

Target Trial Emulation in Pregnancy Research

Louisa Smith

2026-10-30

Introduction

pdf

Today's goals

- Understand the target trial framework and why it's useful in reproductive and perinatal research
- Explore different types of pregnancy research questions that may be addressed with this approach
- Discover some intuition behind the clone-censor-weighting approach through a numeric example

Why pregnancy research is challenging

- Complex timing issues (exposure, outcomes, competing events)
- Immortal time bias is pervasive
- Multiple individuals (pregnant person, fetus, child)
- Need for clarity about *what* we're estimating



One solution

The **target trial framework** forces us to be explicit about who, what, when, and how (I guess where too, but not usually as much of a problem!)

One problem: time zero

Many observational studies don't have a clear "time zero" when treatment assignment occurs

Example: Comparing pregnancy outcomes in:

- People who were vaccinated during pregnancy
- People who weren't vaccinated during pregnancy

When are they "assigned" to exposure groups?

Immortal time bias

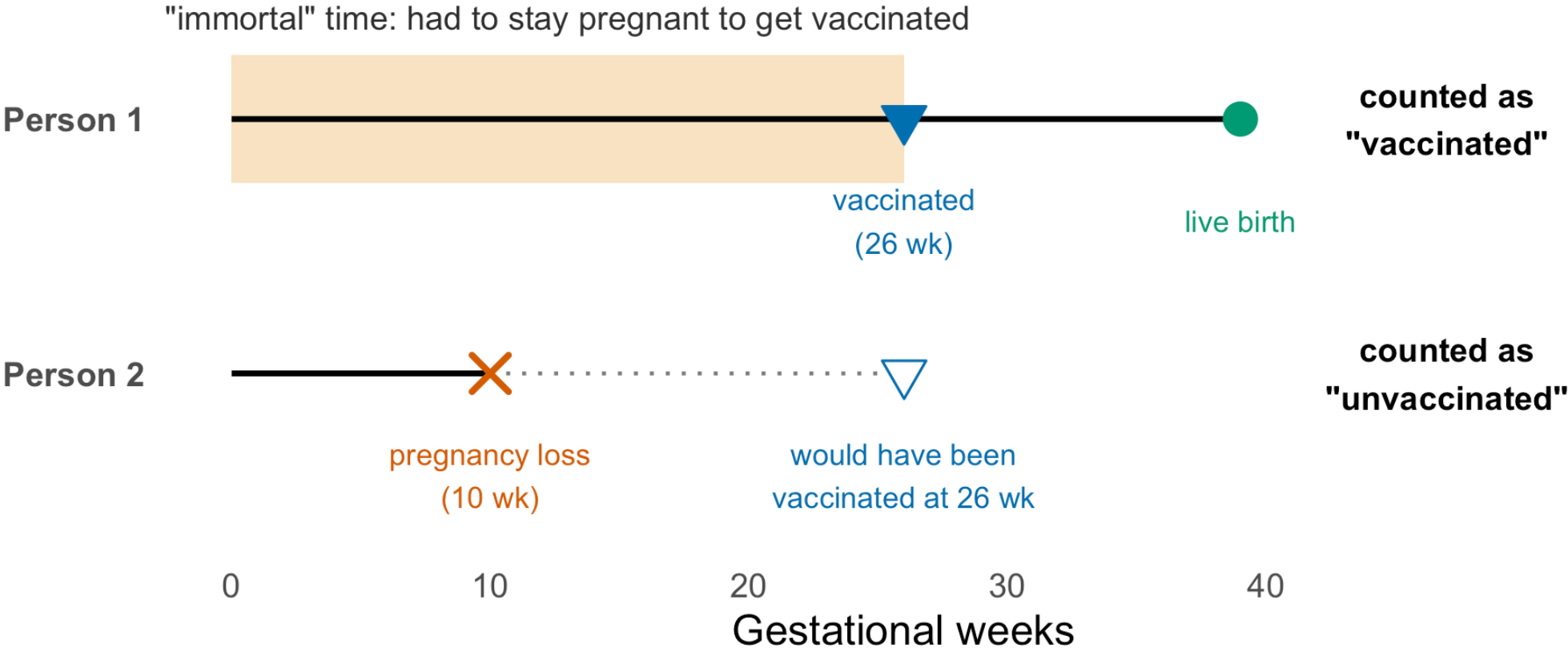
If we define groups based on *what actually happened* during pregnancy:

- “Exposed” group = those who got vaccinated at some point
- “Unexposed” group = those who never got vaccinated

The exposed group had to *survive long enough* (e.g. remain pregnant) to get vaccinated!

- This is a problem whenever the outcome depends on time (i.e., not just survival)

Immortal time



What is a target trial?

A hypothetical randomized trial that would answer your causal question if it could be conducted

Note

The target trial is a *design* concept, not an analysis method. It has guided study design in epidemiology for decades but recently popularized as an explicit framework ([Hernán and Robins 2016](#)).

Why a trial?

Randomized trials have clear advantages for causal inference:

- Randomization at baseline
 - Not the case in observational data no matter what we do, but we can try with good confounder measurement and reasonable eligibility criteria
- Stringent eligibility criteria
 - Everyone who enters the study has equipoise for the treatment strategies being compared

Why a trial?

Randomized trials have clear advantages for causal inference:

- Clear time zero
 - Everyone is assigned to treatment and starts follow-up at the same time
 - We *can* make this happen in observational data with careful design
- Well-defined treatment strategies
 - In order to give participants their assigned treatment, people have to have rules to follow!
 - We can define these rules in observational data too

Why a *target* trial?

We know we can't run the randomized trial we want to conduct to answer our causal question (lack of resources, unethical to randomize, impossible to provide certain treatments/exposures, too many years of follow-up needed, too many treatment strategies to compare, etc.)

- But we can design it hypothetically
- And then try to emulate it as closely as possible with observational data

Emulating a target trial

The observational study should be designed so as to match up with this trial as closely as possible

Warning

Don't jump straight to emulation without carefully thinking through the trial, though it can be helpful to think ahead. Compromises in emulation should be explicit and justified. One recent paper on this topic: Hernán et al. ([2026](#)).

Essential components

1. Eligibility criteria
2. Treatment strategies
3. Assignment procedures
4. Follow-up period
5. Outcome(s)
6. Causal contrast(s)
7. Statistical analysis plan

Note

Recently published guidelines for reporting target trial emulations detail these components: Cashin et al. ([2025](#))

Eligibility criteria: the “who”

Besides making for a clearer question with more practical implications, eligibility criteria can help address confounding in the emulation by ensuring all included have a reasonable chance of getting the treatment strategies being compared*

- We might exclude people with contraindications to treatment, or those who would never consider it
- This often means defining pregnancy status and gestational age at time zero carefully

*Always need *positivity*: everyone included must have some non-zero probability of

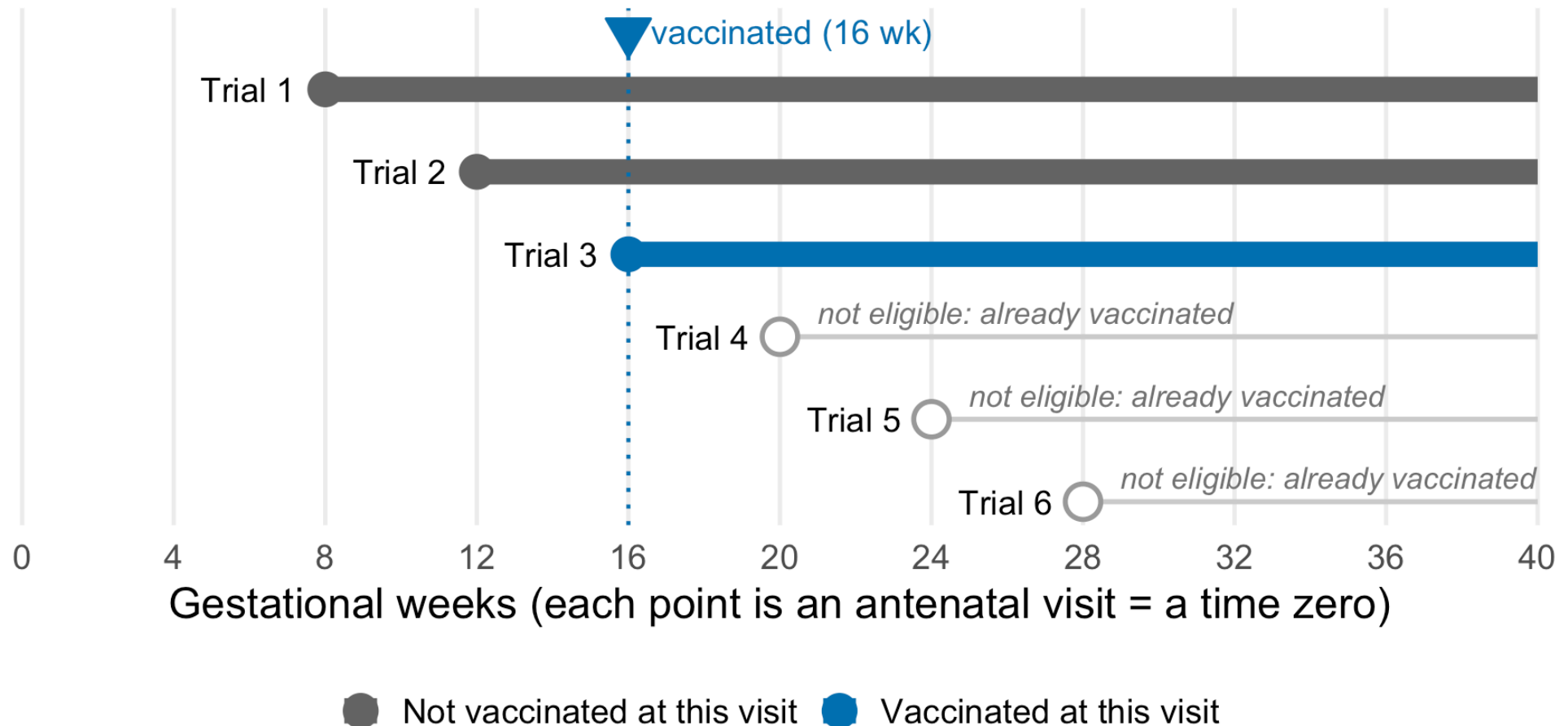
Time zero: the “when”

The eligibility criteria also define *when* people enter the study

- Time zero occurs when people meet eligibility criteria (and in a trial, agree to be randomized)
- We could imagine scenarios where people meet eligibility repeatedly over time (e.g., at every antenatal care visit)
 - We can take this into account when emulating

Repeated eligibility: a sequence of trials

Eligibility at each antenatal visit includes **no prior vaccination**

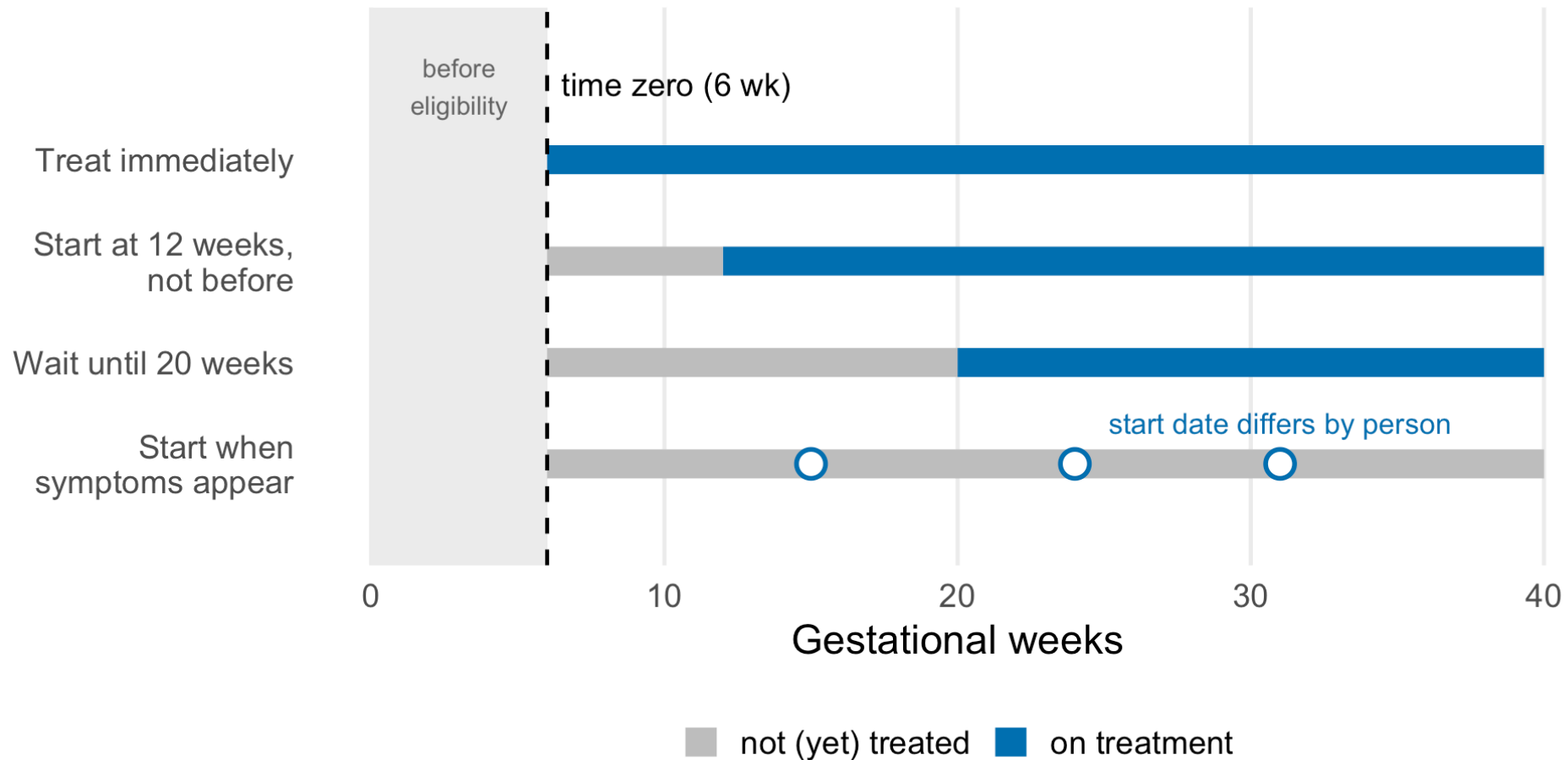


Treatment strategies: the “what” and “how”

Each strategy represents an intervention we could imagine putting in motion **at time zero** for a given treatment arm:

- Immediately upon randomization, tell everyone to get treatment (e.g., a vaccination)
- At 6 weeks gestation (time zero), tell everyone to start treatment at 12 weeks gestation but not before
- Tell everyone to wait until 20 weeks to start treatment
- Tell everyone to start treatment when symptoms appear

Multiple treatment strategies with one time zero



This is easier for some causal questions than others

- Pharmaceutical interventions with fixed timing (e.g., vaccination at week 32 vs. no vaccination)
- Procedure at some known clinical event (e.g., cerclage at diagnosis of short cervix vs. no cerclage)
- Comparisons of two treatments with the same indication



Tip

It's helpful to read through existing randomized trials on similar questions to see how they defined these components, see clinicaltrials.gov for ideas!

Benchmarking to a real trial

PF-07302048 (BNT162 RNA-Based COVID-19 Vaccines)

Protocol C4591015

Final Protocol Amendment 5, 08 March 2022



**A PHASE 2/3, PLACEBO-CONTROLLED, RANDOMIZED, OBSERVER-BLIND
STUDY TO EVALUATE THE SAFETY, TOLERABILITY, AND
IMMUNOGENICITY OF A SARS-COV-2 RNA VACCINE CANDIDATE
(BNT162b2) AGAINST COVID-19 IN HEALTHY PREGNANT WOMEN 18 YEARS
OF AGE AND OLDER**

Pharmaceutical example (Zidan et al. (2025))

Components	Target trial
Causal question	What is the effectiveness of SARS-CoV-2 mRNA vaccine BNT162b2 when taken during pregnancy?
Eligibility criteria	<p>Inclusion criteria:</p> <ol style="list-style-type: none">1. Healthy women ≥ 18 years of age who are between 24 0/7 and 34 0/7 weeks' gestation on the day of planned vaccination, with an uncomplicated, singleton pregnancy.2. Healthy participants determined by medical history, physical examination, and clinical judgment to be appropriate for inclusion in the study.3. Documented negative HIV antibody test. <p>Exclusion criteria:</p> <ol style="list-style-type: none">1. Other medical or psychiatric condition including recent (within the past year) or active suicidal ideation/behavior.2. Previous clinical or microbiological diagnosis of COVID-19.3. Participants with known or suspected immunodeficiency.

Components	Target trial
	<p>4. Bleeding diathesis or condition associated with prolonged bleeding.</p> <p>5. Previous vaccination with any COVID-19 vaccine.</p> <p>6. Current alcohol abuse or illicit drug use.</p> <p>7. Participants who receive treatment with immunosuppressive therapy.</p>
Treatment strategies	<p>1. Two vaccination doses; the first immediately after randomization</p> <p>2. No SARS-CoV-2 vaccination until the end of pregnancy</p>
Assignment procedures	1:1 randomization into the two treatment arms, stratified by gestational week
Follow-up	Since randomization and up to 1 month post delivery
Outcome	SARS-CoV-2 infection as determined by positive PCR test or clinical COVID-19 diagnosis
Causal contrast	Incidence rate ratios; intention-to-treat (one vaccine dose) and per-protocol (two doses)

Other types of questions

It may feel weird to design a target trial for other types of causal questions

- Unethical/impossible to randomize
 - e.g., harmful exposures, social determinants of health

It's worth thinking through anyway to make sure you are clear about your causal question of interest (you don't have to publish it as a "target trial"!)

Non-pharmaceutical example (Smith et al. (2022))

Components	Target trial
Causal question	What is the effect of COVID-19 infection during pregnancy on preterm delivery?
Eligibility criteria	1. Pregnant individuals with gestational age 12-36 weeks. 2. No known previous SARS-CoV-2 infection 3. No previous vaccination for COVID-19
Treatment strategies	1. Symptomatic COVID-19 within a week after enrollment. 2. No SARS-CoV-2 infection for the rest of the pregnancy.
Assignment procedures	Randomization at enrollment, stratified by gestational age (in weeks).
Follow-up	Patients are followed from the time of COVID-19 testing or enrollment (time zero) until delivery, loss to follow-up, or administrative end of follow-up.
Outcome	Preterm delivery, defined as delivery before 37 completed weeks of gestation.

Components	Target trial
Causal contrast	Intention-to-treat effect on the risk ratio and risk difference scales for each gestational week (time zero)

Notes on these components and things to think about in emulation (hopefully not compromises to the integrity of the target trial)

Components	Target trial
Eligibility criteria	<ul style="list-style-type: none">• Based only on pre-baseline characteristics• Generally requires pre-baseline observation window
Treatment strategies	<ul style="list-style-type: none">• Won't be able to emulate actual placebo or blinding (can assign no treatment if realistic)• Some people must have “adhered” to the treatment strategy
Assignment procedures	<ul style="list-style-type: none">• Randomization (within levels of confounders) is always an assumption• Assignment “happens” as soon as someone meets eligibility criteria
Follow-up	<ul style="list-style-type: none">• Monitoring for the outcome throughout follow-up (e.g., SARS-CoV-2 testing) may need to be part of the treatment strategy
Outcome	<ul style="list-style-type: none">• Outcome ascertainment can't be blinded in emulation
Causal contrast	<ul style="list-style-type: none">• Intention-to-treat effect makes sense when “most” of the treatment happens immediately upon randomization

Emulating target trials with clone-censor weighting

This example is somewhat based on an example about comparing duration of treatment in Hernán ([2018](#))

Does vaccination timing in pregnancy affect probability of live birth?

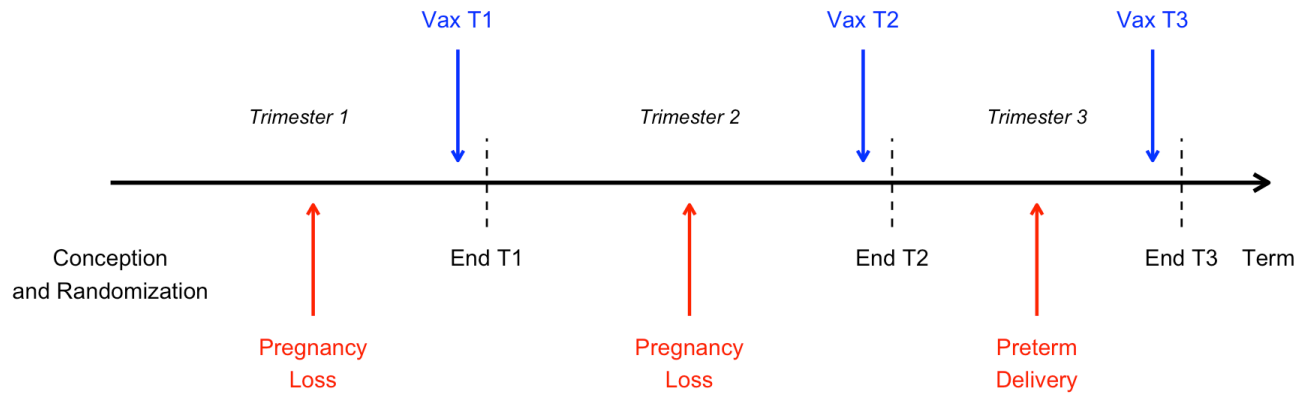
Eligibility: Unvaccinated, at/soon after conception

Treatment strategies:

- Strategy 0: Never vaccinate during pregnancy
- Strategy 1: Vaccinate in trimester 1 only
- Strategy 2: Vaccinate in trimester 2 only
- Strategy 3: Vaccinate in trimester 3 only

Outcome: Live birth (yes/no)

Pregnancy timeline



Note

We are simplifying things by assuming vaccination happens at the end of a trimester, after any pregnancy losses. In reality we need to think about the trimester as a window of time, and that vaccination could happen at any point in that window. Literature on *grace periods* describes how to handle this in emulation.

Randomized trial data

16 pregnant people randomly assigned to 4 strategies:

Person	Assigned strategy	Loss T1	Vax T1	Loss T2	Vax T2	Preterm	Vax T3	Term birth	Live birth
A	0	0	0	0	0	0	0	1	1
B	0	0	0	0	0	1	-	-	1
C	0	0	0	1	-	-	-	-	0
D	0	1	-	-	-	-	-	-	0
E	1	0	1	0	0	0	0	1	1
F	1	0	1	0	0	1	-	-	1
G	1	0	1	1	-	-	-	-	0
H	1	1	-	-	-	-	-	-	0
I	2	0	0	0	1	0	0	1	1
J	2	0	0	0	1	1	-	-	1

Person	Assigned strategy	Loss T1	Vax T1	Loss T2	Vax T2	Preterm	Vax T3	Term birth	Live birth
K	2	0	0	1	-	-	-	-	0
L	2	1	-	-	-	-	-	-	0
M	3	0	0	0	0	0	1	1	1
N	3	0	0	0	0	1	-	-	1
O	3	0	0	1	-	-	-	-	0
P	3	1	-	-	-	-	-	-	0

Randomized trial results

By assigned strategy:

Assigned strategy	N	Live births	Probability
0	4	2	0.5
1	4	2	0.5
2	4	2	0.5
3	4	2	0.5

All strategies have 50% live birth rate (we are operating in a situation where the **null hypothesis** of no effect of vaccination at any time is true)

Moving to observational data

In observational data, we don't see the assigned strategy.

We only see what actually happened:

- When (if) people got vaccinated
- When pregnancy losses occurred
- Whether there was a live birth and when

Let's classify people by **observed** vaccination status and timing...

Observed data

Same 16 people, but now we don't know their assigned strategy:

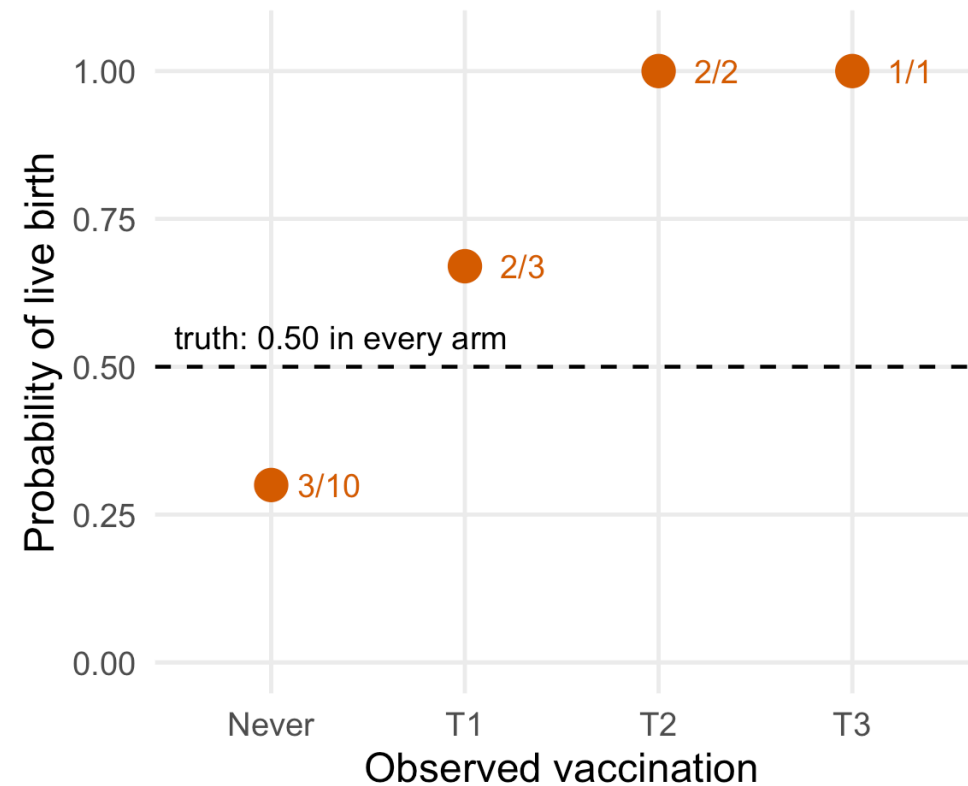
Person	Observed treatment	Loss T1	Vax T1	Loss T2	Vax T2	Preterm	Vax T3	Term birth	Live birth
A	0 (Never)	0	0	0	0	0	0	1	1
B	0 (Never)	0	0	0	0	1	-	-	1
C	0 (Never)	0	0	1	-	-	-	-	0
D	0 (Never)	1	-	-	-	-	-	-	0
E	1 (Vax T1)	0	1	0	0	0	0	1	1
F	1 (Vax T1)	0	1	0	0	1	-	-	1
G	1 (Vax T1)	0	1	1	-	-	-	-	0
H	0 (Never)	1	-	-	-	-	-	-	0
I	2 (Vax T2)	0	0	0	1	0	0	1	1
J	2 (Vax T2)	0	0	0	1	1	-	-	1

Person	Observed treatment	Loss T1	Vax T1	Loss T2	Vax T2	Preterm	Vax T3	Term birth	Live birth
K	0 (Never)	0	0	1	-	-	-	-	0
L	0 (Never)	1	-	-	-	-	-	-	0
M	3 (Vax T3)	0	0	0	0	0	1	1	1
N	0 (Never)	0	0	0	0	1	-	-	1
O	0 (Never)	0	0	1	-	-	-	-	0
P	0 (Never)	1	-	-	-	-	-	-	0

Naive analysis: by achieved vaccination

Classify by when they actually got vaccinated:

Observed vaccination	N	Live births	Probability
0 (Never)	10	3	0.30
1 (Vax T1)	3	2	0.67
2 (Vax T2)	2	2	1.00
3 (Vax T3)	1	1	1.00



The problem: immortal time bias!

Later vaccination appears highly protective!

But: People who got vaccinated later had to **survive** to that point

- All of the pregnancy losses get assigned to the “Never” or “Vax T1” groups
- By the time people are classified as “Vax T2” or “Vax T3”, they have already survived those earlier periods

Think like a randomized trial

In a randomized trial, people are assigned to strategies at time zero – even if they don't get treatment (by choice, not surviving long enough, etc.), they are analyzed in their assigned group*

*In an intention-to-treat analysis. Even under randomized assignment, an “as-treated”

Immortal time bias

Generally the not-treated group will underestimate the true risk, and the treated group will overestimate it (the later treated, or longer duration required, the more the bias):

Strategy	True probability	Naive estimate	Bias
0 (Never)	0.50	0.30	↓
1 (Vax T1)	0.50	0.67	↑
2 (Vax T2)	0.50	1.00	↑↑
3 (Vax T3)	0.50	1.00	↑↑

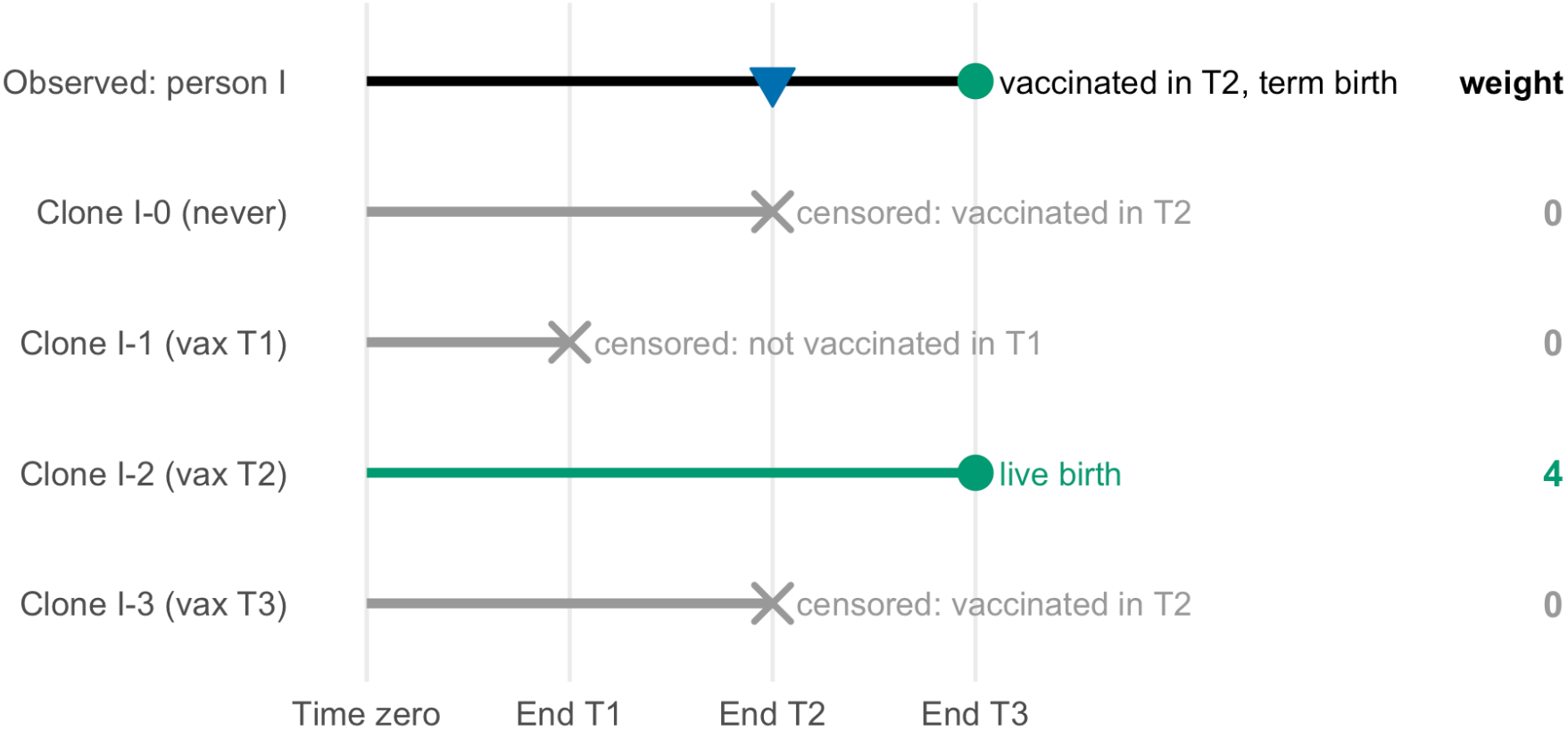
This makes treatment appear to reduce risk when there is actually no effect (or if there were a true effect of treatment, this might mask it)

Emulation via clone-censor-weighting

Pretend you have a randomized trial in which everyone is assigned to all strategies at time zero:

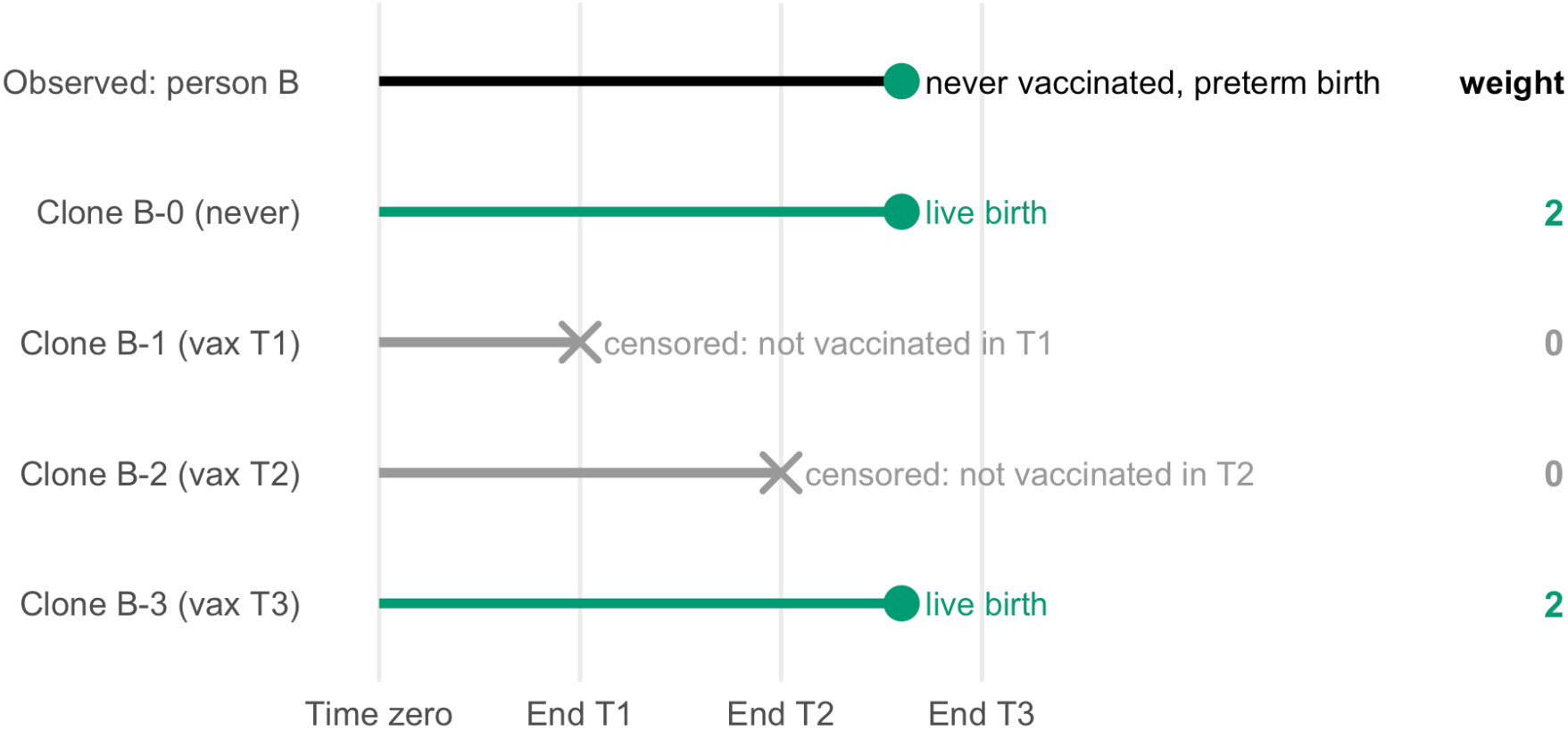
1. **Clone** everyone to all compatible strategies
2. **Censor** clones when they deviate from assigned strategy
3. **Weight** to correct for selection bias from censoring

One person, four clones (person I)



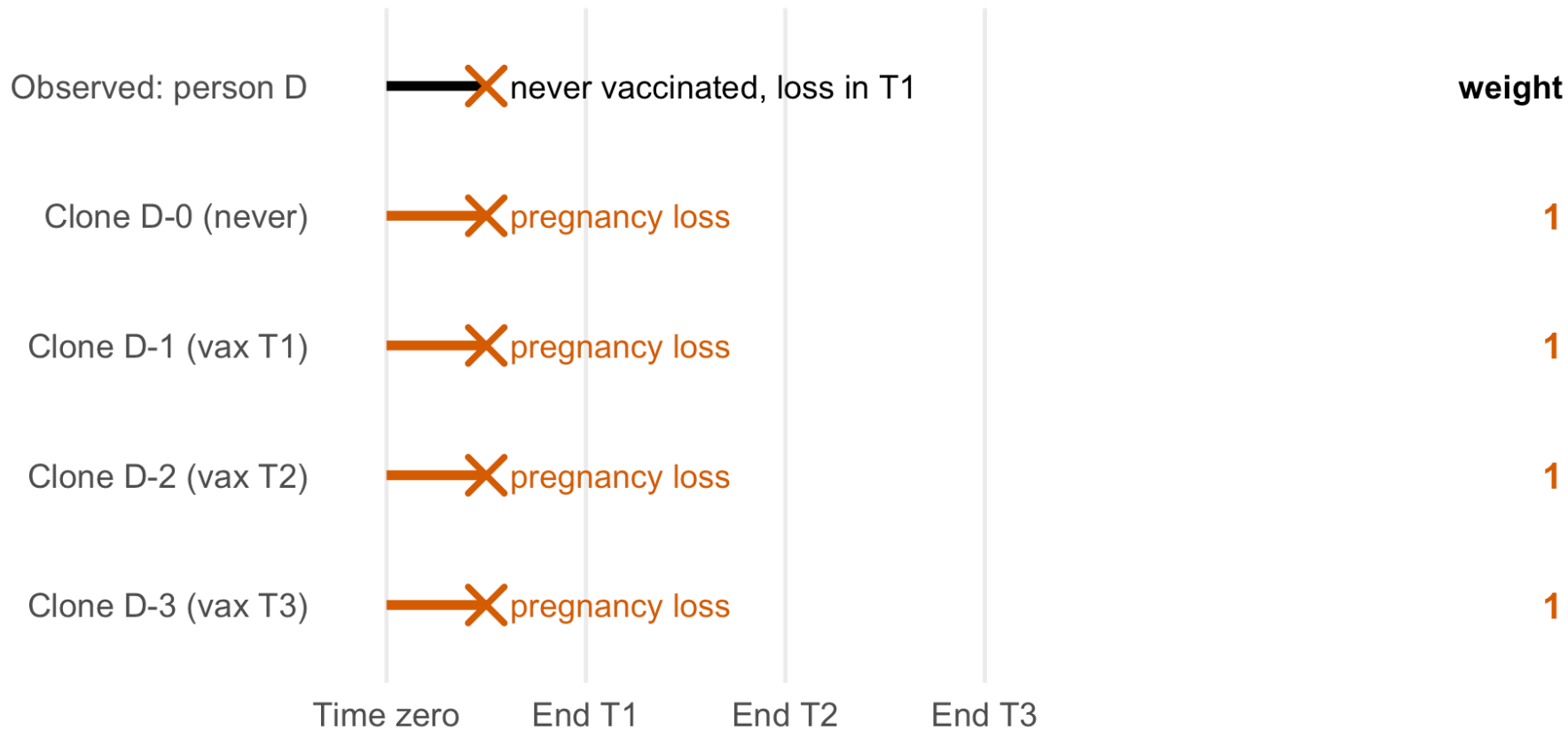
Uncensored clones are weighted up to stand in for censored clones like them

One person, four clones (person B)



The preterm birth comes before any T3 vaccination, so B follows strategies 0 and 3

One person, four clones (person D)



A T1 loss comes before anyone could deviate, so D counts in all four arms

Step 1: Cloning

For each person, create clones for all treatment strategies

Person	Assigned strategy	Loss T1	Vax T1	Loss T2	Vax T2	Preterm	Vax T3	Term birth	Live birth
A-0	0	0	0	0	0	0	0	1	1
B-0	0	0	0	0	0	1	-	-	1
C-0	0	0	0	1	-	-	-	-	0
D-0	0	1	-	-	-	-	-	-	0
E-0	0	0	1	0	0	0	0	1	1
F-0	0	0	1	0	0	1	-	-	1
G-0	0	0	1	1	-	-	-	-	0
H-0	0	1	-	-	-	-	-	-	0
I-0	0	0	0	0	1	0	0	1	1
J-0	0	0	0	0	1	1	-	-	1

Person	Assigned strategy	Loss T1	Vax T1	Loss T2	Vax T2	Preterm	Vax T3	Term birth	Live birth
K-0	0	0	0	1	-	-	-	-	0
L-0	0	1	-	-	-	-	-	-	0
M-0	0	0	0	0	0	0	1	1	1
N-0	0	0	0	0	0	1	-	-	1
O-0	0	0	0	1	-	-	-	-	0
P-0	0	1	-	-	-	-	-	-	0

Treatment strategy Vax T1

Person	Assigned strategy	Loss T1	Vax T1	Loss T2	Vax T2	Preterm	Vax T3	Term birth	Live birth
A-1	1	0	0	0	0	0	0	1	1
B-1	1	0	0	0	0	1	-	-	1
C-1	1	0	0	1	-	-	-	-	0
D-1	1	1	-	-	-	-	-	-	0
E-1	1	0	1	0	0	0	0	1	1
F-1	1	0	1	0	0	1	-	-	1
G-1	1	0	1	1	-	-	-	-	0
H-1	1	1	-	-	-	-	-	-	0
I-1	1	0	0	0	1	0	0	1	1
J-1	1	0	0	0	1	1	-	-	1
K-1	1	0	0	1	-	-	-	-	0

Person	Assigned strategy	Loss T1	Vax T1	Loss T2	Vax T2	Preterm	Vax T3	Term birth	Live birth
L-1	1	1	-	-	-	-	-	-	0
M-1	1	0	0	0	0	0	1	1	1
N-1	1	0	0	0	0	1	-	-	1
O-1	1	0	0	1	-	-	-	-	0
P-1	1	1	-	-	-	-	-	-	0

And so on...

Step 2: Censoring

Censor clones when their observed data becomes incompatible with assigned strategy:

- Strategy 0 (never): censor if vaccinated in T1, T2, or T3
- Strategy 1 (vax T1): censor if *not* vaccinated in T1
- Strategy 2 (vax T2): censor if vaccinated in T1 or *not* vaccinated in T2
- Strategy 3 (vax T3): censor if vaccinated in T1 or T2 or *not* vaccinated in T3

If there is a pregnancy loss in T1, do not censor afterward—we don't know whether they would have gotten vaccinated or not (can contribute to multiple strategies)

Censoring

Practice censoring

Person	Assigned strategy	Loss T1	Vax T1	Loss T2	Vax T2	Preterm	Vax T3	Term birth
A		0	0	0	0	0	0	1
B		0	0	0	0	1	-	-
C		0	0	1	-	-	-	-
D		1	-	-	-	-	-	-
E		0	1	0	0	0	0	1
F		0	1	0	0	1	-	-
G		0	1	1	-	-	-	-
H		1	-	-	-	-	-	-
I		0	0	0	1	0	0	1
J		0	0	0	1	1	-	-
K		0	0	1	-	-	-	-
L		1	-	-	-	-	-	-
M		0	0	0	0	0	1	1
N		0	0	0	0	1	-	-
O		0	0	1	-	-	-	-
P		1	-	-	-	-	-	-

Picture
Data

All 64 clones after censoring

	Assigned strategy			
	0 (never)	1 (vax T1)	2 (vax T2)	3 (vax T3)
A	term birth	censored T1	censored T2	censored T3
B	preterm birth	censored T1	censored T2	preterm birth
C	loss T2	censored T1	loss T2	loss T2
D	loss T1	loss T1	loss T1	loss T1
E	censored T1	term birth	censored T1	censored T1
F	censored T1	preterm birth	censored T1	censored T1
G	censored T1	loss T2	censored T1	censored T1
H	loss T1	loss T1	loss T1	loss T1
I	censored T2	censored T1	term birth	censored T2
J	censored T2	censored T1	preterm birth	censored T2
K	loss T2	censored T1	loss T2	loss T2
L	loss T1	loss T1	loss T1	loss T1
M	censored T3	censored T1	censored T2	term birth
N	preterm birth	censored T1	censored T2	preterm birth
O	loss T2	censored T1	loss T2	loss T2
P	loss T1	loss T1	loss T1	loss T1

■ Live birth ■ Pregnancy loss ■ Censored

Step 3: Weighting

Selection bias introduced by censoring must be corrected

Use **inverse probability weighting**:

- Weight = $1 / (\text{probability of remaining uncensored})$
 - This would be conditional on current values of covariates if we had them
- Transfers weight from censored to uncensored observations

Calculating probability of remaining uncensored

This varies over time, and can be calculated as the product of interval-specific probabilities:

$$\text{Prob(uncensored at time } t) = \prod_{k=0}^t \text{Prob(uncensored at } k \mid \text{uncensored at } k - 1)$$

That is, the probability of still being uncensored at the end of T3 is:

the probability of not being censored in T1

times the probability of not being censored in T2 (given not censored in T1)

times the probability of not being censored in T3 (given not censored in T1 or T2)

Person	Assigned strategy	Loss T1	Vax T1	Loss T2	Vax T2	Preterm	Vax T3	Term birth
A	0	0	0	0	0	0	0	1
B	0	0	0	0	0	1	-	-
C	0	0	0	1	-	-	-	-
D	0	1	-	-	-	-	-	-
E	0	0	1	0	0	0	0	1
F	0	0	1	0	0	1	-	-
G	0	0	1	1	-	-	-	-
H	0	1	-	-	-	-	-	-
I	0	0	0	0	1	0	0	1
J	0	0	0	0	1	1	-	-
K	0	0	0	1	-	-	-	-
L	0	1	-	-	-	-	-	-
M	0	0	0	0	0	0	1	1
N	0	0	0	0	0	1	-	-
O	0	0	0	1	-	-	-	-
P	0	1	-	-	-	-	-	-

 uncensored
 censoring event
 outcome used

interval probability
 of not being censored
 weight calculation

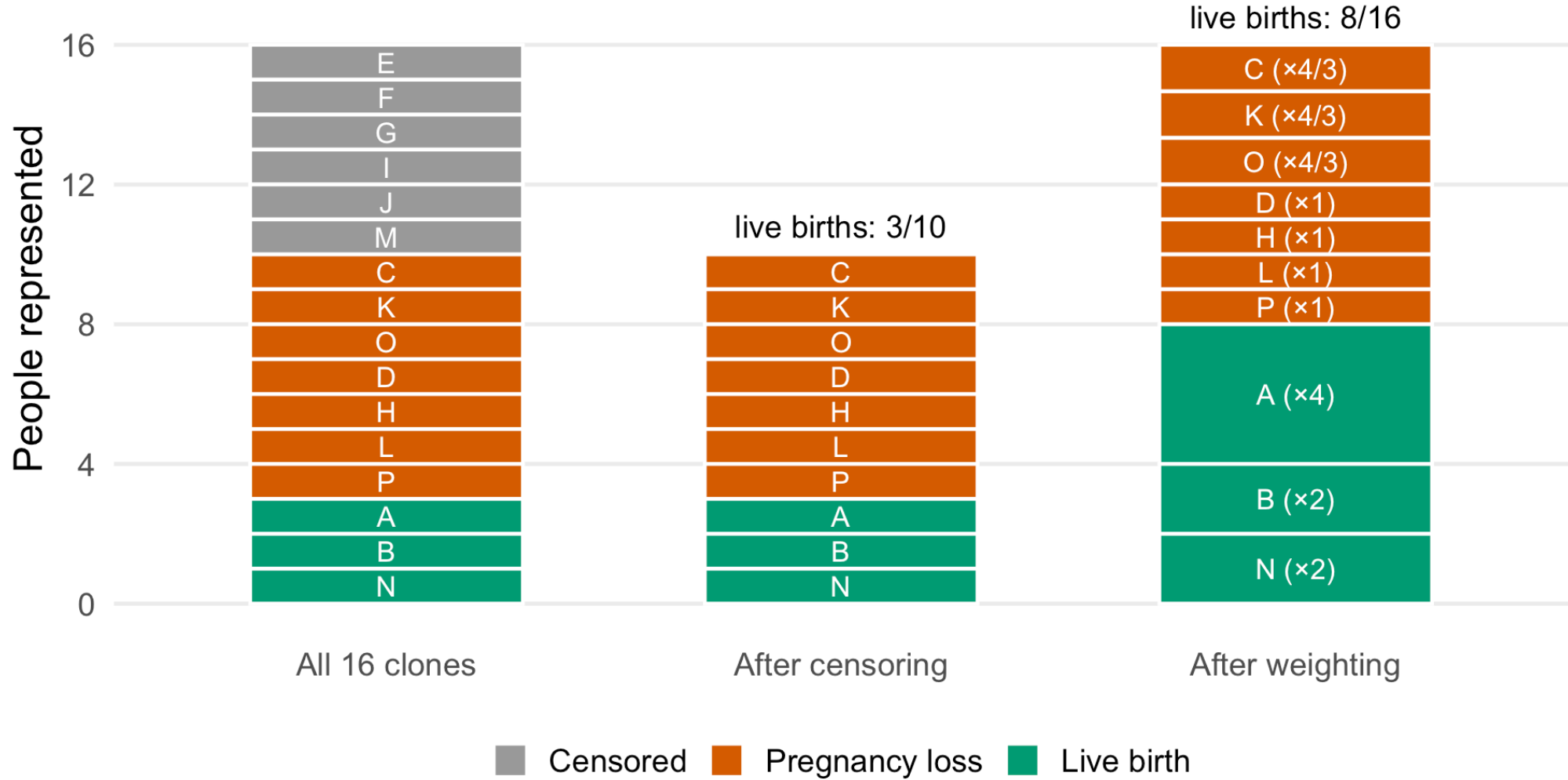
$$\frac{9}{9+3} \quad \frac{4}{4+2} \quad \frac{1}{1+1}$$

$$12/9 \rightarrow 12/9 \times 6/4 \rightarrow 12/9 \times 6/4 \times 2/1 \rightarrow$$

outcome weight

$$\text{weighted live births} = \frac{(4 \times 1 \times 0) + (3 \times \frac{4}{3} \times 0) + (2 \times 2 \times 1) + (4 \times 1 \times 1)}{16} = 0.5$$

Where the weight goes: strategy 0



Person	Assigned strategy	Loss T1	Vax T1	Loss T2	Vax T2	Preterm	Vax T3	Term birth
A		0	0	0	0	0	0	1
B		0	0	0	0	1	-	-
C		0	0	1	-	-	-	-
D		1	-	-	-	-	-	-
E		0	1	0	0	0	0	1
F		0	1	0	0	1	-	-
G		0	1	1	-	-	-	-
H		1	-	-	-	-	-	-
I		0	0	0	1	0	0	1
J		0	0	0	1	1	-	-
K		0	0	1	-	-	-	-
L		1	-	-	-	-	-	-
M		0	0	0	0	0	1	1
N		0	0	0	0	1	-	-
O		0	0	1	-	-	-	-
P		1	-	-	-	-	-	-

 uncensored
 censoring event
 outcome used

interval probability
of not being censored
weight calculation

outcome weight

$$\begin{aligned}
 &\frac{3}{3+9} \quad \frac{1}{1} \quad \frac{1}{1} \\
 &12/3 \rightarrow 12/3 \times 1/1 \rightarrow 12/3 \times 1/1 \times 1/1 \rightarrow
 \end{aligned}$$

1

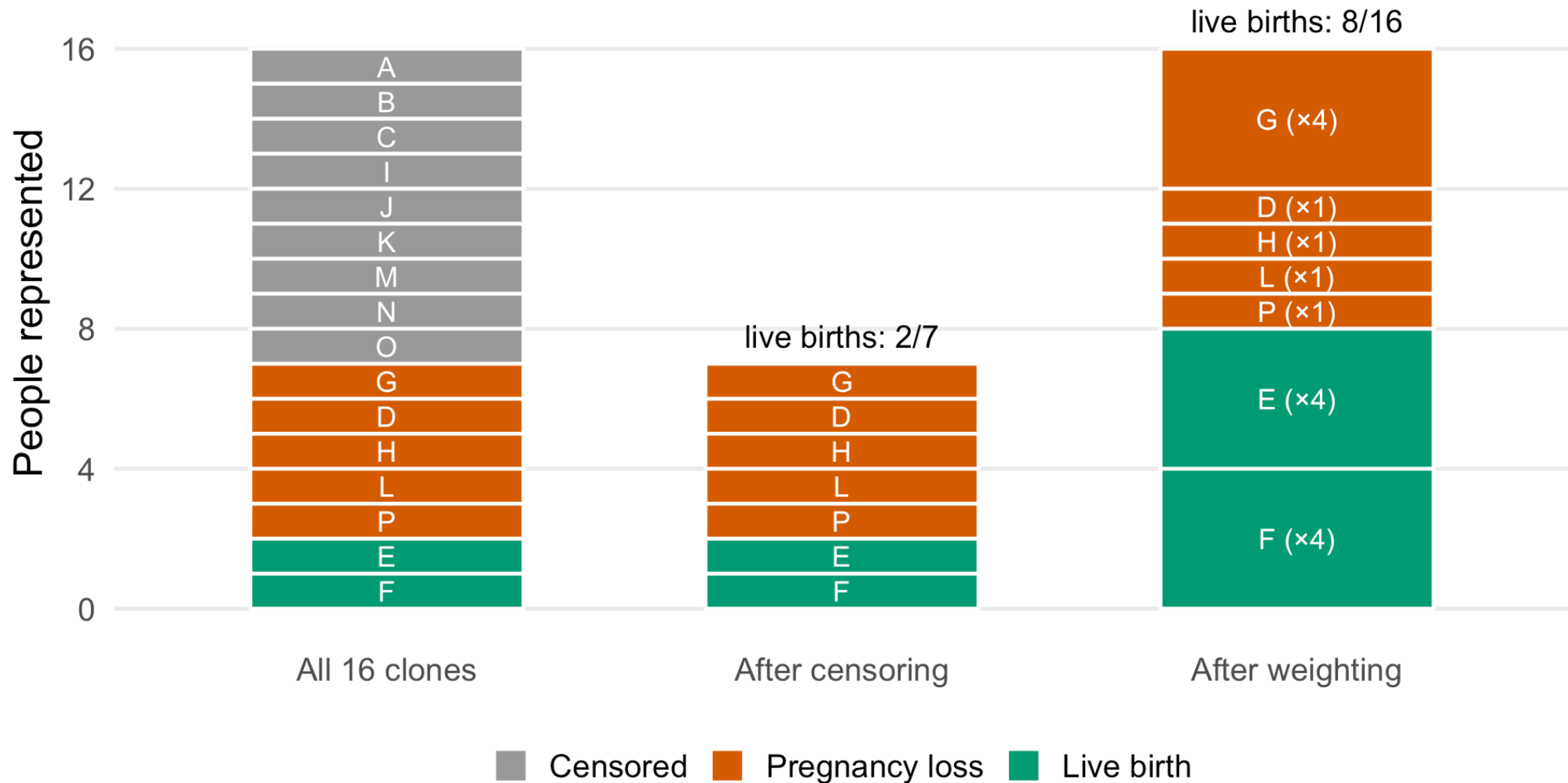
4

4

4

$$\text{weighted live births} = [(4 \times 1 \times 0) + (1 \times 4 \times 0) + (1 \times 4 \times 1) + (1 \times 4 \times 1)] / 16 = 0.5$$

Where the weight goes: strategy 1



Person	Assigned strategy	Loss T1	Vax T1	Loss T2	Vax T2	Preterm	Vax T3	Term birth
A	2	0	0	0	0	0	0	1
B	2	0	0	0	0	1	-	-
C	2	0	0	1	-	-	-	-
D	2	1	-	-	-	-	-	-
E	2	0	1	0	0	0	0	1
F	2	0	1	0	0	1	-	-
G	2	0	1	1	-	-	-	-
H	2	1	-	-	-	-	-	-
I	2	0	0	0	1	0	0	1
J	2	0	0	0	1	1	-	-
K	2	0	0	1	-	-	-	-
L	2	1	-	-	-	-	-	-
M	2	0	0	0	0	0	1	1
N	2	0	0	0	0	1	-	-
O	2	0	0	1	-	-	-	-
P	2	1	-	-	-	-	-	-

 uncensored
 censoring event
 outcome use2

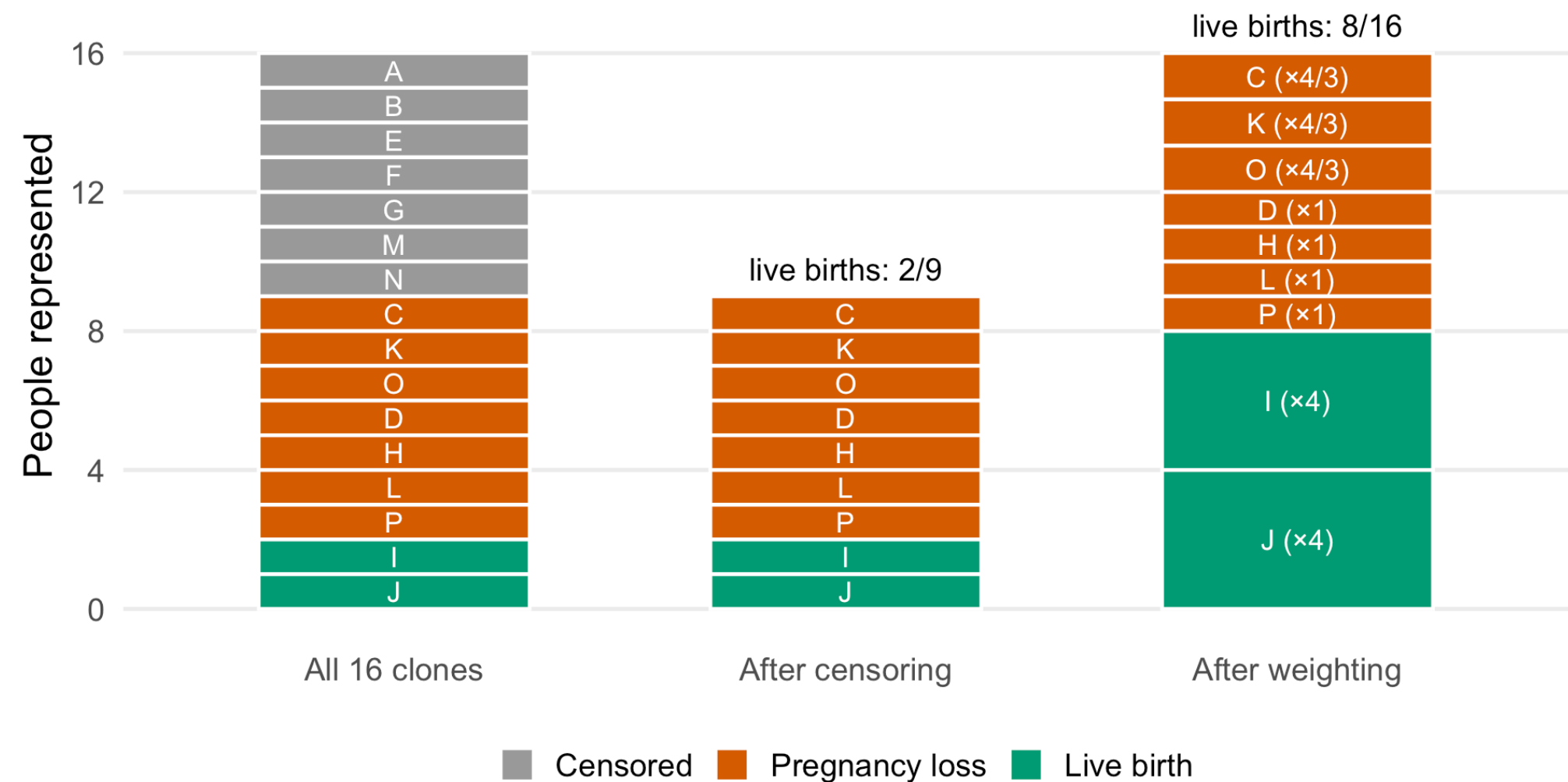
interval probability
of not being censored
weight calculation

outcome weight

$$\begin{aligned}
 &\frac{9}{9+3} \quad \frac{2}{2+4} \quad \frac{1}{1} \\
 &12/9 \longrightarrow 12/9 \times 6/2 \longrightarrow 12/9 \times 6/2 \times 1/4
 \end{aligned}$$

weighted live births = $[(4 \times 1 \times 0) + (3 \times 4/3 \times 0) + (1 \times 4 \times 1) + (1 \times 4 \times 1)] / 16 = 0.5$

Where the weight goes: strategy 2



Person	Assigned strategy	Loss T1	Vax T1	Loss T2	Vax T2	Preterm	Vax T3	Term birth
A	3	0	0	0	0	0	0	1
B	3	0	0	0	0	1	-	-
C	3	0	0	1	-	-	-	-
D	3	1	-	-	-	-	-	-
E	3	0	1	0	0	0	0	1
F	3	0	1	0	0	1	-	-
G	3	0	1	1	-	-	-	-
H	3	1	-	-	-	-	-	-
I	3	0	0	0	1	0	0	1
J	3	0	0	0	1	1	-	-
K	3	0	0	1	-	-	-	-
L	3	1	-	-	-	-	-	-
M	3	0	0	0	0	0	1	1
N	3	0	0	0	0	1	-	-
O	3	0	0	1	-	-	-	-
P	3	1	-	-	-	-	-	-

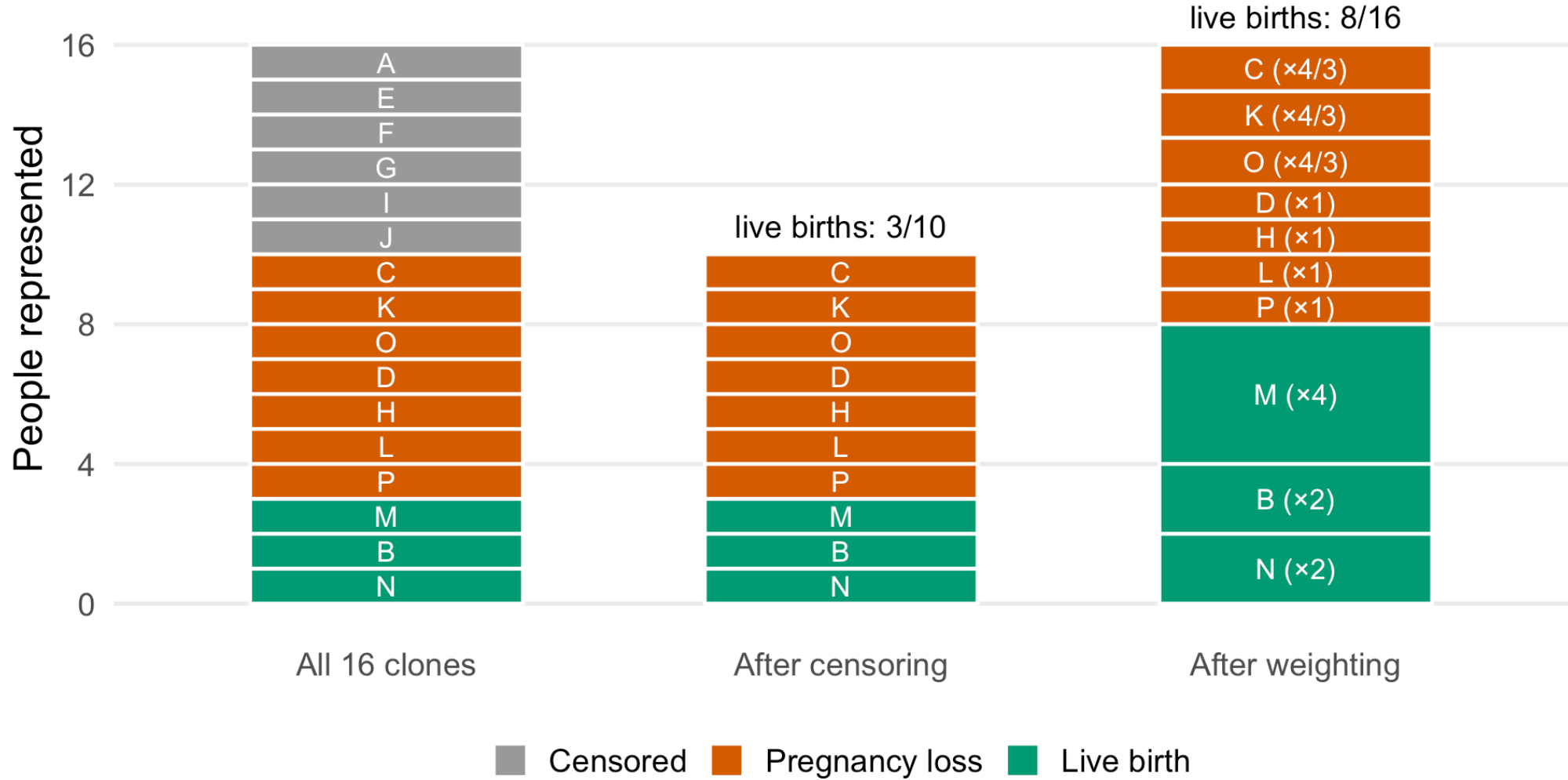
 uncensored
 censoring event
 outcome used

interval probability
of not being censored
weight calculation

outcome weight

$$\begin{aligned}
 & \frac{9}{9+3} \quad \frac{4}{4+2} \quad \frac{1}{1+2} \\
 & \frac{12}{9} \longrightarrow \frac{12}{9} \times \frac{6}{4} \longrightarrow \frac{12}{9} \times \frac{6}{4} \times \frac{2}{1} \\
 & \text{outcome weight: } 1 \quad \frac{4}{3} \quad 2 \quad 4 \\
 & \text{weighted live births} = \frac{(4 \times 1 \times 0) + (3 \times \frac{4}{3} \times 0) + (2 \times 2 \times 1) + (4 \times 1 \times 1)}{16} = 0.5
 \end{aligned}$$

Where the weight goes: strategy 3



Original Data

```
1 library(tidyverse)
2
3 trial_data <- read_csv("ccw-example-data.csv")
4 trial_data
```

A tibble: 16 × 10

	person	assigned	loss_t1	vax_t1	loss_t2	vax_t2	preterm	vax_t3	term	livebirth
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	A	0	0	0	0	0	0	0	1	1
2	B	0	0	0	0	0	1	NA	NA	1
3	C	0	0	0	1	NA	NA	NA	NA	0
4	D	0	1	NA	NA	NA	NA	NA	NA	0
5	E	1	0	1	0	0	0	0	1	1
6	F	1	0	1	0	0	1	NA	NA	1
7	G	1	0	1	1	NA	NA	NA	NA	0
8	H	1	1	NA	NA	NA	NA	NA	NA	0
9	I	2	0	0	0	1	0	0	1	1
10	J	2	0	0	0	1	1	NA	NA	1
11	K	2	0	0	1	NA	NA	NA	NA	0
12	L	2	1	NA	NA	NA	NA	NA	NA	0
13	M	3	0	0	0	0	0	1	1	1
14	N	3	0	0	0	0	1	NA	NA	1

Step 1: Cloning

Each person is cloned into 4 copies (one for each vaccination strategy: 0, 1, 2, 3)

```
1 cloned_data <- trial_data %>%  
2   crossing(strategy = 0:3) %>%  
3   relocate(strategy, .after = person)  
4  
5 cloned_data %>%  
6   count(strategy)
```

A tibble: 4 × 2

	strategy	n
	<int>	<int>
1	0	16
2	1	16
3	2	16
4	3	16

16 people × 4 strategies = 64 rows

Step 2: Censoring

```
1 censored_data <- cloned_data %>%
2   mutate(
3     # T1: only at risk if survived T1
4     censored_t1 = case_when(
5       loss_t1 == 1 ~ NA, # Already had outcome
6       strategy %in% c(0, 2, 3) & vax_t1 == 1 ~ TRUE, # Deviated by vaccination
7       strategy == 1 & vax_t1 == 0 ~ TRUE, # Deviated by not vaccinating
8       .default = FALSE # Followed strategy
9     ),
10
11    # T2: only at risk if uncensored and unvaccinated at T1 and survived T2
12    censored_t2 = case_when(
13      is.na(censored_t1) | censored_t1 ~ NA, # Already censored or had outcome
14      loss_t2 == 1 ~ NA, # Had outcome at T2
15      vax_t1 == 1 ~ NA, # Already had vax at T1
16      strategy %in% c(0, 3) & vax_t2 == 1 ~ TRUE, # Deviated by vaccination
17      strategy == 2 & vax_t2 == 0 ~ TRUE, # Deviated by not vaccinating
18      .default = FALSE # Followed strategy
```

Censoring Summary

Different numbers contribute to each strategy:

```
1 censored_data %>%
2   group_by(strategy) %>%
3   summarise(
4     n_total = n(),
5     censored_t1 = sum(censored_t1 == TRUE, na.rm = TRUE),
6     censored_t2 = sum(censored_t2 == TRUE, na.rm = TRUE),
7     censored_t3 = sum(censored_t3 == TRUE, na.rm = TRUE),
8     total_censored = sum(censored),
9     uncensored = sum(!censored)
10  )
```

A tibble: 4 × 7

	strategy	n_total	censored_t1	censored_t2	censored_t3	total_censored	uncensored
	<int>	<int>	<int>	<int>	<int>	<int>	<int>
1	0	16	3	2	1	6	10
2	1	16	9	0	0	9	7
3	2	16	3	4	0	7	9
4	3	16	3	2	1	6	10

Step 3a: Calculate Censoring Probabilities

Probability of vaccination can be used to calculate interval-specific censoring probabilities

1. Set up the data so that people who aren't eligible to be censored at a given timepoint don't contribute (have **NA** for vaccination status and/or subset to those not previously censored or vaccinated)
2. Model the probability of treatment (i.e., a propensity score model!)

Vaccination in T1 = censoring (strategy 0, 2, and 3) or not (strategy 1)

```
1 trial_data |>
2   filter(!is.na(vax_t1)) |>
3   pull(vax_t1, name = person)
```

A	B	C	E	F	G	I	J	K	M	N	O
0	0	0	1	1	1	0	0	0	0	0	0

```
1 mod_vax_t1 <- glm(vax_t1 ~ 1, data = trial_data, family = binomial())
2 p_vax_t1 <- predict(mod_vax_t1, type = "response")[1] # all have same predi
3 p_vax_t1
```

1
0.25

Vaccination in T2 = censoring (strategy 0 and 3) or not (2)

(For strategy 1, already censored if not vaccinated in T1 so no one “at risk for” censoring here)

```
1 trial_data |>
2   filter(vax_t1 == 0, !is.na(vax_t2)) |>
3   pull(vax_t2, name = person)
```

```
A B I J M N
0 0 1 1 0 0
```

```
1 mod_vax_t2 <- glm(vax_t2 ~ 1, data = trial_data, family = binomial(),
2                   subset = vax_t1 == 0)
3 p_vax_t2 <- predict(mod_vax_t2, type = "response")[1]
4 p_vax_t2
```

```
1
0.3333333
```

Vaccination in T3 = censoring (strategy 0) or not (3)

(For strategies 1 and 2, already censored if not vaccinated in T1 or T2 so no one “at risk for” censoring here)

```
1 trial_data |>
2   filter(vax_t1 == 0, vax_t2 == 0, !is.na(vax_t3)) |>
3   pull(vax_t3, name = person)
```

A M
0 1

```
1 mod_vax_t3 <- glm(vax_t3 ~ 1, data = trial_data, family = binomial(), subse
2 p_vax_t3 <- predict(mod_vax_t3, type = "response")[1]
3 p_vax_t3
```

1
0.5

Step 3b: Calculate Weights

Weight = $1 / (\text{cumulative probability of not being censored})$

```
1 weighted_data <- censored_data %>%
2   mutate(
3     # Probability of not being censored at each time point
4     prob_not_cens_t1 = case_when(
5       is.na(censored_t1) ~ 1, # Not at risk
6       strategy == 1 ~ p_vax_t1, # Strategy 1: needs vax at T1
7       strategy %in% c(0, 2, 3) ~ 1 - p_vax_t1, # No vax at T1
8       TRUE ~ 1
9     ),
10
11    prob_not_cens_t2 = case_when(
12      is.na(censored_t2) ~ 1, # Not at risk
13      strategy == 2 ~ p_vax_t2, # Strategy 2: needs vax at T2
14      strategy %in% c(0, 3) ~ 1 - p_vax_t2, # No vax at T2
15      TRUE ~ 1
16    ),
17
18    prob_not_cens_t3 = case_when(
```

Example: Strategy 0 (Never vaccinate)

Uncensored individuals and their weights:

```
1 weighted_data %>%  
2   filter(strategy == 0, !censored) %>%  
3   select(person, livebirth, weight)
```

A tibble: 10 × 3

	person	livebirth	weight
	<chr>	<dbl>	<dbl>
1	A	1	4.00
2	B	1	2.00
3	C	0	1.33
4	D	0	1
5	H	0	1
6	K	0	1.33
7	L	0	1
8	N	1	2.00
9	O	0	1.33
10	P	0	1

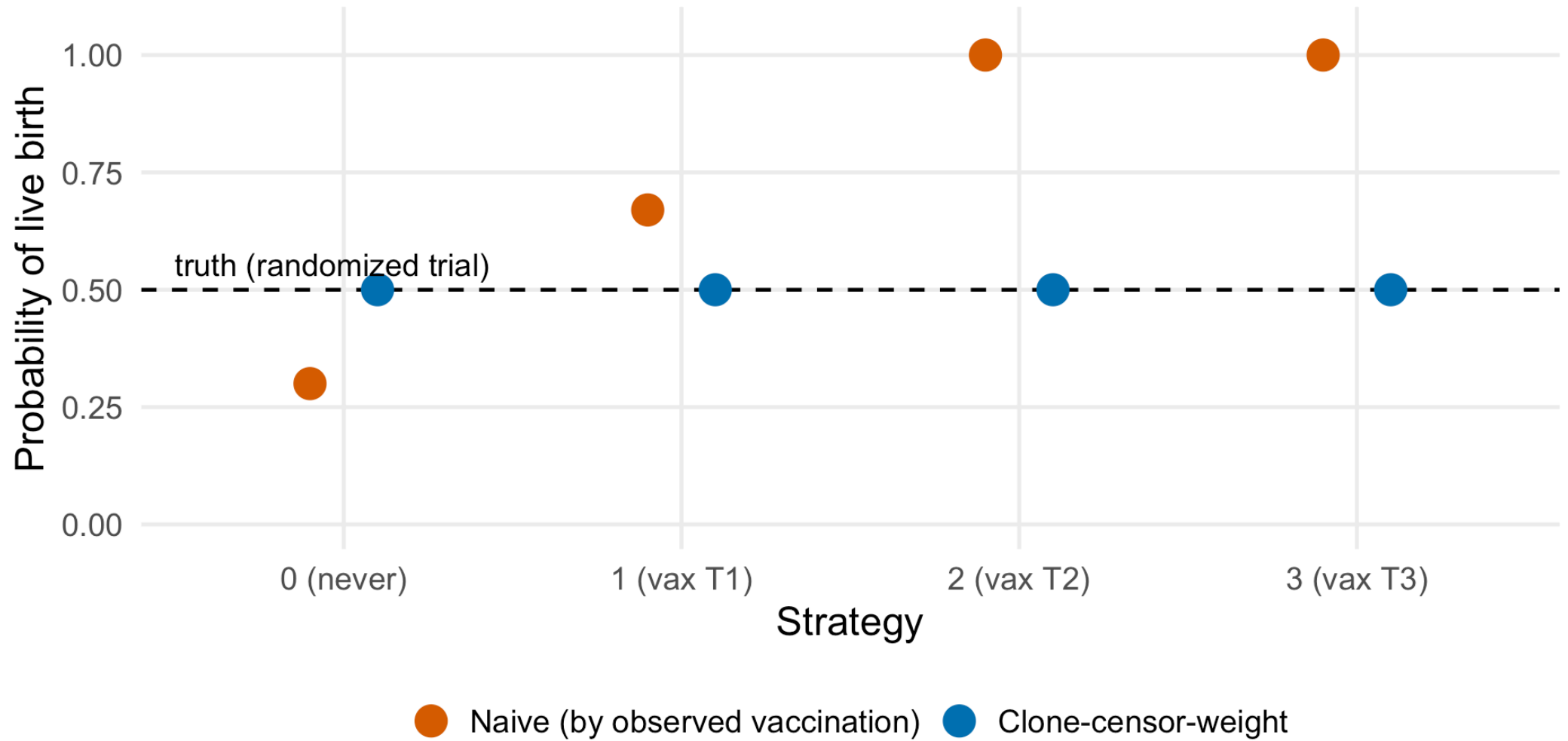
Step 3c: Calculate Weighted Outcomes

```
1 weighted_data %>%
2   filter(!censored) %>%
3   group_by(strategy) %>%
4   summarise(
5     sum_weights = sum(weight),
6     weighted_livebirths = sum(livebirth * weight),
7     risk_livebirth = weighted_livebirths / sum_weights
8   )
```

A tibble: 4 × 4

	strategy	sum_weights	weighted_livebirths	risk_livebirth
	<int>	<dbl>	<dbl>	<dbl>
1	0	16.0	8.00	0.500
2	1	16.0	8.00	0.500
3	2	16.0	8.00	0.500
4	3	16.0	8.00	0.500

Back to the truth



Why it works

The three steps:

1. **Cloning** eliminates immortal time bias by assigning strategies at time zero
2. **Censoring** ensures clones follow their assigned strategy
3. **Weighting** corrects for selection bias introduced by censoring

Key assumptions

- No unmeasured confounding (of baseline treatment and treatment continuation/discontinuation, i.e., time-varying confounding)
- Correct specification of censoring models
 - There are many different modeling assumptions we could make, e.g., one model for vaccination with a term for time, or separate models at each time point
- Positivity (some probability of continuing the treatment strategy at each time)

When to use clone-censor-weighting

- Treatment duration comparisons
- Sustained treatment strategies that evolve over time
- Variable timing of exposure
- Threshold-based or dynamic treatment rules
- Any strategy where assignment isn't identifiable at time zero
- Multiple cycles or sequential treatment decisions

Practical considerations

- Descriptive analysis of treatment patterns
- Check positivity (can strategies actually be followed?)
- How will you define confounders (both baseline and time-varying)
- Start with simple examples to develop code
- Check weight distributions

Extensions and advanced topics

- Grace periods for treatment initiation
- Nested sequential trials
- Joint strategies (treatment + monitoring)

Helpful/interesting papers

Hernán et al. (2008) Cain et al. (2010) Young et al. (2011)
Hernán et al. (2016) Hernán and Robins (2016) Labrecque
and Swanson (2017) Hernán (2018) Caniglia et al. (2019)
Dickerman et al. (2019) Chiu et al. (2020) Maringe et al.
(2020) Ben-Michael et al. (2021) Wanis et al. (2024) Gaber
et al. (2024) Cashin et al. (2025) Fu et al. (2026) Moreno-
Betancur et al. (2026) Hernán et al. (2026)

Discussion/Questions

1. What pregnancy research questions are you working on?
2. How might you apply target trial thinking?
3. What challenges do you anticipate?
4. What tools or resources would be most helpful?

Thank you!

email: l.smith@northeastern.edu

References

- Ben-Michael, Eli, Avi Feller, and Elizabeth A. Stuart. 2021. “A Trial Emulation Approach for Policy Evaluations with Group-Level Longitudinal Data.” *Epidemiology* 32 (4): 533–40. <https://doi.org/10.1097/ede.0000000000001369>.
- Cain, Lauren E., James M. Robins, Emilie Lanoy, Roger Logan, Dominique Costagliola, and Miguel A. Hernán. 2010. “When to Start Treatment? A Systematic Approach to the Comparison of Dynamic Regimes Using Observational Data.” *The International Journal of Biostatistics* 6 (2): 1–42. <https://doi.org/10.2202/1557-4679.1212>.
- Caniglia, Ellen C., James M. Robins, Lauren E. Cain, et al. 2019. “Emulating a Trial of Joint Dynamic Strategies: An Application to Monitoring and Treatment of HIV-Positive Individuals.” *Statistics in Medicine* 38 (13): 2428–46. <https://doi.org/10.1002/sim.8120>.
- Cashin, Aidan G., Harrison J. Hansford, Miguel A. Hernán, et al. 2025. “Transparent Reporting of Observational Studies Emulating a Target Trial: The TARGET Statement.” *BMJ* 390 (September): e087179. <https://doi.org/10.1136/bmj-2025-087179>.

- Chiu, Yu-Han, Mats J. Stensrud, Issa J. Dahabreh, et al. 2020. "The Effect of Prenatal Treatments on Offspring Events in the Presence of Competing Events: An Application to a Randomized Trial of Fertility Therapies." *Epidemiology* 31 (5): 636. <https://doi.org/10.1097/EDE.0000000000001222>.
- Dickerman, Barbra A., Xabier García-Albéniz, Roger Logan, Spiros Denaxas, and Miguel A. Hernán. 2019. "Avoidable Flaws in Observational Analyses: An Application to Statins and Cancer." *Nature Medicine* 25 (10): 1601–6. <https://doi.org/10.1038/s41591-019-0597-x>.
- Fu, Edouard L., Michael O. Harhay, Sebastian Schneeweiss, Rishi J. Desai, and Miguel A. Hernán. 2026. "Starting Right: Aligning Eligibility and Treatment Assignment at Time Zero When Emulating a Target Trial." *BMJ* 392 (January): e084909. <https://doi.org/10.1136/bmj-2025-084909>.
- Gaber, Charles E., Kent A. Hanson, Sodam Kim, Jennifer L. Lund, Todd A. Lee, and Eleanor J. Murray. 2024. "The Clone-Censor-Weight Method in Pharmacoepidemiologic Research: Foundations and Methodological Implementation." *Current Epidemiology Reports*, ahead of print, February. <https://doi.org/10.1007/s40471-024-00346-2>.
- Hernán, Miguel A. 2018. "How to Estimate the Effect of Treatment Duration on Survival Outcomes Using Observational Data." *BMJ* 360 (February): k182. <https://doi.org/10.1136/bmj.k182>.

- Hernán, Miguel A., Alvaro Alonso, Roger Logan, et al. 2008. "Observational Studies Analyzed Like Randomized Experiments: An Application to Postmenopausal Hormone Therapy and Coronary Heart Disease." *Epidemiology* 19 (6): 766–79. <https://doi.org/10.1097/ede.0b013e3181875e61>.
- Hernán, Miguel A., Barbra A. Dickerman, Sonja A. Swanson, and Issa J. Dahabreh. 2026. "Where Do Target Trials Come from? Specifying the Protocol of a Target Trial When Repurposing Data for Causal Inference." *Epidemiology* 37 (3): 282. <https://doi.org/10.1097/EDE.0000000000001951>.
- Hernán, Miguel A., and James M. Robins. 2016. "Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available." *American Journal of Epidemiology* 183 (8): 758–64. <https://doi.org/10.1093/aje/kwv254>.
- Hernán, Miguel A., Brian C. Sauer, Sonia Hernández-Díaz, Robert Platt, and Ian Shrier. 2016. "Specifying a Target Trial Prevents Immortal Time Bias and Other Self-Inflicted Injuries in Observational Analyses." *Journal of Clinical Epidemiology* 79: 70–75. <https://doi.org/10.1016/j.jclinepi.2016.04.014>.
- Labrecque, Jeremy A., and Sonja A. Swanson. 2017. "Target Trial Emulation: Teaching Epidemiology and Beyond." *European Journal of Epidemiology* 32 (6): 473–75. <https://doi.org/10.1007/s10654-017-0293-4>.
- Maringe, Camille, Sara Benitez Majano, Aimilia Exarchakou, et al. 2020. "Reflection on Modern Methods: Trial Emulation in the Presence of Immortal-Time Bias. Assessing the Benefit of Major Surgery for Elderly Lung Cancer Patients Using

Observational Data.” *International Journal of Epidemiology* 49 (5): 1719–29.

<https://doi.org/10.1093/ije/dyaa057>.

Moreno-Betancur, Margarita, Rushani Wijesuriya, and John B. Carlin. 2026. “The Ideal Trial: Defining Causal Estimands That Balance Relevance and Feasibility in Target Trial Emulations and Actual Randomized Trials.” *Epidemiology* 37 (2): 153.

<https://doi.org/10.1097/EDE.0000000000001933>.

Smith, Louisa H., Camille Y. Dollinger, Tyler J. VanderWeele, Diego F. Wyszynski, and Sonia Hernández-Díaz. 2022. “Timing and Severity of COVID-19 During Pregnancy and Risk of Preterm Birth in the International Registry of Coronavirus Exposure in Pregnancy.” *BMC Pregnancy and Childbirth* 22 (1): 775.

<https://doi.org/10.1186/s12884-022-05101-3>.

Wanis, Kerollos Nashat, Aaron L Sarvet, Lan Wen, et al. 2024. “Grace Periods in Comparative Effectiveness Studies of Sustained Treatments.” *Journal of the Royal Statistical Society Series A: Statistics in Society*, January, qnae002.

<https://doi.org/10.1093/jrssa/qnae002>.

Young, Jessica G., Lauren E. Cain, James M. Robins, Eilis J. O’Reilly, and Miguel A. Hernán. 2011. “Comparative Effectiveness of Dynamic Treatment Regimes: An Application of the Parametric g-Formula.” *Statistics in Biosciences* 3 (1): 119–43.

<https://doi.org/10.1007/s12561-011-9040-7>.

Zidan, Mahmoud, Nhung T. H. Trinh, Anteneh Desalegn, et al. 2025. “BNT162b2 mRNA COVID-19 Vaccine Effectiveness in Pregnancy: Emulating Trial