum_l \Pr[Y=1|A=a_Y,L=l,D=0]\Pr[D=0|A=1,L=l]f_L(l) \\
&\qquad -\sum_l \Pr[Y=1|A=a_Y,L=l,D=0]\Pr[D=0|A=0,L=l]f_L(l).
\end{aligned}$$

Causal Effect 5. The total (direct) effect in the principal stratum of always survivors is defined as $\Pr[Y^{a=1} = 1 | D^{a=1} = D^{a=0} = 0] - \Pr[Y^{a=0} = 1 | D^{a=1} = D^{a=0} = 0]$. There are several approaches (requiring different assumptions) for identifying the principal stratum effect. Here, we describe an approach for bounding [3] and an approach for point identifying the effect in the principal stratum of always survivors [4].

Bounding of the principal stratum effect [3] is possible under the following conditions:

1. Monotonicity: For all individuals, $D^{a=1} \geq D^{a=0}$, which implies that $\Pr[D^{a=1} = 0, D^{a=0} = 1] = 0$.
2. $\Pr[Y^{a=0} = 1 | D = 0, A = 1] - \Pr[Y^{a=0} = 1 | D = 0, A = 0] = \alpha \geq 0$.
3. Exchangeability: $(Y^a, D^a) \perp\!\!\!\perp A$ for each $a \in \{0, 1\}$.
4. Positivity: $\Pr[A = a, D = 0] > 0$ for each $a \in \{0, 1\}$.
5. Consistency: if $A = a$ and $D = 0$, then $Y^a = Y$ and $D^a = D$ for each $a \in \{0, 1\}$.

Using the above conditions, we have

$$\begin{aligned}
\Pr[Y^a = 1 | D = 0, A = 1] &= \Pr[Y^a = 1 | D^{a=1} = 0, A = 1] \text{ (by consistency)} \\
&= \frac{\Pr[Y^a = 1, D^{a=1} = 0 | A = 1]}{\Pr[D^{a=1} = 0 | A = 1]} \\
&= \frac{\Pr[Y^a = 1, D^{a=1} = 0]}{\Pr[D^{a=1} = 0]} \text{ (by exchangeability)} \\
&= \Pr[Y^a = 1 | D^{a=1} = 0] \\
&= \Pr[Y^a = 1 | D^{a=1} = 0, D^{a=0} = 0] \text{ (by monotonicity).}
\end{aligned}$$

The total effect (equal to the direct effect) of the event among the principal stratum can be written as

$$\begin{aligned}
&\Pr[Y^{a=1} | D^{a=1} = D^{a=0} = 0] - \Pr[Y^{a=0} | D^{a=1} = D^{a=0} = 0] \\
&\quad = \Pr[Y^{a=1} = 1 | D = 0, A = 1] - \Pr[Y^{a=0} = 1 | D = 0, A = 1] \text{ (by monotonicity)} \\
&\quad = \Pr[Y^{a=1} = 1 | D = 0, A = 1] - (\Pr[Y^{a=0} = 1 | D = 0, A = 0] + \alpha) \text{ (by assumption 2)} \\
&\quad = \Pr[Y = 1 | D = 0, A = 1] - \Pr[Y = 1 | D = 0, A = 0] - \alpha \text{ (by consistency)} \\
&\quad \leq \Pr[Y = 1 | D = 0, A = 1] - \Pr[Y = 1 | D = 0, A = 0].
\end{aligned}$$

Thus, upper bound for effect of treatment on event of interest among the principal stratum

is $\Pr[Y=1|D=0,A=1]-\Pr[Y=1|D=0,A=0]$, which is equal to the unadjusted estimate of event restricted to live births.

Point identification of the principal stratum effect is possible using a regression-based approach [4]: this approach does not require the monotonicity assumption, but requires a cross-world counterfactual independence condition and parametric assumptions about the data generating mechanism [4]. Specifically, we assume that the following conditions hold:

1. $Y^a \perp\!\!\!\perp D^{1-a} | (D^a=0, A=a, L, U)$ for $a \in \{0,1\}$.
2. $\text{logit}\Pr[D=0|A,U,L] = \gamma U + v(A,L)$, where $v(A,L)$ is an unrestricted function.
3. $\log\Pr[Y=1|D=0,A,L,U] = \beta_0 + \beta_1 A + U + b_l(L)$, where $b_l(L)$ is an unrestricted function.
4. $\mathrm{E}[U|A,L] = \mathrm{E}[U|L]$.
5. $\Delta_d \perp\!\!\!\perp (A,L) | D=d$ for $d \in \{0,1\}$, where $\Delta_d = U - \mathrm{E}[U|A,L,D=d]$ is a shift in the distribution of U.
6. Positivity: if $f_{L,A}(l,a) \neq 0$, then $\Pr[D=0|A=a,L=l] > 0$.
7. Consistency: if $A=a$ and $D=0$, then $Y^a = Y$ and $D^a = D$.

We observed that

$$
\begin{aligned}
&\mathrm{E}[Y|A=a,L,D=0] \\
&\quad = \mathrm{E}[\mathrm{E}[Y|A=a,L,U,D=0]|A=a,L,D=0] \\
&\quad = \mathrm{E}[\exp(\beta_0+\beta_1 a + U + b_l(L))|A=a,L,D=0] \text{ (by assumption 3)} \\
&\quad = \exp\{\beta_0+\beta_1 a + b_l(L)\}\mathrm{E}[\exp(U)|A=a,L,D=0] \\
&\quad = \exp\{\beta_0+\beta_1 a + b_l(L)\}\mathrm{E}[\exp(\Delta_d)|A=a,L,D=0]E[\exp(E[U|A,D,L])|A=a,L,D=0] \\
&\quad (\text{ using the model in assumption 5: } U = \Delta_d + \mathrm{E}[U|A,D,L]) \\
&\quad = \exp\{\beta_0+\beta_1 a + b_l(L)\}\mathrm{E}[\exp(\Delta_d)|D=0]\mathrm{E}[\exp(\mathrm{E}[U|A=a,D=0,L])|A=a,L,D=0] \\
&\quad (\text{ by assumption 5}) \\
&\quad = \exp\{\beta_0+\beta_1 a + b_l(L) + \mathrm{E}[U|A=a,L,D=0]\}\mathrm{E}[\exp(\Delta_d)|D=0]
\end{aligned}
$$

Next, we use the Theorem of [7] to rewrite $\mathrm{E}[U|A=a,L,D=0]$

$$
\begin{aligned}
&\mathrm{E}[U|A=a,L,D=0] \\
&\quad= \frac{\mathrm{E}[U\exp(\gamma U)|A=a,L,D=1]}{\mathrm{E}[\exp(\gamma U)|A=a,L,D=1]} \text{ (by (5), see below proof of intermediate step)} \\
&\quad= \frac{\partial}{\partial\gamma}\log\mathrm{E}[\exp(\gamma U)|A=a,L,D=1] \\
&\quad= \frac{\partial}{\partial\gamma}\log\{\mathrm{E}[\exp(\gamma\{\Delta_d + E[U|A=a,L,D=1]\})|A=a,L,D=1]\} \text{ (by assumption 5)} \\
&\quad= \frac{\partial}{\partial\gamma}\log\{\mathrm{E}[\exp(\gamma\Delta_d)|A=a,L,D=1]\exp(\gamma\mathrm{E}[U|A=a,L,D=1])\} \\
&\quad= \frac{\partial}{\partial\gamma}\{\gamma\mathrm{E}[U|A=a,L,D=1] + \log\mathrm{E}[\exp(\gamma\Delta_d)|A=a,L,D=1]\} \\
&\quad= \mathrm{E}[U|A=a,L,D=1] + \frac{\partial}{\partial\gamma}\log\mathrm{E}[\exp(\gamma\Delta_d)|A,L,D=1)] \\
&\quad= \mathrm{E}[U|A=a,L,D=1] + \frac{\partial}{\partial\gamma}\log\mathrm{E}[\exp(\gamma\Delta_d|D=1)] \text{ (by assumption 5)}
\end{aligned}
$$

therefore

$$
\begin{aligned}
&\mathrm{E}[U|A=a,L,D=0] - \mathrm{E}[U|A=a,L,D=1] \\
&\quad= \frac{\partial}{\partial\gamma}\log\mathrm{E}[\exp(\gamma\Delta_d|D=1)] \\
&\quad= \beta_{al}
\end{aligned} \tag{1}
$$

Furthermore, from the law of total expectation

$$
\begin{aligned}
&\mathrm{E}[U|A=a,L,D=0] \\
&\quad= \mathrm{E}[U|A=a,L] - \{\mathrm{E}[U|A=a,L,D=1] - \mathrm{E}[U|A=a,L,D=0]\}\Pr[D=1|A=a,L] \\
&\quad= \mathrm{E}[U|A=a,L] + \beta_{al}\Pr[D=1|A=a,L] \\
&\quad= \mathrm{E}[U|L] + \beta_{al}\Pr[D=1|A=a,L] \text{ (by assumption 4)}
\end{aligned} \tag{2}
$$

We therefore conclude that

$$
\begin{aligned}
&\mathrm{E}[Y|A=a,L,D=0] \\
&\quad= \exp\{\beta_0 + \beta_1 a + b_l(L) + \mathrm{E}[U|A=a,L,D=0]\}\mathrm{E}[\exp(\Delta_d)|D=0] \\
&\quad= \exp\{\beta_0 + \beta_1 a + b_l(L) + \mathrm{E}[U|L] + \beta_{al}\Pr[D=1|A,L]\}\mathrm{E}[\exp(\Delta_d)|D=0] \text{ (by (2))} \\
&\quad= \exp\{\beta_1 a + b_l^*(L) + \beta_{al}\Pr[D=1|A,L]\}
\end{aligned}
$$

where we define $b_l^*(L) = \beta_0 + b_l(L) + \mathrm{E}[U|L] + \log\mathrm{E}[\exp(\Delta_d)|D=0]$

Thus, it follows that

$$\log \Pr[Y=1|A,L,D=0] = \beta_1 A + b_l^* L + \beta_{al} P[D=1|A,L]$$

and we can conclude that the risk ratio comparing treatment $a=1$ versus $a=0$ in the principal stratum is $\exp(\beta_1)$.

Proof of the intermediate step: We will now show that

$$\mathrm{E}[U|A=a,L,D=0] = \frac{\mathrm{E}[U\exp(\gamma U)|A,L,D=1]}{\mathrm{E}[\exp(\gamma U)|A,L,D=1]}. \tag{3}$$

Using the model in assumption 2, we have that

$$\begin{aligned} & \operatorname{logit}\Pr[D=0|A,U,L] = \gamma U + v(A,L) \\ \Rightarrow \quad & \frac{\Pr[D=0|A,L,U]}{\Pr[D=1|A,L,U]} = \exp(\gamma U)\exp(v(A,L)). \end{aligned} \tag{4}$$

Starting with the right-hand-side of (3) and using the definition of expectation,

$$\begin{aligned} & \frac{\mathrm{E}[U\exp(\gamma U)|A,L,D=1]}{\mathrm{E}[\exp(\gamma U)|A,L,D=1]} \\ & \quad = \frac{\int u\exp(\gamma u) f(u|A,L,D=1)du}{\int \exp(\gamma u') f(u'|A,L,D=1)du'} \\ & \quad = \frac{\int u\exp(\gamma u)\frac{\Pr[D=0|A,L]}{\Pr[D=1|A,L]}\frac{\Pr[D=1|A,L,U=u]}{\Pr[D=0|A,L,U=u]} f(u|A,L,D=0)du}{\int \exp(\gamma u')\frac{\Pr[D=0|A,L]}{\Pr[D=1|A,L]}\frac{\Pr[D=1|A,L,U=u']}{\Pr[D=0|A,L,U=u']} f(u'|A,L,D=0)du'} \\ & \quad = \frac{\int u\exp(\gamma u)\frac{\Pr[D=1|A,L,U=u]}{\Pr[D=0|A,L,U=u]} f(u|A,L,D=0)du}{\int \exp(\gamma u')\frac{\Pr[D=1|A,L,U=u']}{\Pr[D=0|A,L,U=u']} f(u'|A,L,D=0)du'} \\ & \quad = \int u f(u|A,L,D=0)du \text{ (by (4))} \\ & \quad = \mathrm{E}[U|A,L,D=0] \end{aligned} \tag{5}$$

where we use the fact that,

$$f(u|A,L,D=1) = \frac{\Pr[D=0|A,L]}{\Pr[D=1|A,L]}\frac{\Pr[D=1|A,L,U=u]}{\Pr[D=0|A,L,U=u]} f(u|A,L,D=0).$$

References

[1] Jessica G Young, Mats J Stensrud, Eric J Tchetgen Tchetgen, and Miguel A Hernán. A causal framework for classical statistical estimands in failure time settings with competing events. *Statistics in Medicine*, 39(8):1199–1236, 2020.

[2] Mats J Stensrud, Jessica G Young, Vanessa Didelez, James M Robins, and Miguel A Hernán. Separable effects for causal inference in the presence of competing risks. *Journal of the American Statistical Association, In Press*, 2020.

[3] Yasutaka Chiba and Tyler J VanderWeele. A simple method for principal strata effects when the outcome has been truncated due to death. *American journal of epidemiology*, 173(7):745–751, 2011.

[4] Eric J Tchetgen Tchetgen, Kelesitse Phiri, and Roger Shapiro. A simple regression-based approach to account for survival bias in birth outcomes research. *Epidemiology*, 26(4):473–480, 2015.

[5] A Philip Dawid. Conditional independence in statistical theory. *Journal of the Royal Statistical Society: Series B (Methodological)*, 41(1):1–15, 1979.

[6] Judea Pearl. Causality: Models, reasoning and inference cambridge university press. *Cambridge, MA, USA,*, 9:10–11, 2000.

[7] Eric J Tchetgen Tchetgen and Kathleen E Wirth. A general instrumental variable framework for regression analysis with outcome missing not at random. *Biometrics*, 73(4):1123–1131, 2017.

eAppendix 2:

Separable Effects

To define the separable direct and indirect effects, we have made explicit assumptions about a decomposed version of the treatment.[1] As an illustration, we suggest one way to reason about a decomposition of letrozole in the AMIGOS trial. However, note that these conditions can be considerably relaxed.[2]

Suppose that letrozole, an aromatase inhibitor, can bind to aromatase-cytochrome P450 that is expressed in mother's ovaries and fetuses' tissues, and suppose that letrozole exerts effects on neonatal complications through two distinct pathways (Figure 1B). In the first pathway, letrozole binds to receptors of the fetus that interfere with fetal development (without having effects on achieving a live birth), which increases the risk of neonatal complications. This would occur if letrozole interrupts the normal aromatase function in fetal tissues that express P450 receptors (including fetal liver, intestine, skin, and brain) without having effects on achieving a live birth.[3] In the second pathway, letrozole binds to P450 receptors in the mother's ovaries and enhances single-follicle recruitment, which may affect whether a mother achieves a live birth but not the risk of neonatal complications. A possible justification of this pathway is the following: suppose that letrozole inhibits the mother's aromatization of androgen to estrogen and thus reduces circulating estrogen, which in turn increases FSH secretion and stimulates ovarian follicular growth.[4] If the affinity of letrozole to P450 receptors in the mother and the fetus could be removed selectively, then we can conceptualize a decomposed versions of letrozole: one component A_D (targeting the receptors of mother only) exerts effects on live birth

eAppendix 2:

Separable Effects

To define the separable direct and indirect effects, we have made explicit assumptions about a decomposed version of the treatment.[1] As an illustration, we suggest one way to reason about a decomposition of letrozole in the AMIGOS trial. However, note that these conditions can be considerably relaxed.[2]

Suppose that letrozole, an aromatase inhibitor, can bind to aromatase-cytochrome P450 that is expressed in mother's ovaries and fetuses' tissues, and suppose that letrozole exerts effects on neonatal complications through two distinct pathways (Figure 1B). In the first pathway, letrozole binds to receptors of the fetus that interfere with fetal development (without having effects on achieving a live birth), which increases the risk of neonatal complications. This would occur if letrozole interrupts the normal aromatase function in fetal tissues that express P450 receptors (including fetal liver, intestine, skin, and brain) without having effects on achieving a live birth.[3] In the second pathway, letrozole binds to P450 receptors in the mother's ovaries and enhances single-follicle recruitment, which may affect whether a mother achieves a live birth but not the risk of neonatal complications. A possible justification of this pathway is the following: suppose that letrozole inhibits the mother's aromatization of androgen to estrogen and thus reduces circulating estrogen, which in turn increases FSH secretion and stimulates ovarian follicular growth.[4] If the affinity of letrozole to P450 receptors in the mother and the fetus could be removed selectively, then we can conceptualize a decomposed versions of letrozole: one component A_D (targeting the receptors of mother only) exerts effects on live birth

and one component A_Y (targeting the receptors of fetus only) exerts effects on neonatal complications.

Suppose that we also could describe a decomposition of gonadotropin into a component that only exerts effects on neonatal complications (A_Y) and a component that only exerts effects on achieving a live birth (A_D). Then, we can define separable direct and indirect effects by considering a (hypothetical) four-arm trial in which both the component that affects neonatal complications (A_Y) and the component that affects live births (A_D) are randomly assigned.[1] The first separable direct effect is defined by the contrast of outcomes between the arm receiving both components of letrozole versus the arm receiving the component of letrozole that affects life birth and the component of gonadotropin that affects neonatal complications. Similarly, the second separable direct effect is defined by the contrast of outcomes between the arm receiving the component of gonadotropin that affects life births and the component of letrozole that affects neonatal complications versus the arm receiving both components of gonadotropin. The separable indirect effects are defined analogously.[1]

References:

1. Stensrud MJ, Young JG, Didelez V, Robins JM, Hernán MA. Separable Effects for Causal Inference in the Presence of Competing Risks. *Journal of the American Statistical Association, In Press* 2020.
2. Stensrud MJ, Hernán MA, Robins JM, Tchetgen Tchetgen EJ, Didelez V, Young JG. Generalized interpretation and identification of separable effects in competing event settings, Forthcoming, 2020.
3. Toda K, Simpson ER, Mendelson CR, Shizuta Y, Kilgore MW. Expression of the gene encoding aromatase cytochrome P450 (CYP19) in fetal tissues. *Mol Endocrinol* 1994;**8**(2):210-7.
4. Pavone ME, Bulun SE. Clinical review: The use of aromatase inhibitors for ovulation induction and superovulation. *J Clin Endocrinol Metab* 2013;**98**(5):1838-44.