# Bias bounds *and* target trials *for* causal inference *in* observational epidemiology

Dissertation Defense
Louisa Smith
May 7, 2021

# Zoom defense protocol

and rough schedule

3:00-4:00: Louisa presents
4:00-4:45: Committee questions
4:45-4:55: Audience questions
4:55-5:00: Committee deliberates

Please keep your video and microphone off during the presentation, unless you are on the committee.

If you have problems or questions, ask them in the chat and someone will help you out.

# How do we decide if something causes something else?

exposure
treatment
intervention

increases
decreases
harms
benefits

outcome
(risk, incidence,
rate, hazard)

## causal inference

**experimental studies (e.g., randomized controlled trials)**

**observational studies**

design them like experimental studies to avoid bias

see how sensitive the results are to possible biases

**Paper 1**

**Paper 2**

**Paper 3**

# Overview

## Paper 1

Does COVID-19 during pregnancy increase the risk of preterm birth?

## Paper 2

Can we improve prostate cancer survival with more targeted treatment strategies?

## Paper 3

How much bias could change the conclusions we draw from observational studies?

# Paper 1

Does COVID-19 during pregnancy increase the risk of preterm birth?

1

## Potential problems

Why is this a complex question to answer?

2

## Target trial

How can thinking about experimental studies help address those problems?

3

## Analytic methods

Apply target trial thinking to a large, observational cohort study.

4

## Conclusions

What did we learn about COVID-19 and preterm birth?

# COVID-19 during pregnancy

- Infections known to be harmful during pregnancy
- Pregnant people may be at higher risk of infection or, more likely, bad outcomes from infection
  - Immune state
- Preterm birth (delivery before 37 weeks' gestation) is an outcome of concern because it's a leading cause of neonatal death and has potential long-term outcomes
  - Viruses: flu, some data from SARS and MERS outbreaks

**TABLE 4** Mode of delivery and preterm birth in pregnant women in the included studies.

| Authors | Gestational age, w | Mode of delivery | | | Preterm birth |
|---|---|---|---|---|---|
| | | Cesarean for maternal COVID-19 infection | Cesarean for obstetric indication | Vaginal delivery | |
| Chen et al.[4] | 36–39 | 9[a]/9 | 7[a]/9 | – | 4/9 |
| Chen et al.[20] | 38–40 | – | 2/5 | 3/5 | 0/5 |
| Dong et al.[11] | 37 | 1/1 | – | – | 0/1 |
| Fan et al.[12] | 36, 37 | 2/2 | – | – | 1/2 |
| Lee et al.[18] | 37 | – | 1/1 | | 0/1 |
| Li et al.[13] | 35 | – | 1[b]/1 | – | 1/1 |
| Liu et al.[14] | Nr | 5[c]/10 | 5[d]/10 | – | 6/10 |
| Liu et al.[15] | Nr | 9/11 | 1/11 | 1/11 | Nr |
| Wang et al.[16] | 40 | 1/1 | – | – | 0/1 |
| Wang et al.[6] | 30 | 1/1 | – | – | 1/1 |
| Yu et al.[17] | 37–41 | 7/7 | – | – | 0/7 |
| Zeng et al.[19] | Nr | 6/6[g] | – | – | Nr |
| Zhu et al.[5] | 31–39 | 1/9 | 6[e]/9 | 2[f]/9 | 6 (2 twins)/10 |

Abbreviation: Nr, not reported.

# Some initial case series

Parazzini F, Bortolus R, Mauri PA, Favilli A, Gerli S, Ferrazzi E. Delivery in pregnant women infected with SARS-CoV-2: A fast review. *International Journal of Gynecology & Obstetrics*. 2020;150(1):41-46.

# Estimates of preterm risk

- If someone gets COVID-19 at week 39 of pregnancy and delivers soon after, that is not a preterm birth – but it doesn't mean COVID-19 doesn't cause preterm birth
- % preterm LOW if COVID-19 preferentially leads to hospitalization later in pregnancy
  - Preferentially counting people who were already past the preterm threshold at infection

# Estimates of preterm risk

- If someone gets COVID-19 at week 19 of pregnancy and is soon released from the hospital (with ongoing pregnancy), we don't know yet whether they will have a preterm delivery
- % preterm HIGH if we ignore people who haven't yet delivered
  - Preferentially counting short pregnancies that finished soon enough for us to assess whether they were preterm or not

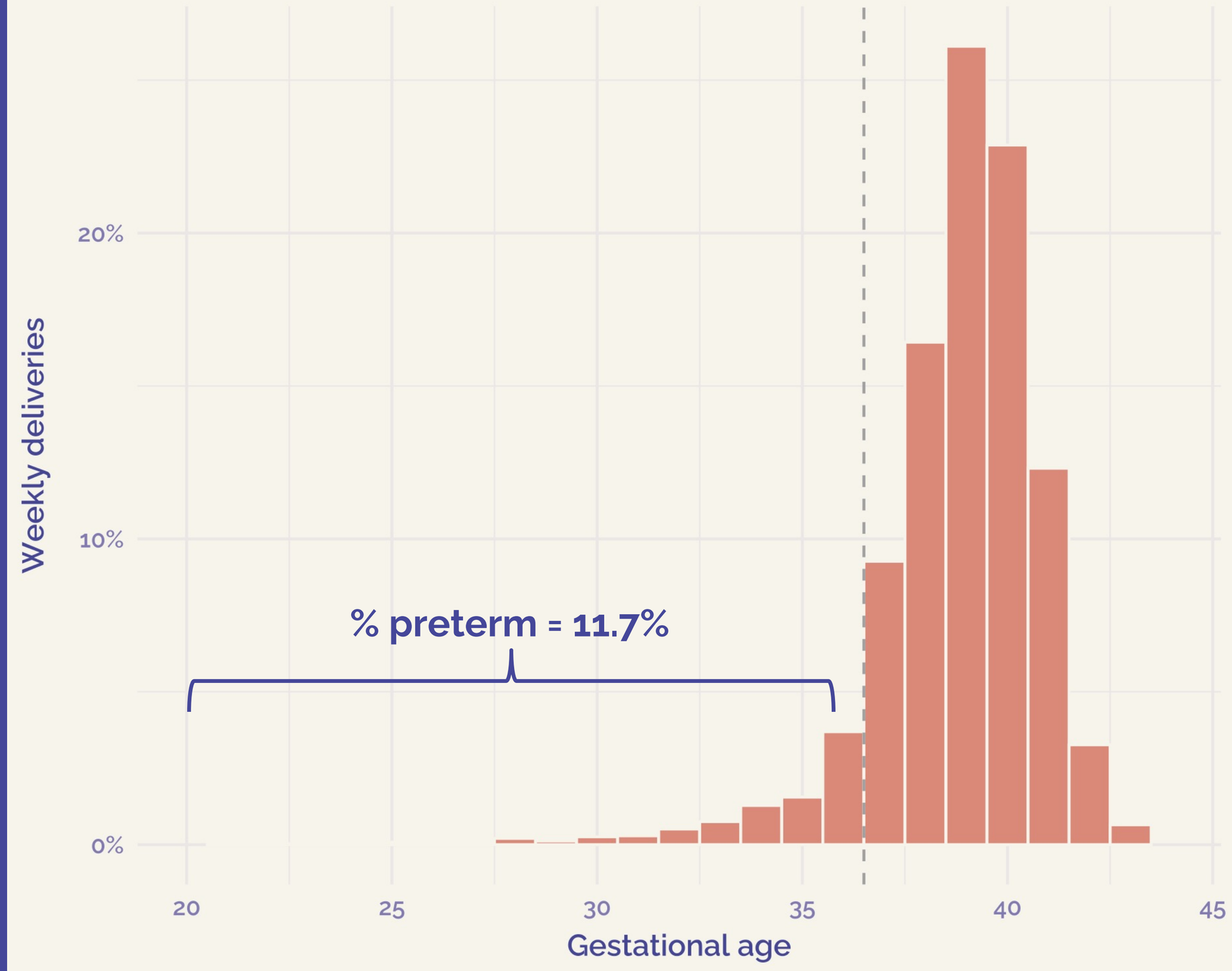
# Comparative measures

- Maybe the estimates of *absolute* risk have problems, but what about measures of *relative* risk?
- What if we count the preterm deliveries among people who had COVID-19 in pregnancy and compare to those who never did?

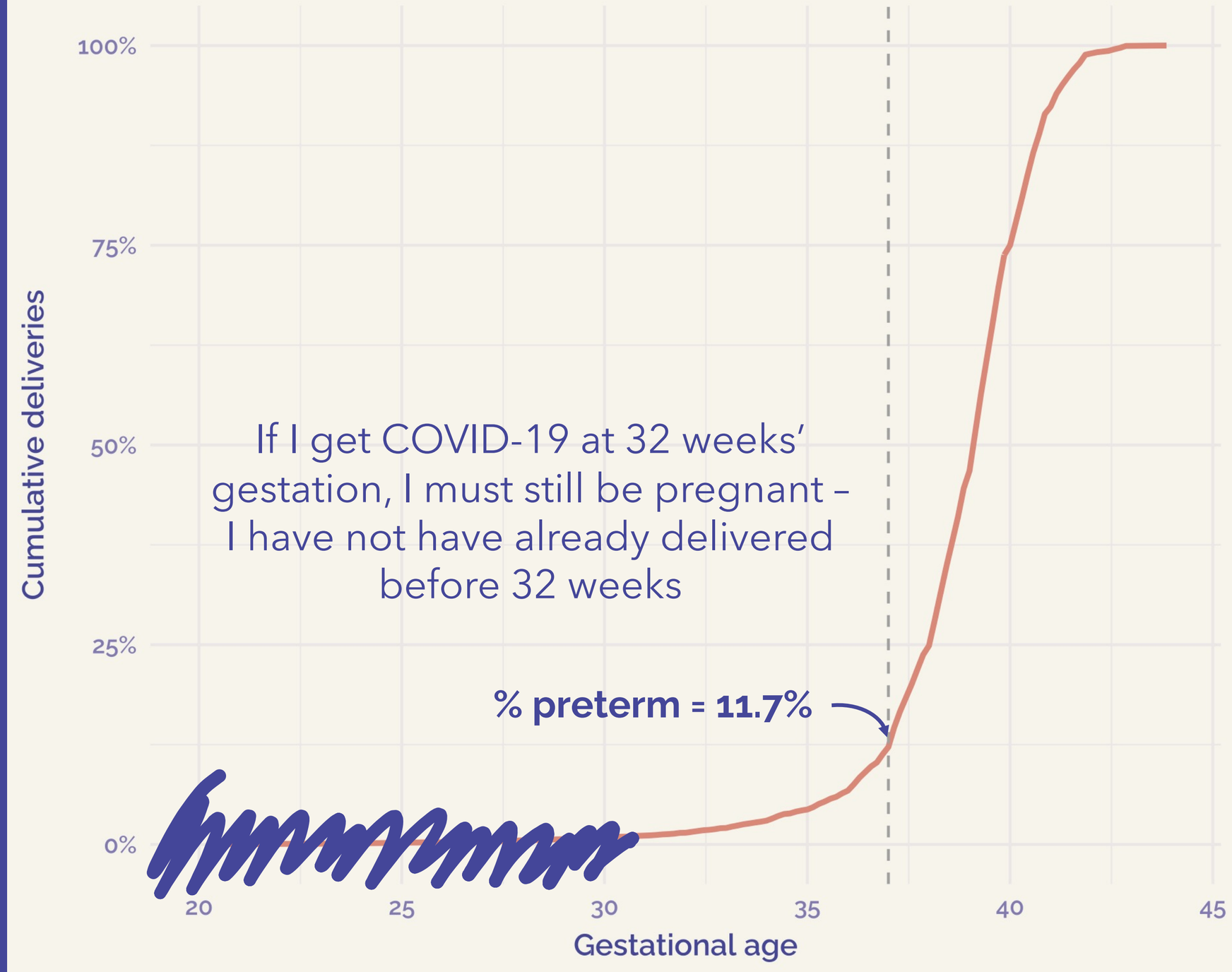
# Immortal time bias

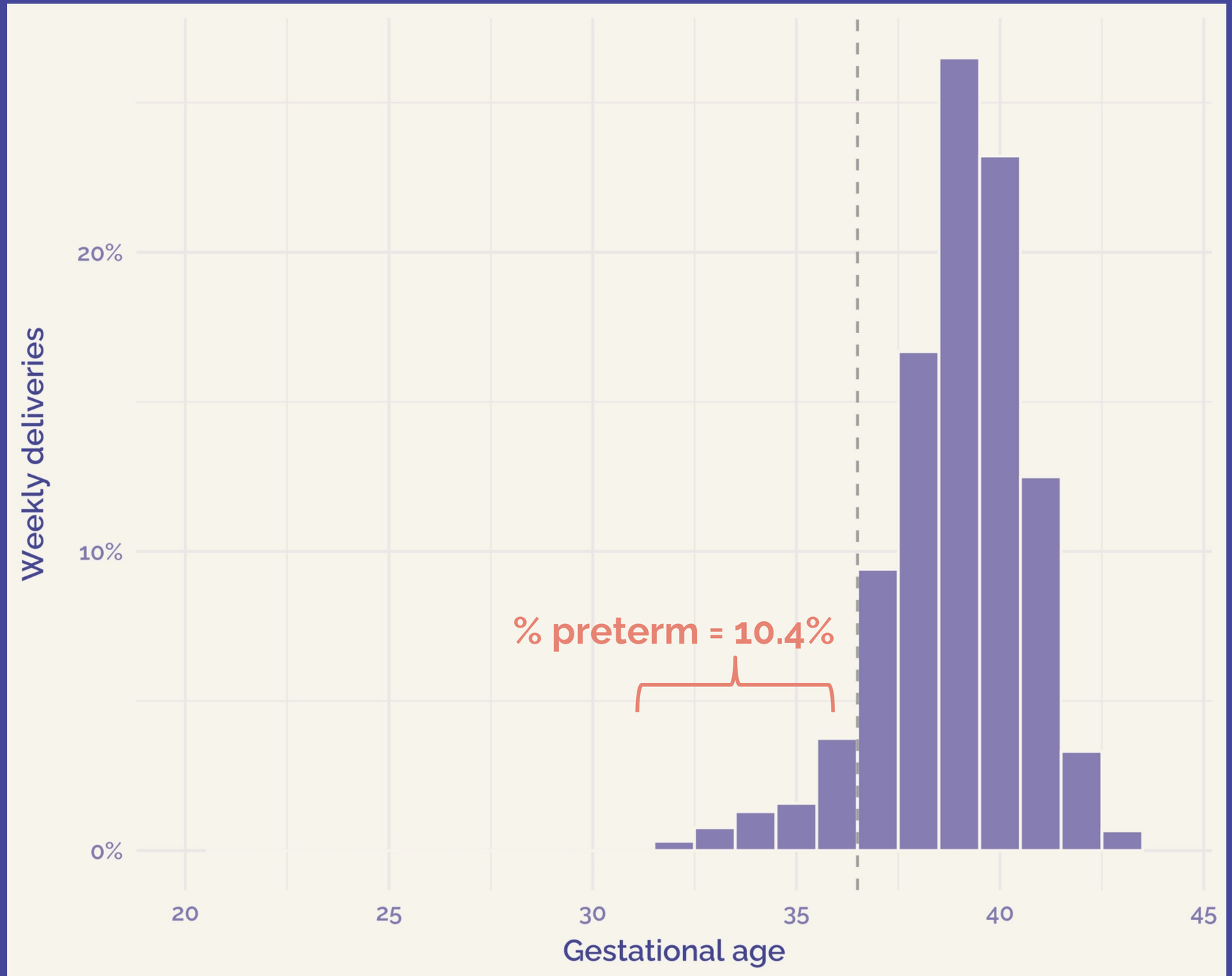
- If we just look at deliveries overall, we may underestimate the effect of COVID-19 on preterm birth
- Shorter pregnancies are less likely to have been affected by COVID-19… just because they were shorter!
- Exposure is (in part) *defined* by the requirement that a pregnancy last long enough to get COVID-19
  - This isn't a requirement for the unexposed comparison group
  - We need both the exposed and unexposed groups to start from the same time zero

# Weekly deliveries

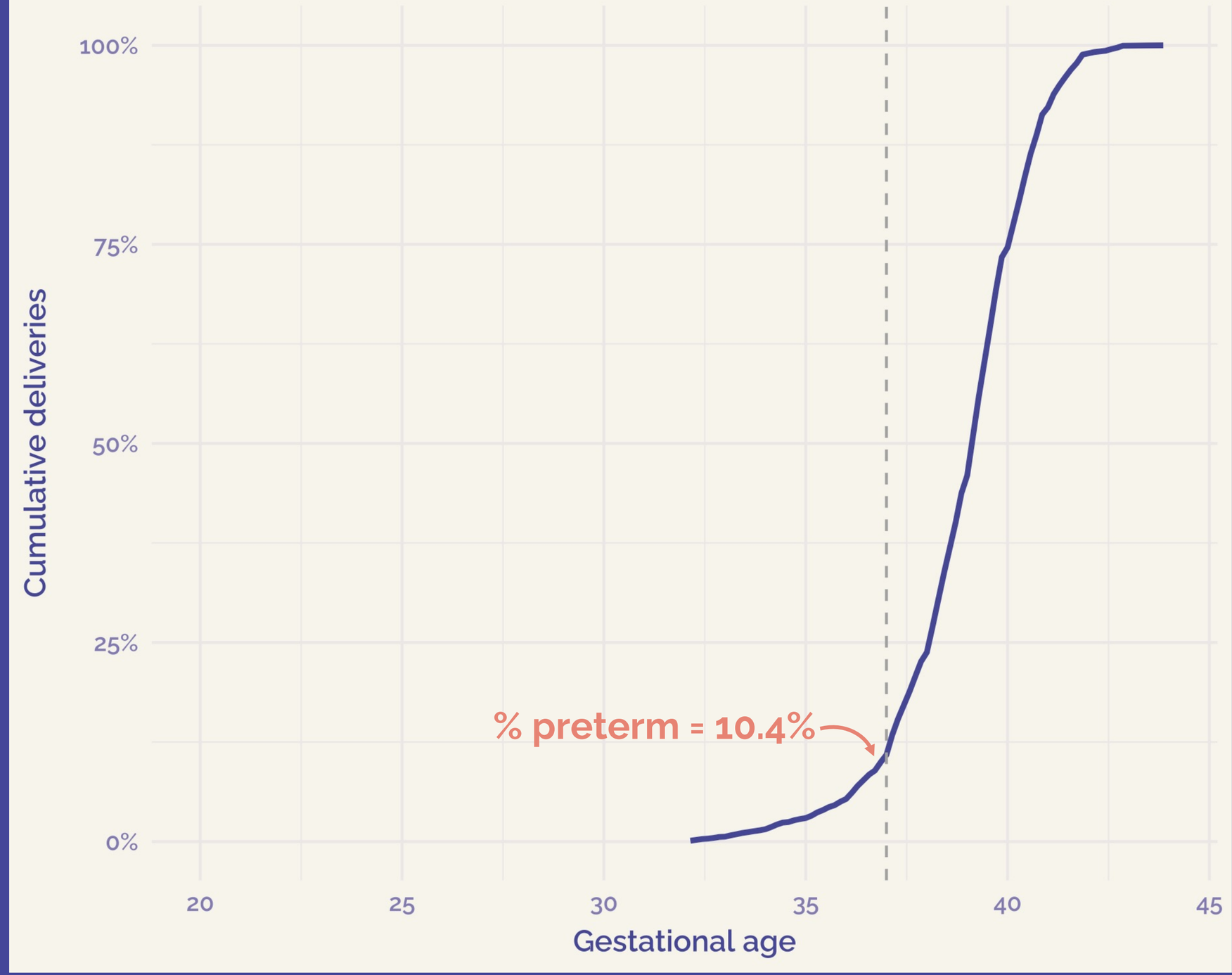


# Cumulative deliveries

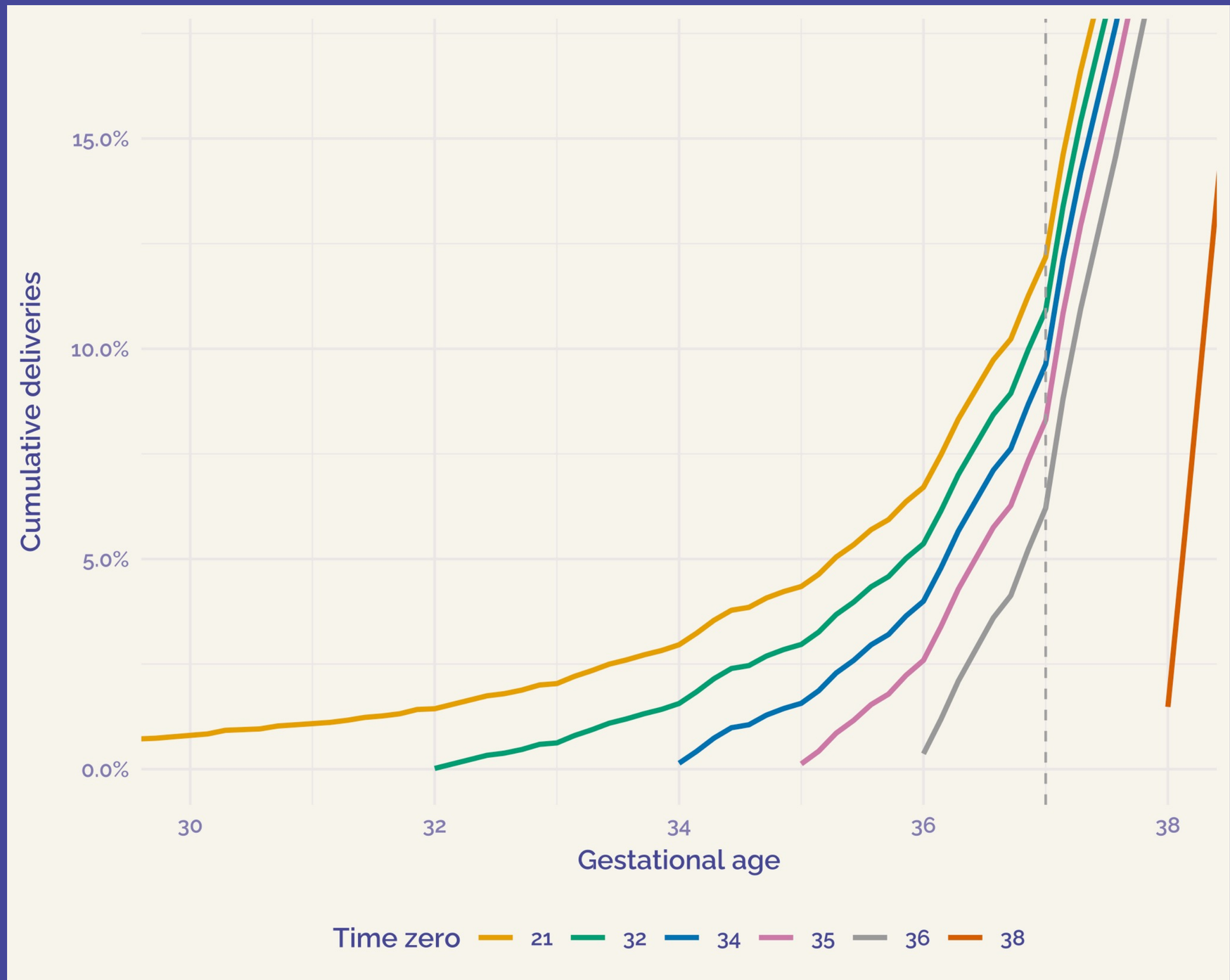





# Weekly deliveries of pregnancies ongoing at 32 weeks




# Cumulative deliveries of pregnancies ongoing at 32 weeks




# Cumulative deliveries starting at different “time zeros”

# How can we avoid having to think through all this?

- Think through how you would design a randomized controlled trial to answer you question instead

- A *target trial*

# Example: How do we know if a vaccine works?

- Who is eligible?
  - Excluded groups due to worries about effectiveness (immunosuppressed), protected groups (pregnancy), etc.
- What are the treatment strategies?
  - Get a vaccine shot, return 3-4 weeks later for another vaccine shot
  - Get a placebo shot, return 3-4 weeks later for another placebo shot
- How is treatment decided?
  - Flip of a coin (50%?)
- How is the outcome measured?
  - Symptomatic, test positive

# Apply this thinking to our question

- Who is eligible?
  - Pregnant, at e.g., 33 weeks' gestation, never had COVID-19
  - Create different eligibility group for every week of gestation (time zero)
- What are the treatment strategies?
  - Get COVID-19 that week (even assign severity!)
  - Don't get COVID-19
- How is treatment decided?
  - Randomly, but some groups may have higher chance of being assigned COVID-19, or getting severe vs. mild COVID-19 than others
- How is the outcome measured?
  - Follow everyone until delivery and measure gestational age at that point

# Now *emulate* in observational data

- At each week of gestation (time zero), choose the people who developed COVID-19 that week
  - i.e., multiple “trials” throughout gestation
- At that same week of gestation, choose the participants whose pregnancies were ongoing but who didn’t have COVID-19
  - Those people might participate in multiple “trials”
- Compare risks of preterm birth after they deliver!

# Comparison

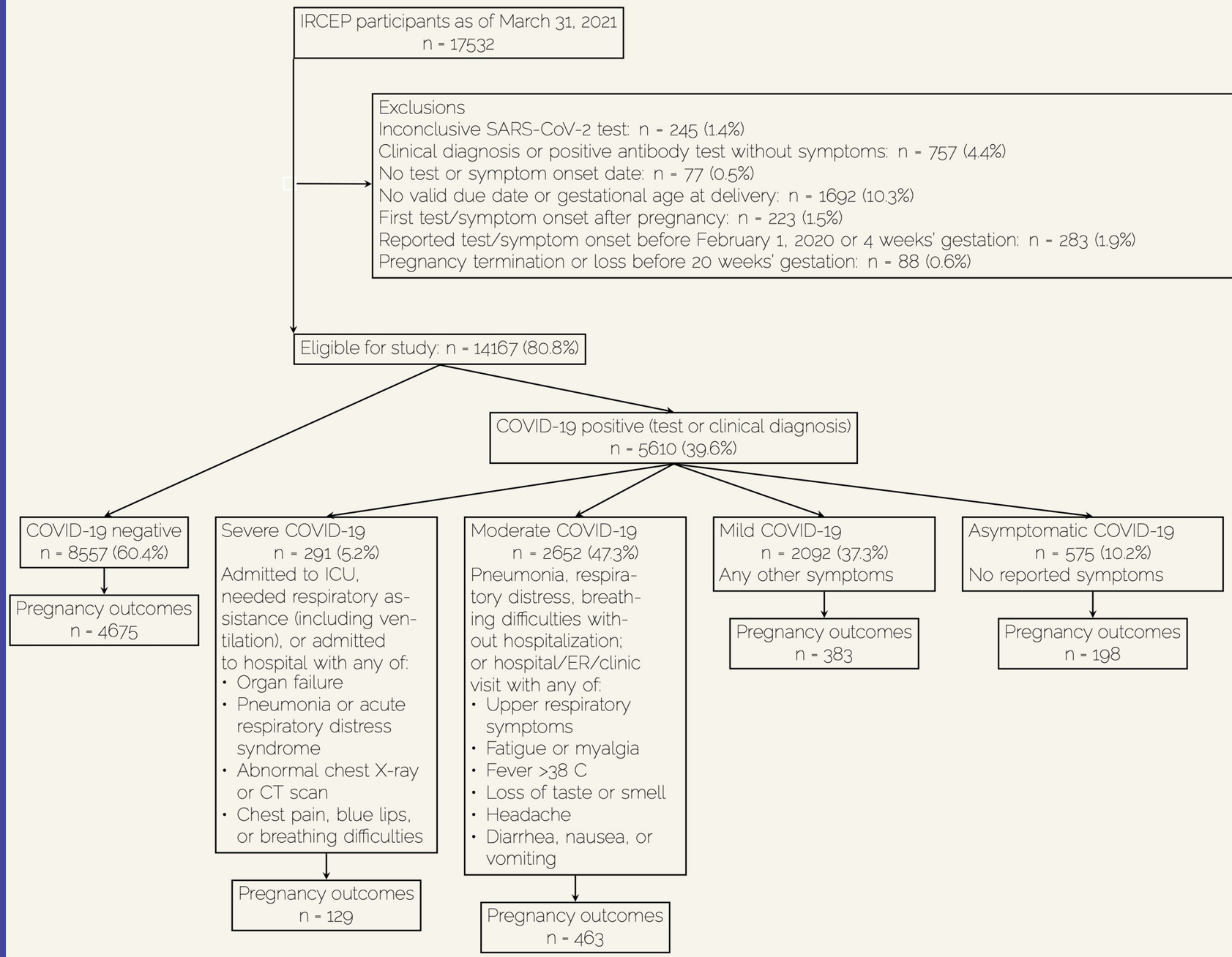
- BUT… a direct comparison only works in a trial because the exposure is randomized
  - Whether you are assigned COVID-19 or not doesn't have anything to do with your underlying risk of preterm birth
- In observational data, it may be that people who are more likely to get COVID-19 are already at higher or lower risk of preterm birth
  - This is *confounding*
  - To isolate risk due to COVID-19, we should only compare people who have the same underlying risk of preterm birth
  - Confounders depend on study population

# Study design and population

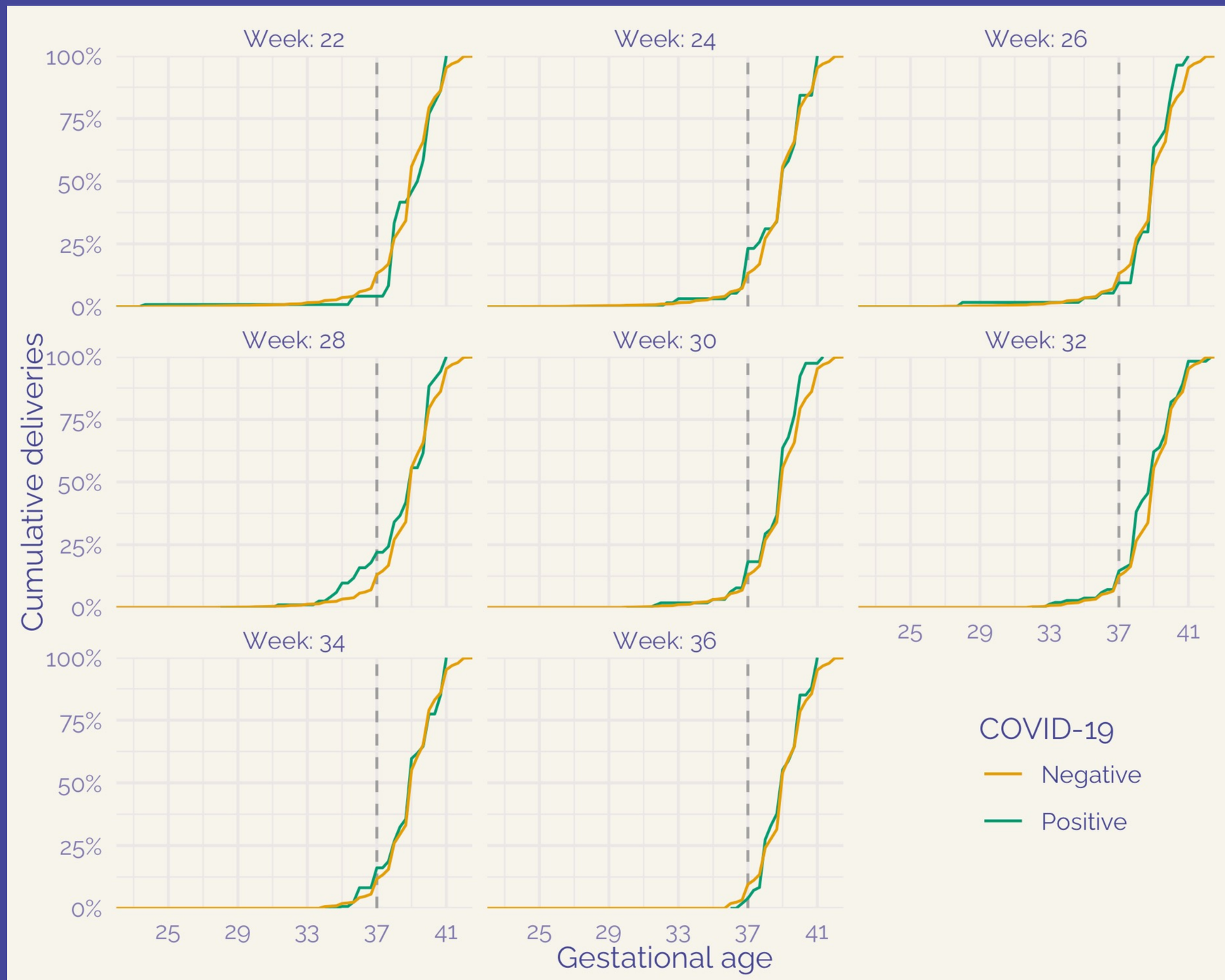
IRCEP
International Registry of Coronavirus Exposure in Pregnancy

- Enrollment during pregnancy (*prospective*) or within 6 months afterward (*retrospective*)
- Must have had a COVID-19 test or clinical diagnosis of COVID-19 during pregnancy
  - These are different eligibility criteria than our target trial, which is designed specifically for preterm delivery
- Study is advertised online in countries around the world
- Survey modules completed via internet
  - Demographics, reproductive and health history, COVID-19 symptoms/tests/treatments, pregnancy outcomes, infant outcomes at birth and 3 months

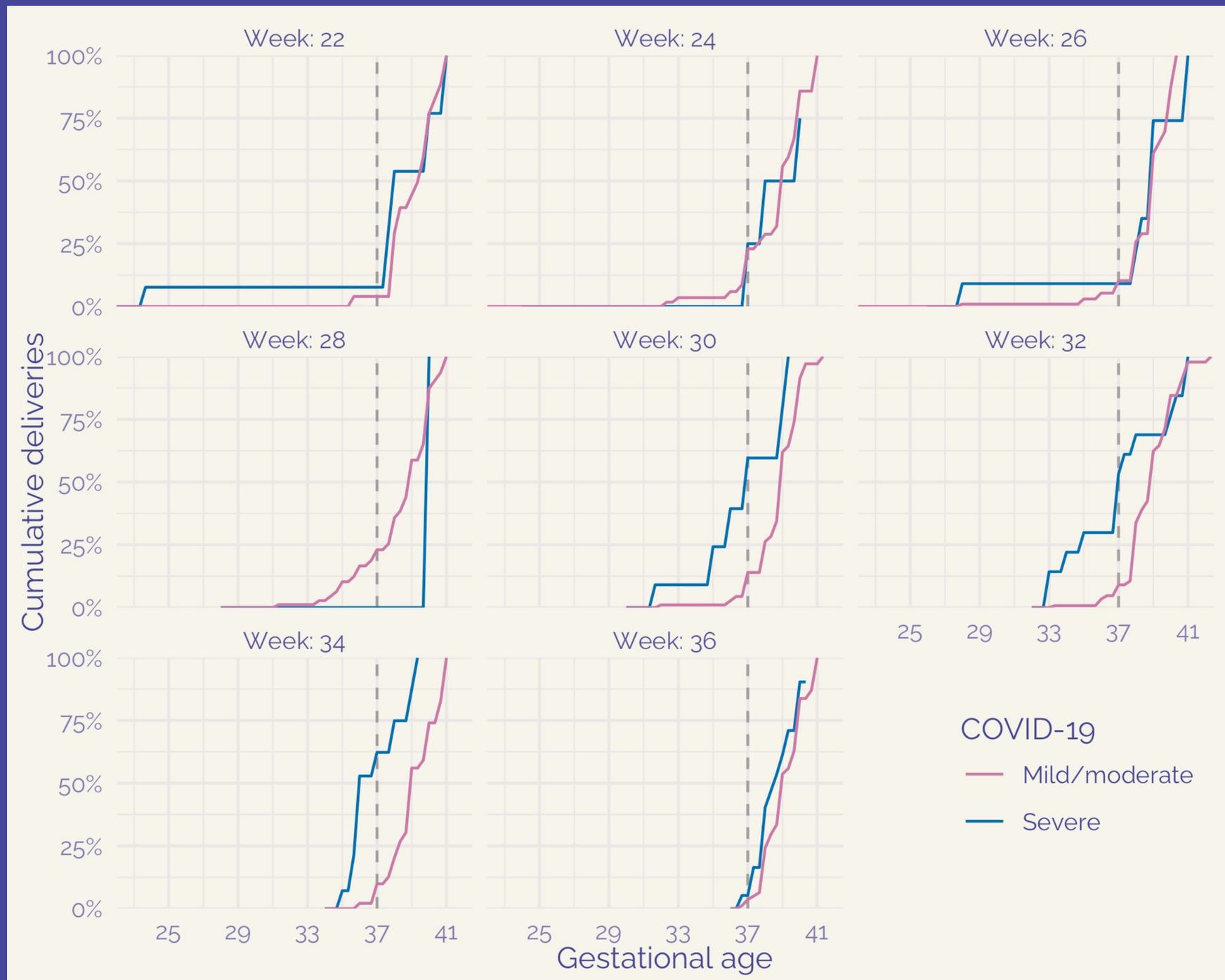
# Participants in IRCEP



# Unadjusted cumulative deliveries



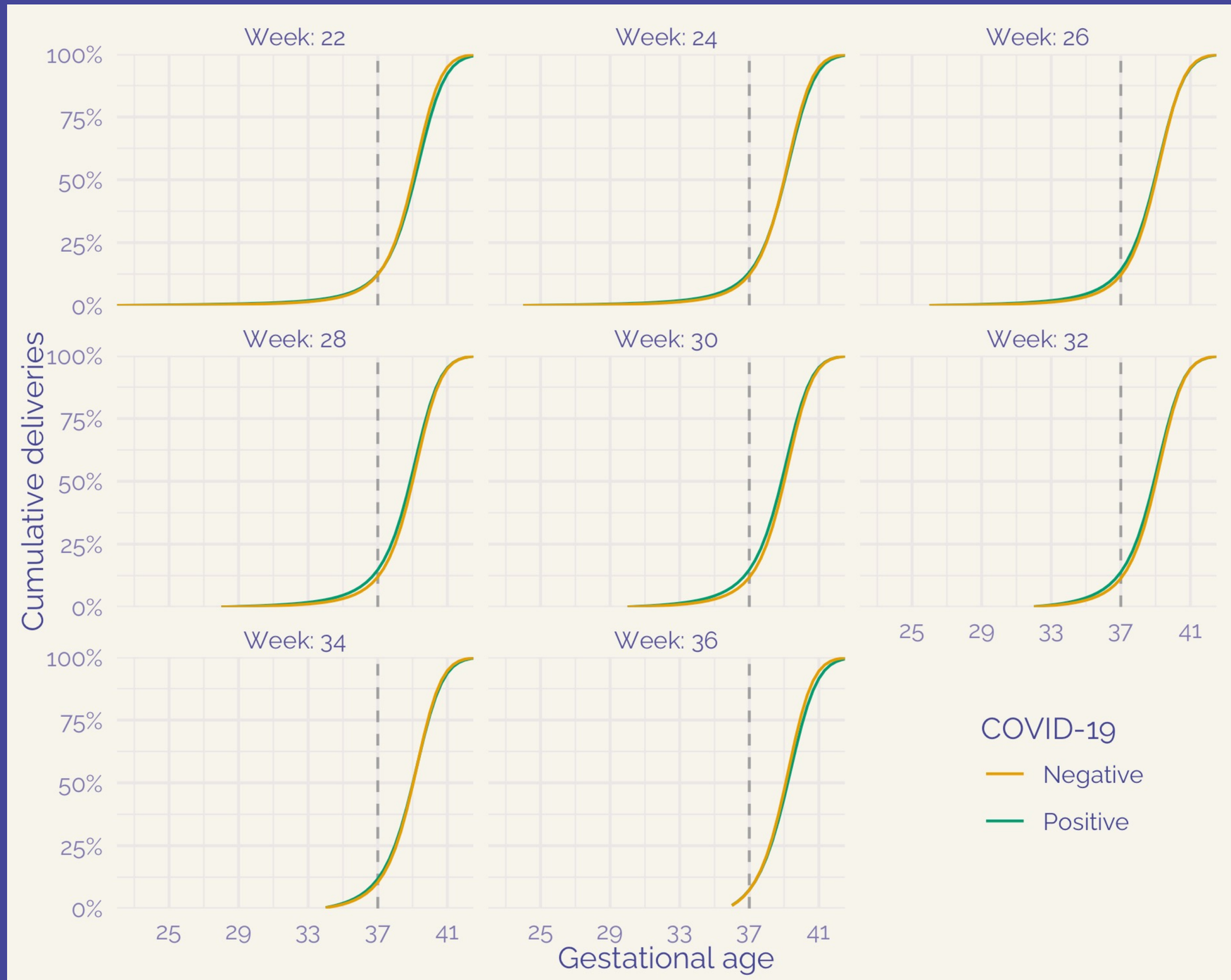
# Unadjusted cumulative deliveries



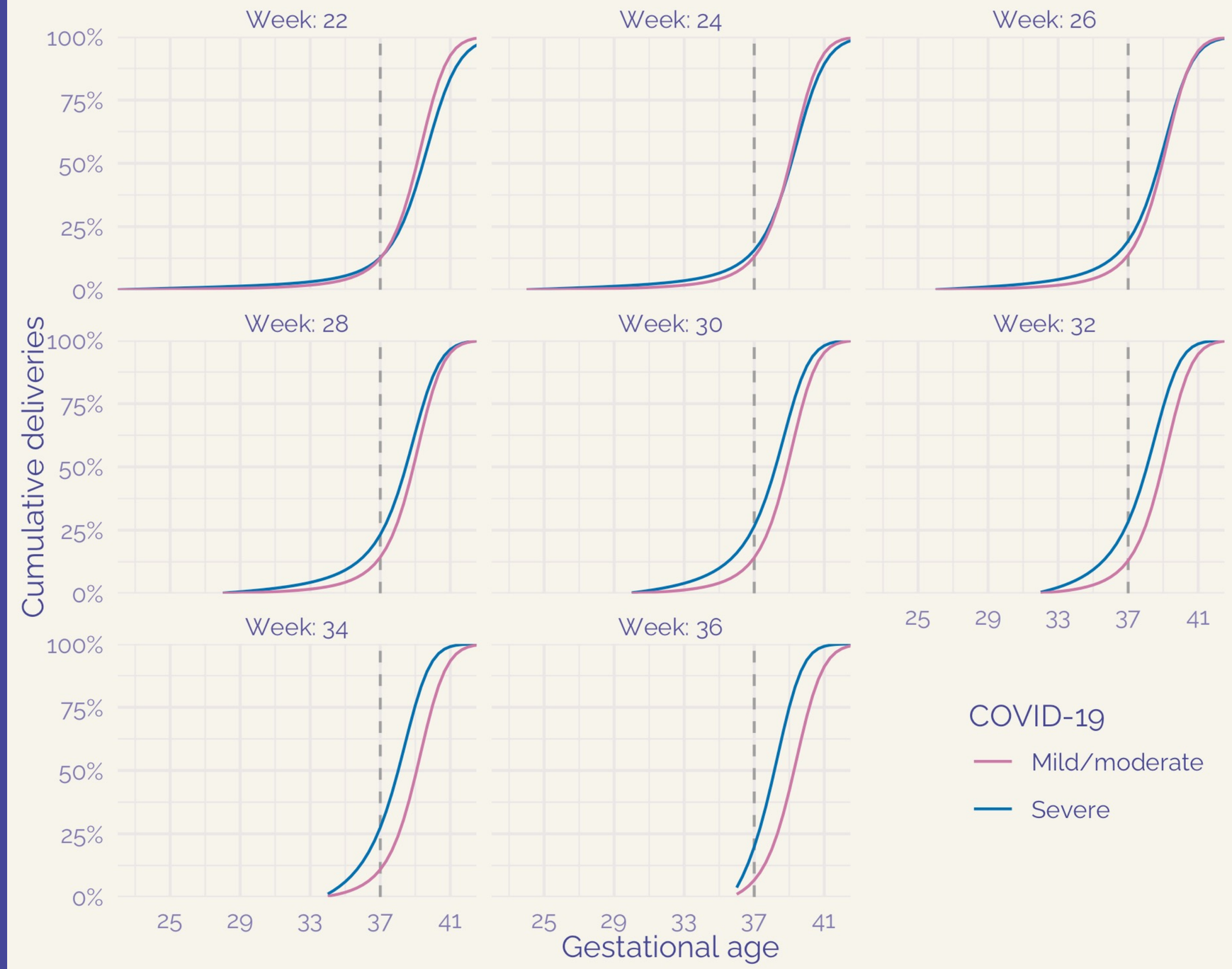
# Standardized cumulative delivery curves

- Fit a model for daily hazard of delivery:
  - "What's the probability I'll deliver at 34 weeks + 2 days if I'm still pregnant at 34 weeks + 1 day?"
  - Conditional on confounders (continent, maternal age, pre-pregnancy BMI, parity, race/ethnicity, pre-existing condition, healthcare coverage, reason for testing), infection/severity, time since infection
  - Allow delivery rate to vary over gestational age (cubic splines), and effects of infection to vary over gestational age as well (interaction terms)
- For every "time zero" week, estimate delivery hazards in weeks 20+ given observed distribution of covariates in test-negative participants still pregnant
  - Had everyone been negative but still pregnant that week, positive with mild/moderate infection that week, positive with severe infection that week

# Adjusted cumulative deliveries



# Adjusted cumulative deliveries




# Risks over pregnancy

# Delivery with severe COVID-19: patient perspective

Me practicaron una cesárea a de emergencia mientras estaba en coma inducido por covid 19, tenía 24 semas de gestación.

I had an emergency cesarean while I was in an induced coma from covid 19 at 42 weeks gestation.

Tive o covid no período de 6 meses e meio de gestação, precisou de um parto emergência, após parto coma induzido de 14 dias

I had covid at 6 and a half months' gestation, I needed an emergency delivery, after delivery I was in an induced coma for 14 days.

I unfortunately gave my baby covid 19 after delivery I went to itu and on a ventilator was very poorly I didnt meet my ba y till she was 2 weeks old I also diagnosed with PTSD

I unfortunately gave my baby covid 19 after delivery I went to itu and on a ventilator was very poorly I didnt meet my ba y till she was 2 weeks old I also diagnosed with PTSD

…mon état respiratoire s'aggravait alors l'équipe médicale a décidé de me faire accoucher (césarienne en urgence). Mon bébé … né prématurément à 34SA, a passé 3 semaines en néonatalogie. Je n'ai pu aller le voir en néonatalogie que 14 jours après résultat positif covid. Très difficile de ne pas voir son bébé

… my respiratory condition worsened so the medical team decided to deliver (emergency cesarean). My baby, born prematurely at 34 weeks, spent 3 weeks in neonatal care. I only was able to see him 14 days after the positive covid test. Very hard not to see your own baby.

# Strengths and Limitations

## Loss to follow-up

Outcomes are missing for most prospective participants – some still pregnant, others lost to follow-up.

## Gestational-age-specific absolute and relative risks

We considered effects throughout gestation, including early infections.

## Self-report

We did not have clinical measures to classify severity (e.g., oxygen levels).

## Additional analyses

We conducted multivariable regression analyses and a case-time-control analysis to support our findings, along with sensitivity analyses.

# Paper 2

Can we improve prostate cancer survival with more targeted treatment strategies?

1

## Proposing treatment strategies

What if we based prostate cancer treatment on biomarker characteristics?

2

## Refining treatment strategies

What questions do we need to answer to fully define the strategy?

3

## Methods for time-varying confounding

How can we adjust for lack of adherence?

4

## Conclusions

Can we find a strategy that minimizes all-cause mortality?

# Prostate cancer

- When treated early, usually good outcomes with either surgery or radiation
- Sometimes cancer recurs
  - It may present as overt metastases (bones, lungs, liver)
    - Treat with androgen deprivation therapy and possibly chemotherapy, radiation, etc.
    - Androgen deprivation therapy = drugs or surgery to reduce hormones
  - Or as a rise in prostate-specific antigen (PSA)
    - Treat with androgen deprivation therapy?

# Treatment considerations

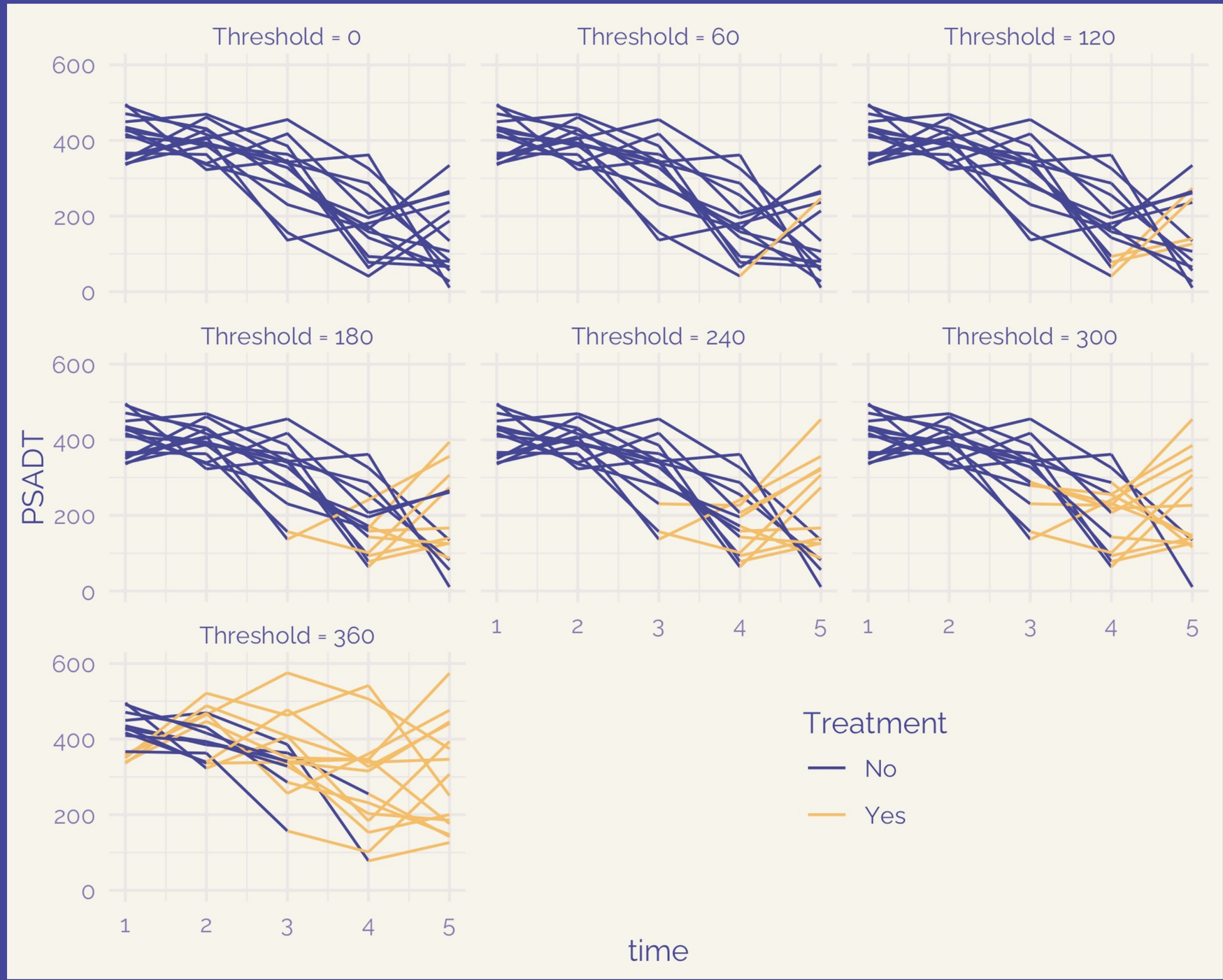
- Biochemical recurrence of prostate cancer might not lead to death from prostate cancer
  - It can be slow growing
  - It can occur toward the end of the natural lifespan
- Treatment is with androgen deprivation therapy
  - This causes side effects, negative quality of life
  - May be expensive, time-consuming, etc.
- No definite improvement in treating immediate vs. waiting ~2 years
- Goal: treat only if you need it!

# More nuanced treatment strategy

- Start treatment not only when there are clear signs of progression (may be too late in some cases), but also when PSA characteristics indicate that cancer is growing more quickly
- A common measure of growth is PSA doubling time (PSADT)
  - If PSA is rising quickly, it doubles in a short amount of time
  - We can estimate PSADT using PSA from current and most recent measurements
    - If PSA has risen 54% in the 37 days since the last appointment, it's on the road to doubling in x days
    - Lower PSADT is a bad sign
    - If PSA is flat or dropping, PSADT is undefined (consider it infinitely high)
- Why not base treatment initiation on PSADT?

# Treatment strategy based on PSADT

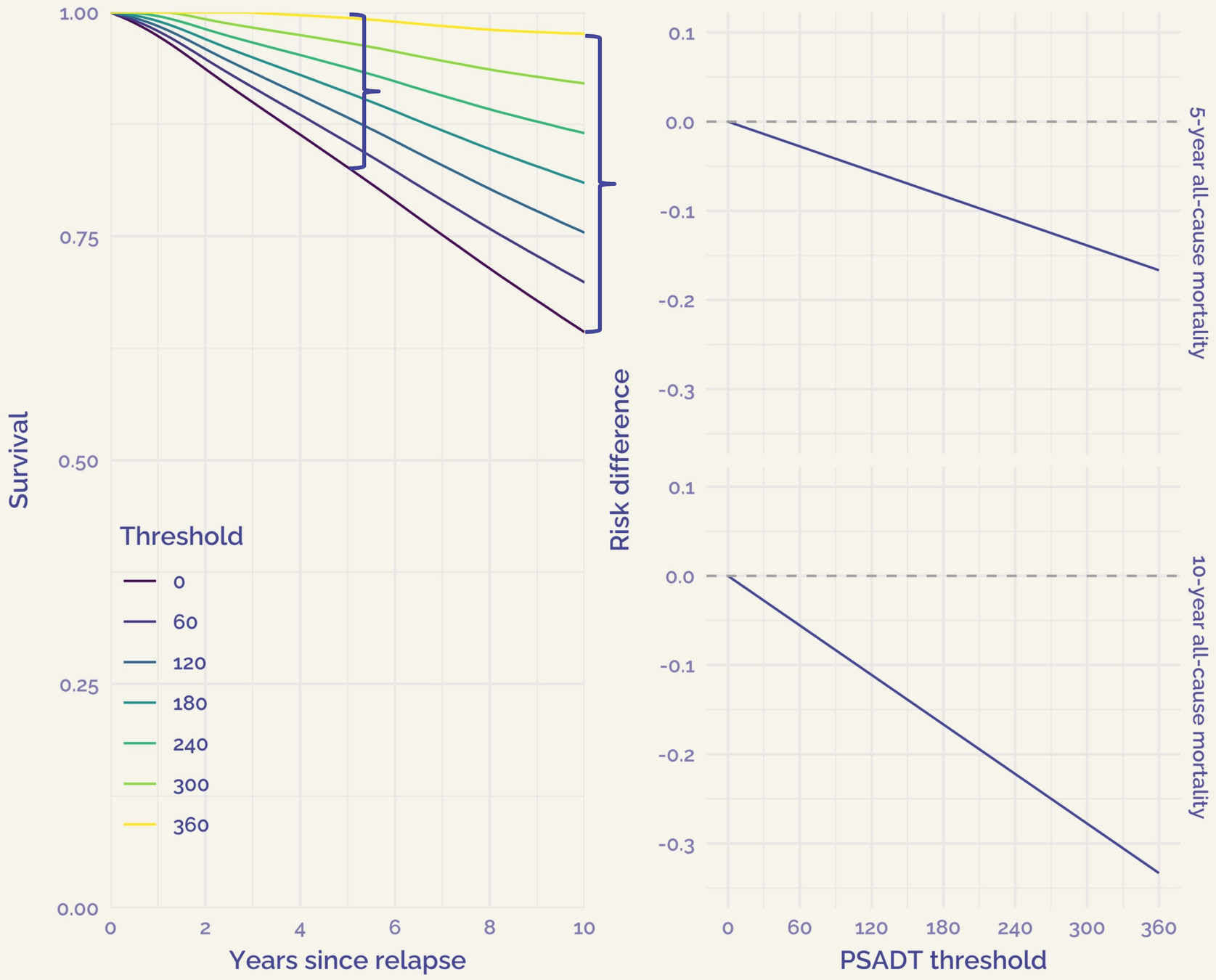
- "Start androgen deprivation therapy the first time PSADT drops below $x$ days."
  - If PSA is slow growing, its doubling time may never fall below x days – so never need treatment
  - Patients whose PSA is growing the fastest will get treated the soonest – possibly as soon as their second appointment after enrollment into the trial
- One treatment arm for each value of $x$:
  - Start the first time PSADT drops below 360 days (more people treated)
  - Start the first time PSADT drops below 300 days
  - Start the first time PSADT drops below 30 days
  - Start the first time PSADT drops below 0 days (no one is treated)




# Treatment at PSADT threshold

# Analyze the trial data: intention-to-treat

- Within each group assigned to PSADT threshold $x$ (for $x = 0, 10, \ldots, 360$)
  - Compute survival curve
  - Compute risk of all-cause mortality at 5 and 10 years
- Compare survival curves risks (risk ratios or risk differences) between treatment arms



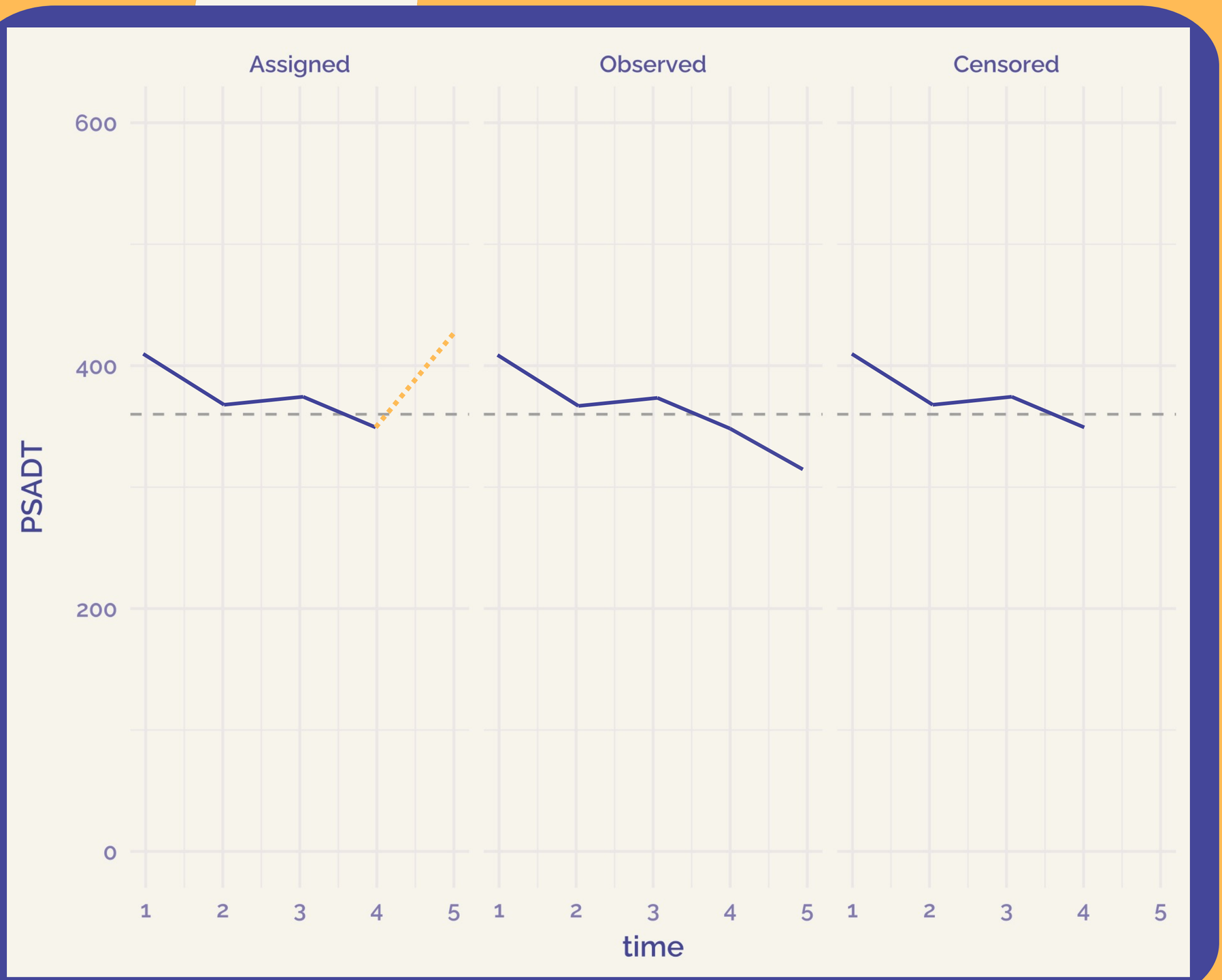




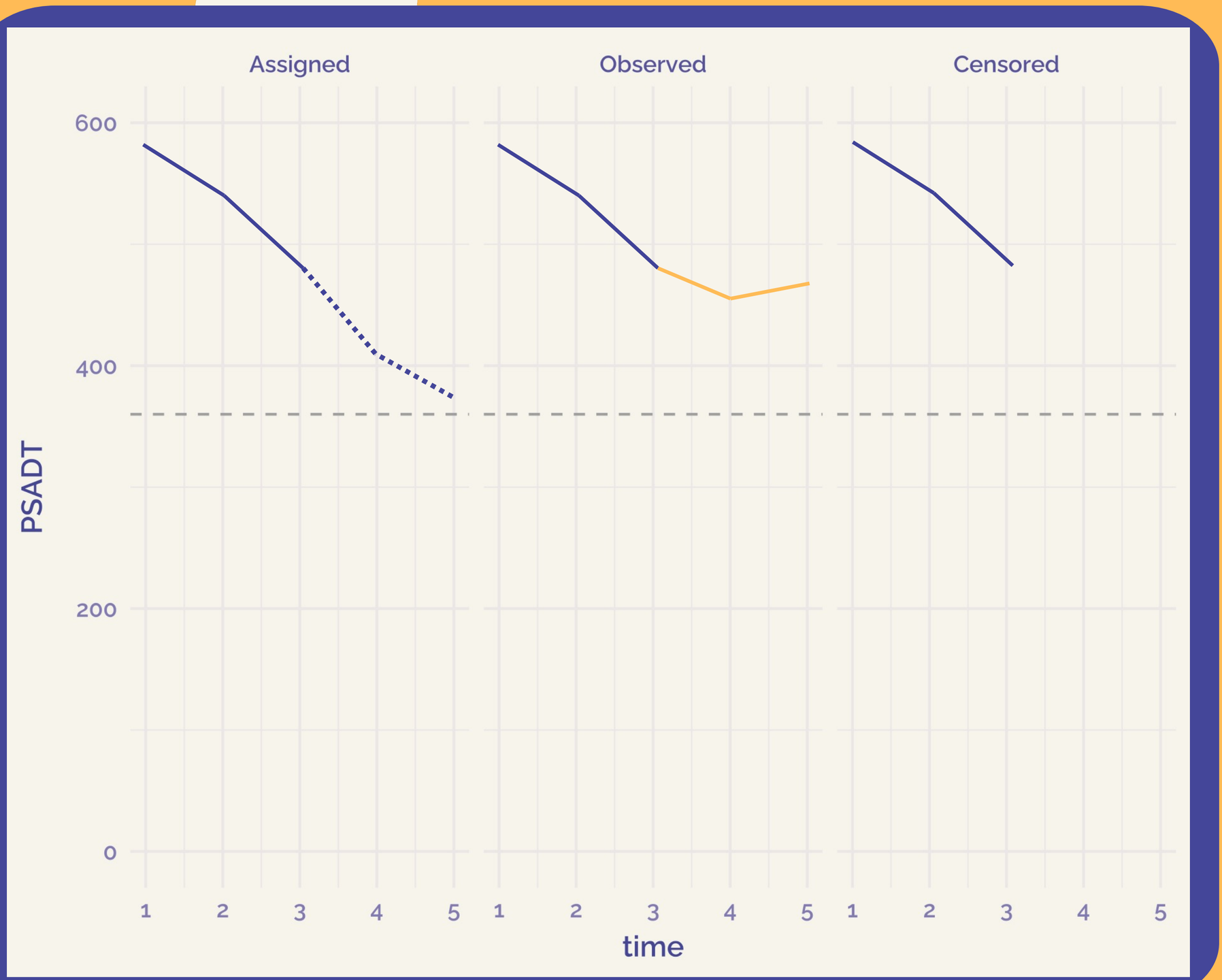
# Comparisons using trial data

# Analyze the trial data: per-protocol

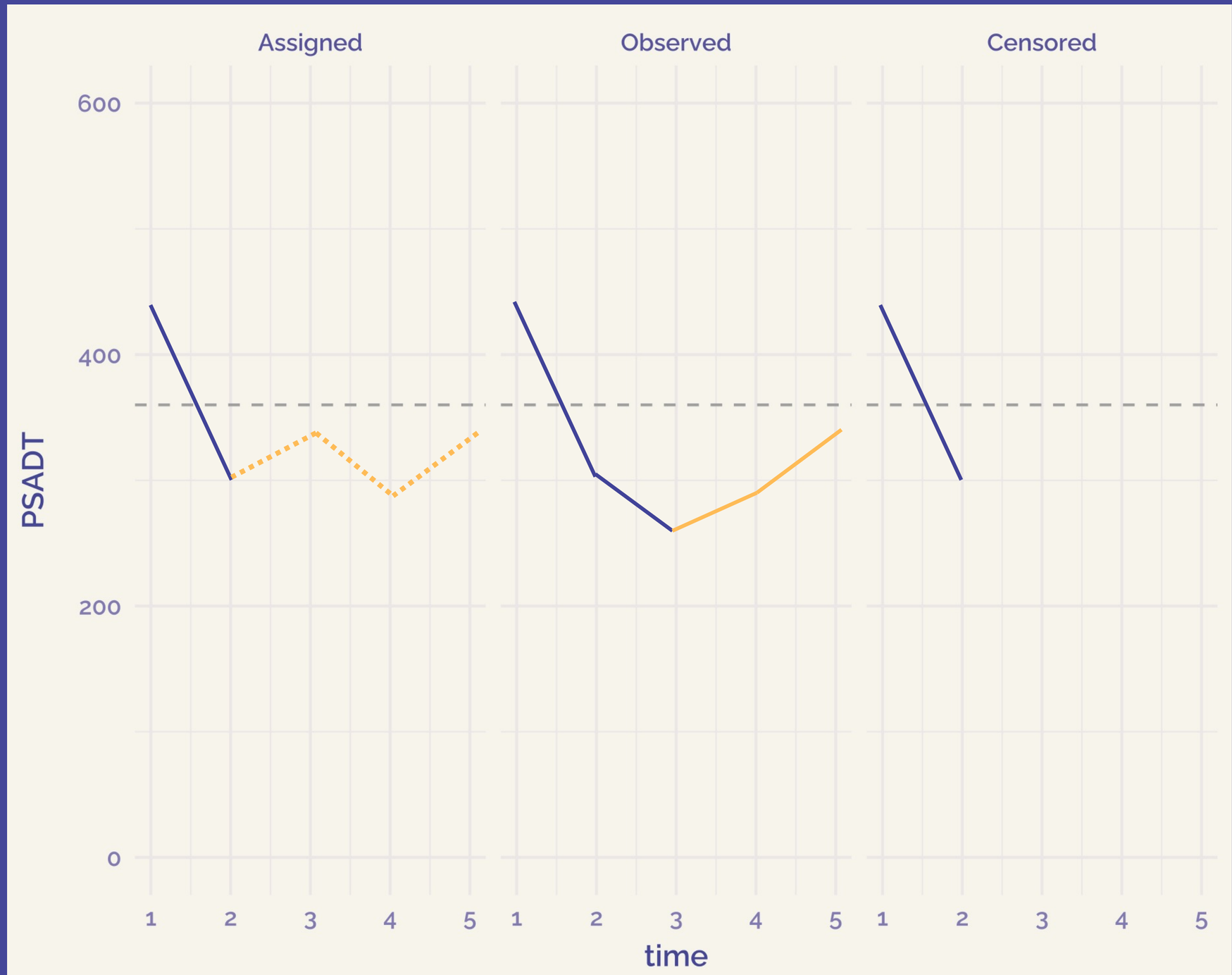
- We might want to also estimate a per-protocol effect: what if everyone actually followed their treatment strategy?
- In the COVID-19 trial we didn't really worry about this – we imagined that everyone who was randomized to either get COVID-19 or not get it actually followed through with it
  - In a real trial, depending on the treatment and the trial design, people who are randomized to get some treatment might not actually get it (or if randomized to placebo, for example, might find a way to get the drug)
- Non-adherence to the treatment strategy is more likely the longer and harder the treatment strategy is!




# Per-protocol: Threshold of 360




# Per-protocol: Threshold of 360




# Per-protocol: Threshold of 360

# Refining the treatment strategy

"Start androgen deprivation therapy the first time PSADT drops below $x$ days."

- Treatment may be clinically indicated in other settings.
- Treatment may not immediately follow the drop in PSADT.
  - If not, what is the pattern of initiation?
- Treatment can only start if PSA is monitored.
- Treatment may not continue forever after initiated.
  - If not, how long should it last?

# Refined treatment strategy

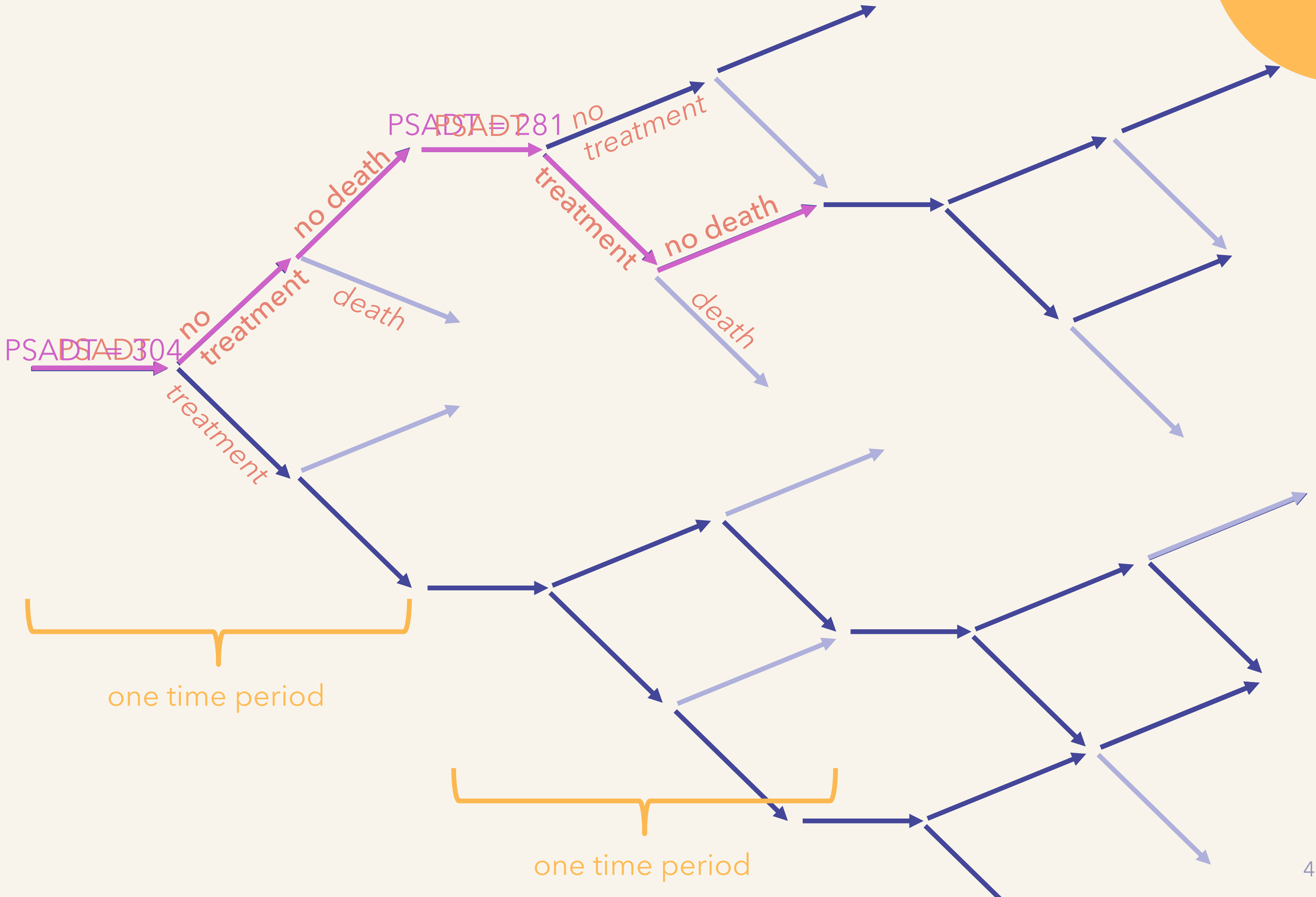
"Start androgen deprivation therapy with equal probability within the three months following the first time PSADT drops below $x$ days, or if a patient shows other signs of progression based on imaging or severe symptoms. Participants must visit their physician for tests, imaging, and or symptom assessment in addition to completing surveys at home not less than once every 2 years. Treatment duration is decided by physician and patient, but once treatment is discontinued, it is not to be restarted."

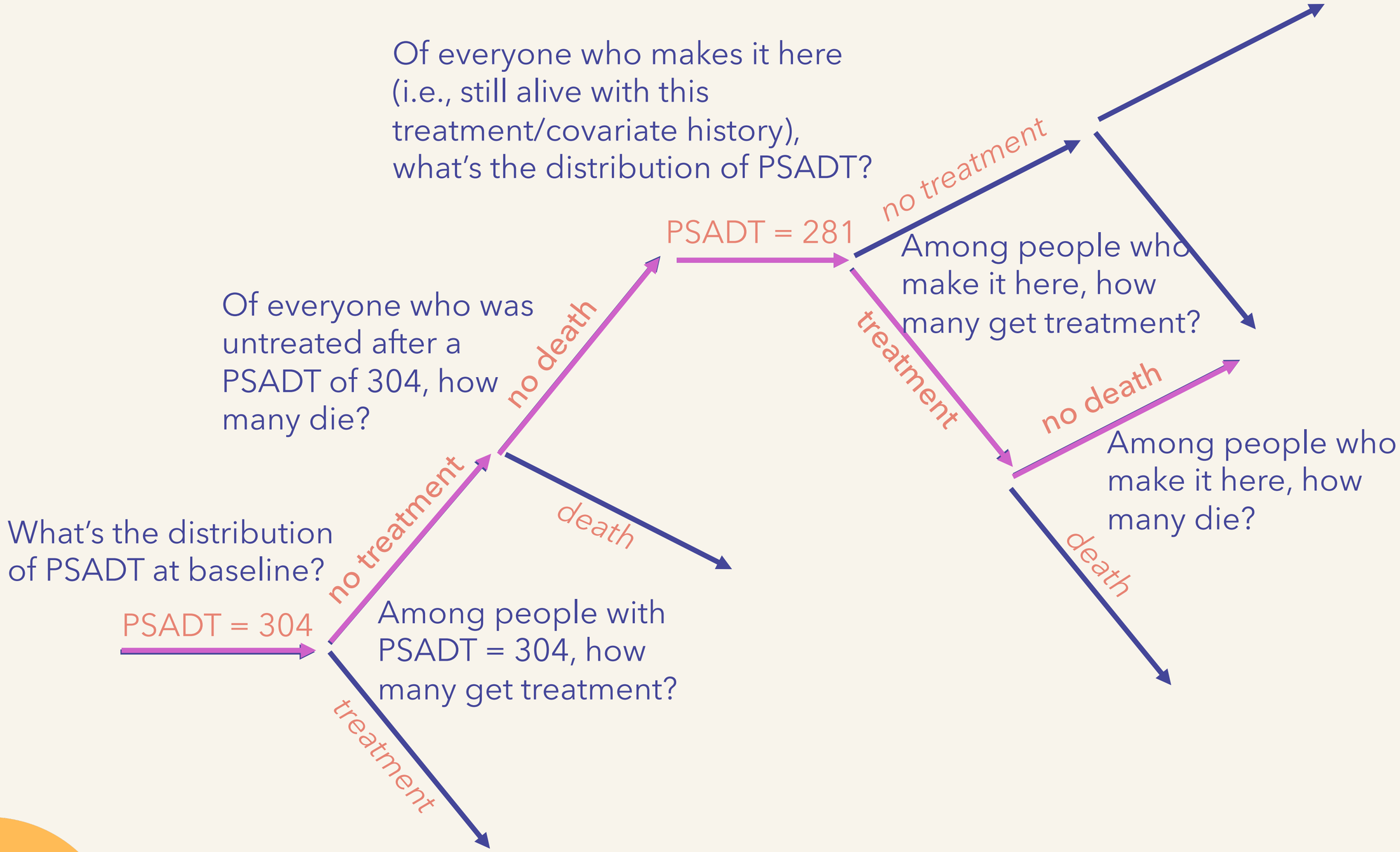
# Per-protocol effect

- Have to use methods that can account for time-varying confounding
- Before we were worried about confounding for an exposure at a single timepoint
  - Are people who get COVID-19 different from those who do not (with respect to their counterfactual risk of preterm delivery)?
- Now the **exposure occurs over time** (every physician visit – treat or not treat?), so we have to worry about how people are different over time and how that may affect whether or not they are exposed

# Methods that can be used with time-varying confounding include…

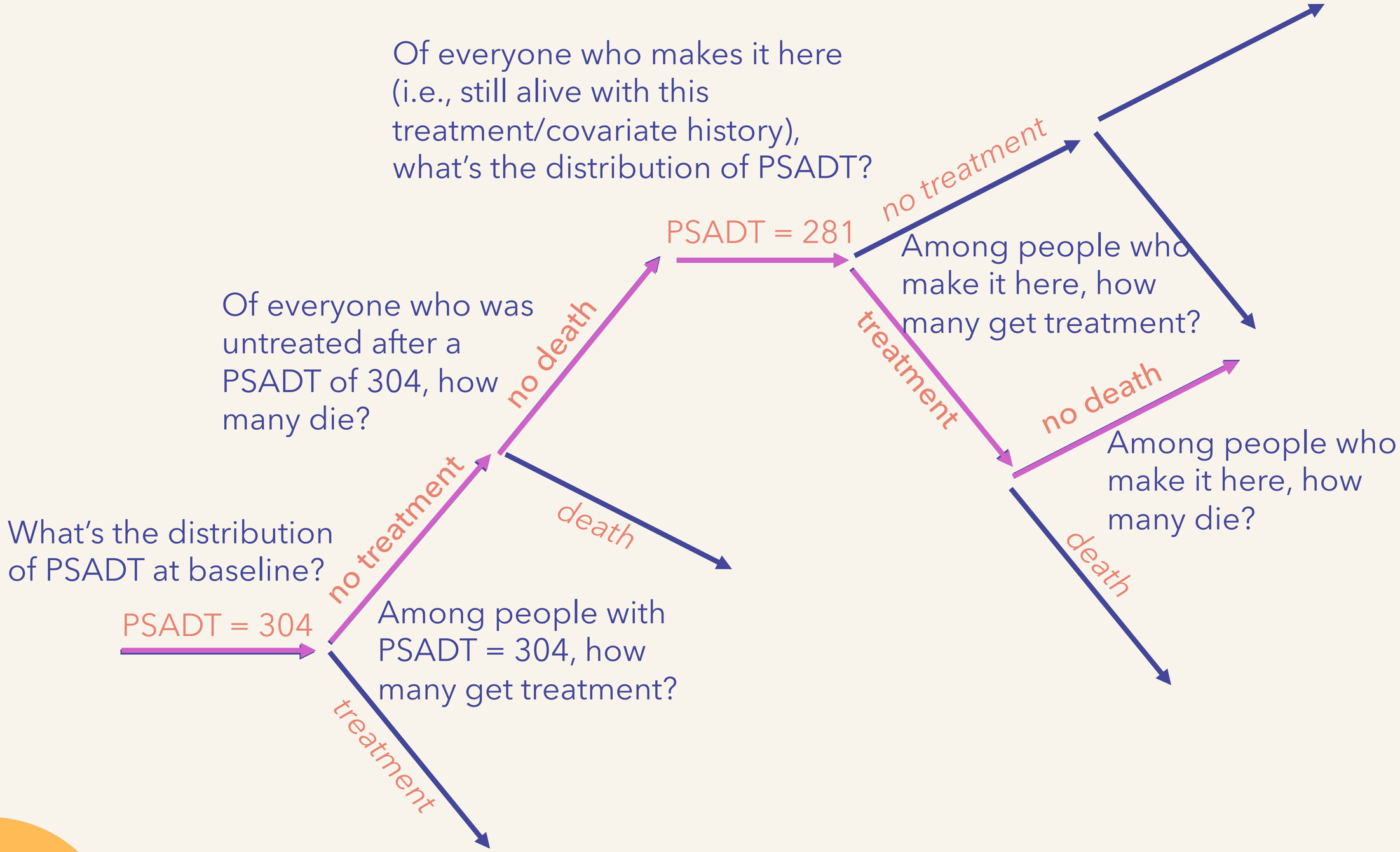
- G-formula
  - Specifically, we'll use the parametric g-formula
- Inverse probability weighting
  - Which we'll use to fit a dynamic marginal structural model


PSADT=304
PSADT=281
no treatment
no death
death
one time period
one time period


What's the distribution of PSADT at baseline?
PSADT = 304
no treatment
death
treatment
Among people with PSADT = 304, how many get treatment?
Of everyone who was untreated after a PSADT of 304, how many die?
no death
PSADT = 281
no treatment
death
treatment
Among people who make it here, how many get treatment?
Of everyone who makes it here (i.e., still alive with this treatment/covariate history), what's the distribution of PSADT?
no death
no treatment
death
Among people who make it here, how many die?

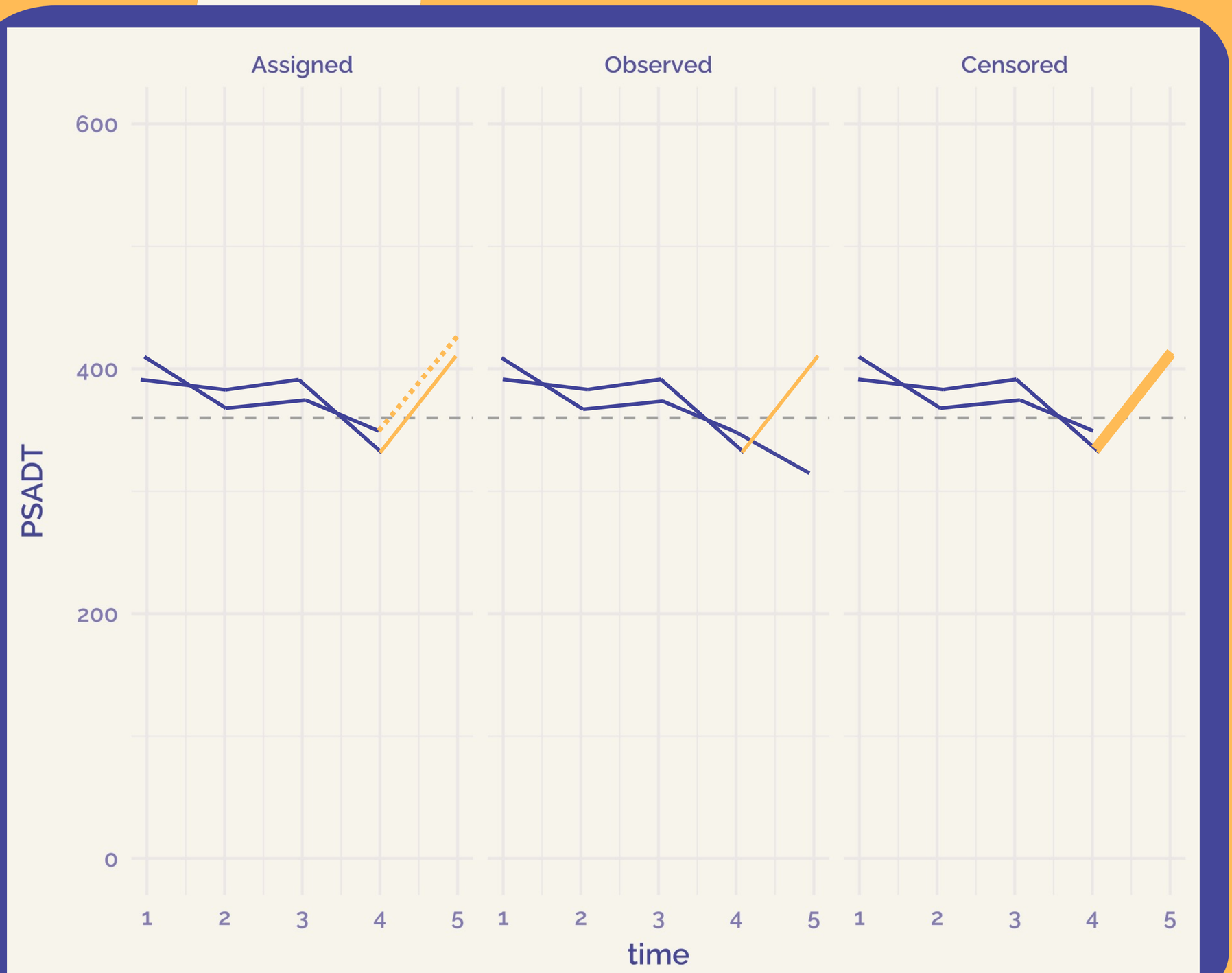
# Parametric g-formula

- Fit **models** for clinic visits, time-varying covariates and all-cause mortality (within each of the 37 treatment arms)
- Draw a large number of observations from the baseline distribution (*start a lot of paths*)
- Use Monte Carlo simulation to progressively assign clinic visits and other time-varying confounders based on models (*choose forks probabilistically*)
  - Assign new PSA, symptom values only when a clinic visit is assigned
  - PSADT is computed directly from most recent and previous assigned values
- Assign treatment according to strategy (*force treatment forks*)
- Use predicted probabilities from outcome model to compute survival curves and risk ratios


What's the distribution of PSADT at baseline?
PSADT = 304
no treatment
death
treatment
Among people with PSADT = 304, how many get treatment?
Of everyone who was untreated after a PSADT of 304, how many die?
no death
PSADT = 281
no treatment
death
treatment
Among people who make it here, how many get treatment?
Of everyone who makes it here (i.e., still alive with this treatment/covariate history), what's the distribution of PSADT?
no death
no treatment
death
Among people who make it here, how many die?

# IP weighting

- Censor everyone who deviates from strategy they were assigned to
  - Treatment before eligible, no treatment by the end of grace period after eligibility, resume treatment after discontinuation, stop monitoring
- Weight uncensored observations by the inverse of the probability they remained uncensored
  - Probabilities estimated from a pooled logistic model for treatment (among untreated) across all timepoints, with splines for month and for PSDAT
  - To approximate uniform initiation, weights adjusted during grace period a factor of $^1/_4$ if initiating in the first month, $^1/_3$ if initiating in the second, $^1/_2$ in the third; and by $^3/_4$, $^1/_3$ , $^1/_2$, respectively, if not yet initiating
  - Censoring due to discontinued monitoring (loss to follow-up) → weights estimated separately




# Per-protocol: Threshold of 360

# Dynamic marginal structural model

- We have a lot of treatment strategies defined by $x$ – so we may not have a lot of people in each one
- Regress indicator of all-cause mortality on $x$, pooled across all timepoints, using the censoring weights
  - $\text{logit}\,\Pr(Y_k^x = 1 \mid Y_{k-1}^x = 0, L_0) = \beta_k + \beta_1 x + \beta_2 L_0$
  - Splines for $x$, time
  - Baseline covariates for precision
- Use hazards estimated from model to compute survival curves, risk ratios comparing different values of $x$

# Aligning the protocol makes the methods comparable

- The two estimators are based on different components of the joint density of the observable data
  - We fit different models
- The estimands are the same because the treatment strategy we defined is the same
- If we left, e.g., duration of treatment unspecified:
  - IP weighting: Don’t censor anyone after treatment initiation
  - G-formula: Keep assigning people treatment forever? Choose a distribution for treatment duration and assign based on that?
  - Strategy requires balance: realistic (censor fewer people) vs. complex (fully specify patterns of treatment starting and stopping)

# When emulating in observational data

- We don’t know treatment assignment. For the g-formula, we can’t estimate the models within each treatment arm but otherwise proceed as in the trial. For IP weighting, we let everyone contribute to each treatment strategy:
  - After fitting treatment and censoring models, make 37 copies of the dataset
  - Censor for deviations for the strategy $x = 0,10,\ldots,360$
  - Compute weights separately for each strategy (using models estimated in the original data)
  - Add a column for $x$

# CaPSURE study

- > 14,000 participants newly diagnosed with prostate cancer from over 40 US clinics
- Physicians provided clinical data (medications, lab tests, imaging)
- Participants complete follow-up survey every 6-12 months
- Eligibility for our target trial: biochemically recurrent prostate cancer after initial radiation/surgery

# Participants in CaPSURE study

Inclusions
9,684 patients with a histological diagnosis of prostate adenocarcinoma, PSA measurements, and imaging tests after diagnosis and no orchiectomy at the time of primary treatment.
8,903 staged ≤ T3aN0M0.
7,658 treated with curative intention (5,416 with RP ± EBRT and 2,242 with EBRT and/or brachytherapy).
1,710 relapsed by PSA (958 in the subset treated with RP ± EBRT and 752 in the subset treated with EBRT and/or brachytherapy).

Exclusions
10 patients underwent orchiectomy before PSA relapse.
275 patients received ADT in the 12 months preceding PSA relapse, or for > 12 months ever.
41 patients presented overt relapse on bone scan, abdominopelvic CT scan or pelvic MRI.
2 patients suffered from severe cancer-related symptoms at relapse.
11 patients had progressed according to their clinical notes at relapse.
81 patients were estimated to have a life expectancy of less than 5 years.
61 patients had a PSA doubling time < 30 days at relapse.

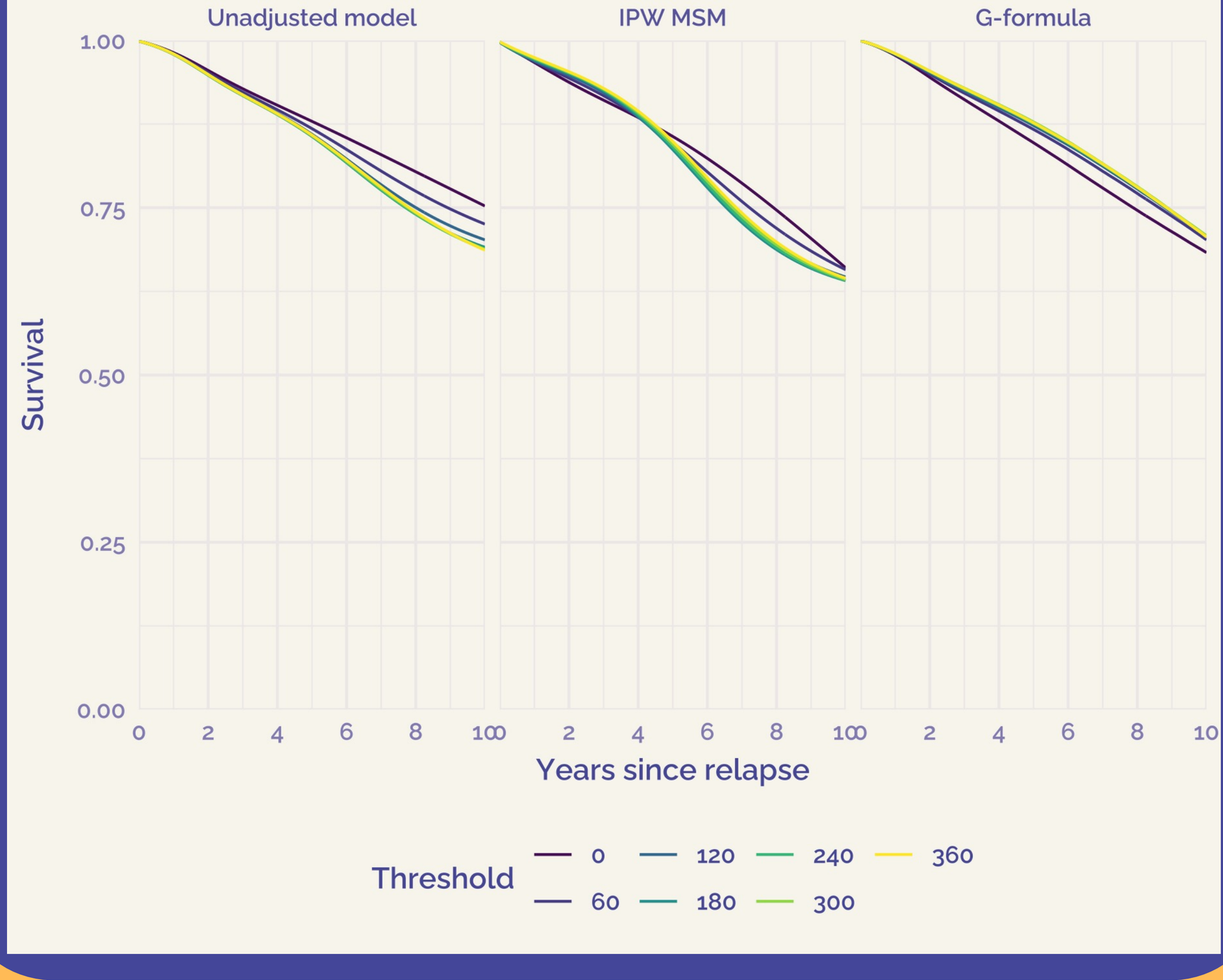
1,229 eligible patients

Treatment initiated
347 patients initiated ADT at some point during 40,946 months of follow-up

141 died during follow-up (69 died of prostate cancer) and 206 were censored.

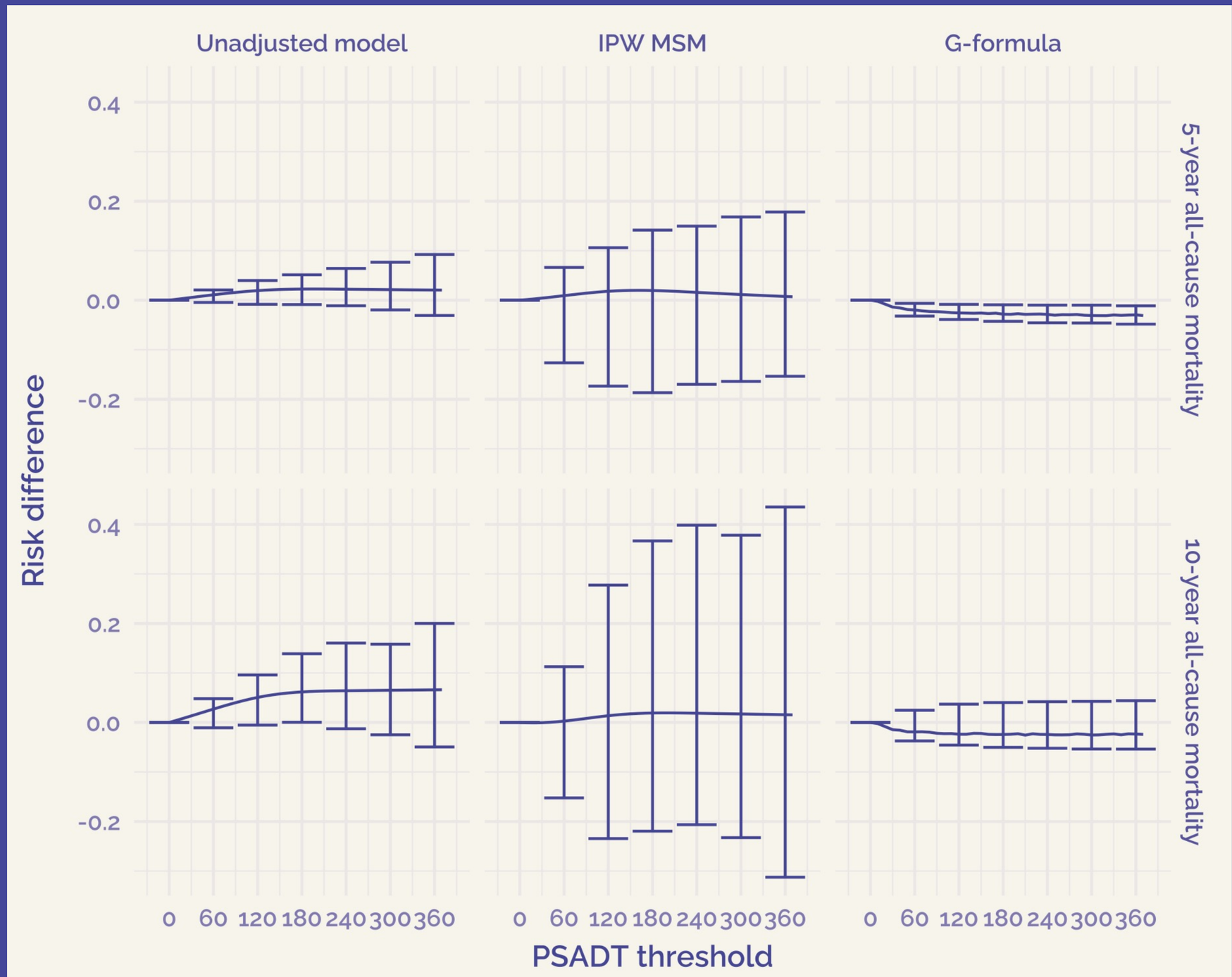
No treatment initiated
882 patients never initiated ADT during 86,157 months of follow-up

150 died during follow-up (17 died of prostate cancer) and 732 were censored.




# Survival curve

# Risk differences



# Strengths and Limitations

## Imprecision

Few participants who followed any given strategy = lots of censoring, few outcomes.

## Over precision?

Model misspecification in the g-formula particularly means confidence intervals small but probably biased.

## Fully specified strategy

Estimates are comparable because estimands are comparable.

# Paper 3

How much bias would change our conclusions from observational studies?

1

## Biases in epidemiology

Where do we go wrong, particularly when using observational data?

2

## A bound for multiple biases

How can we think about confounding, selection, and misclassification simultaneously?

3

## Interpretation and examples

How do we interpret and specify the parameters defining the bound?

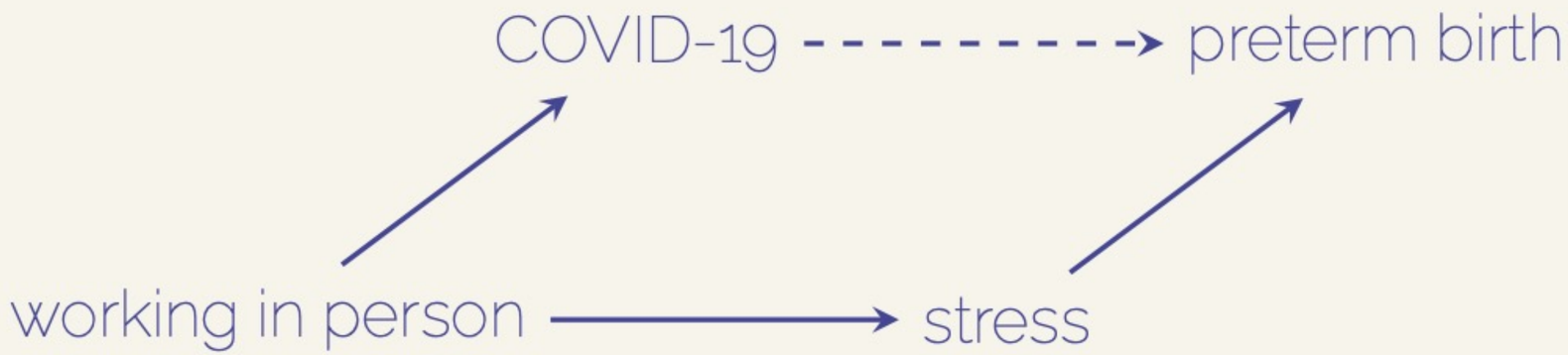
4

## Software

How can we make this easy to do?

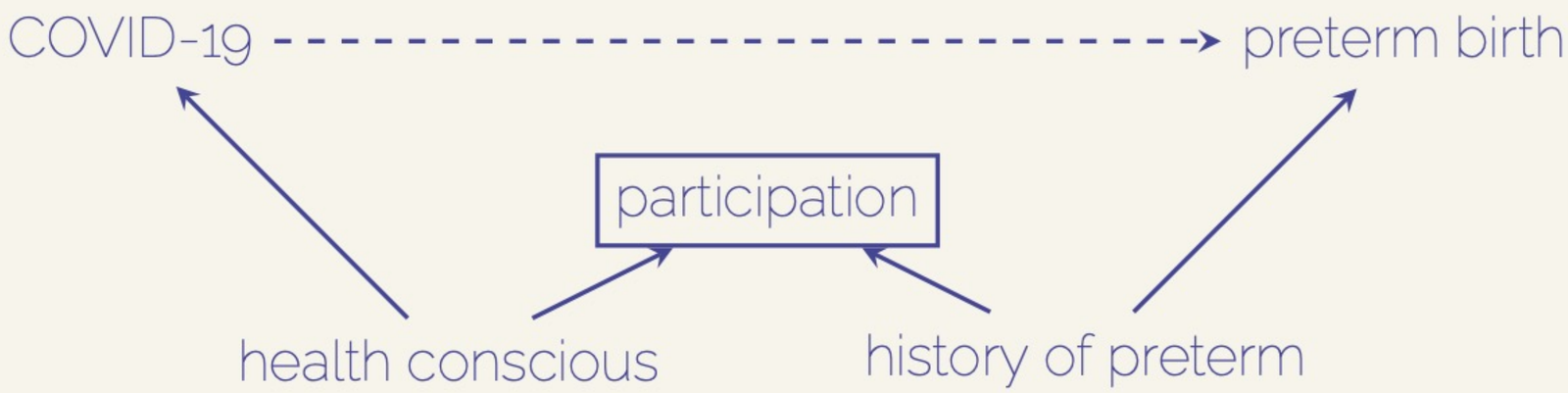
# Observational studies have limitations

- Confounding
  - Most exposures are not randomly distributed, so exposed and unexposed groups tend to differ across many factors
  - We suspected this with COVID-19: people who are infected may be at higher risk of preterm for other reasons



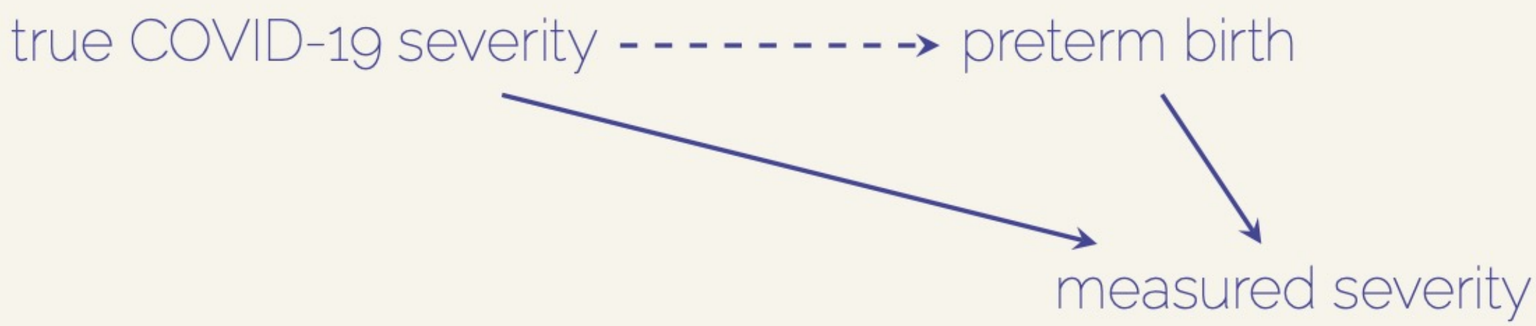
# Observational studies have limitations

- Selection bias
  - People who are recruited, agree to participate, have no missing data, etc. may be different than those who don't show up in the data
  - Imagine that people are more likely to enroll in IRCEP if they are particularly worried about their health (and are therefore less likely to get COVID-19), or if they have a history of  preterm so are motivated to participate in pregnancy research



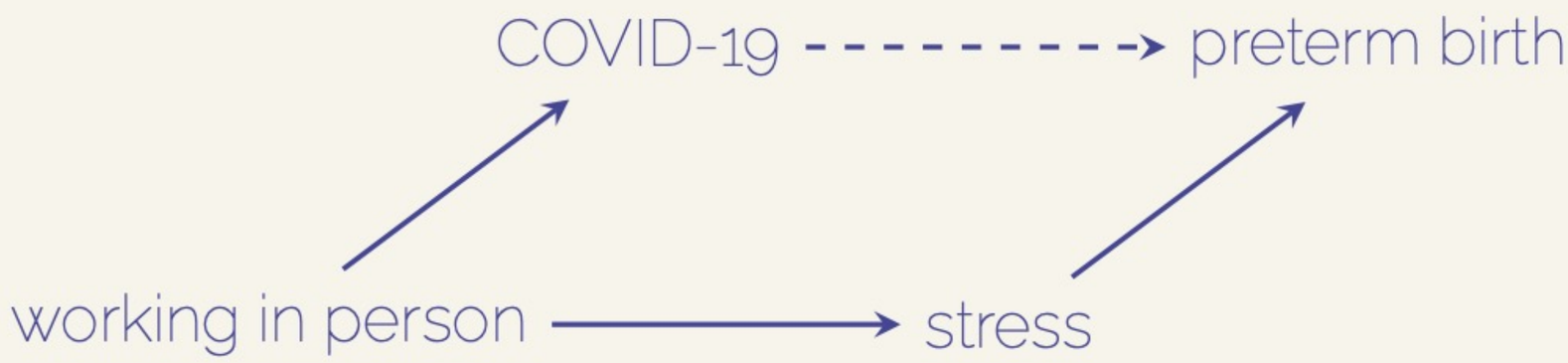
# Observational studies have limitations

- (Differential) misclassification
  - Exposures and outcomes aren't usually measured perfectly
  - This is an even bigger problem if the extent of misclassification of one variable depends on another
  - For example, if someone delivered prematurely and enrolled retrospectively, they might be more likely to recall their symptoms during pregnancy because they attribute the early delivery to COVID-19 (exposure misclassification)
  - Outcome misclassification would occur if gestational age were reported such that, e.g., people with severe disease were more likely to be categorized as delivering preterm



# Bias analysis, briefly

- Even if we can't directly deal with these biases by measuring the factors causing them, what if we tried to figure out how much they affected our results?
- For example, bias due to unmeasured confounding depends on relationship between confounder and exposure and between confounder and outcome, and prevalence of confounder



# Bias analysis, briefly

- Probabilistic methods assign distributions to these relationships
- Bounds were developed that depend on these relationships, and were extended to remove assumptions
- These ideas also extend to other biases
  - Most focus on one bias at a time
  - Focus here on bound for *total* bias from confounding, selection bias, differential misclassification

# Assessing possible bias with a bound

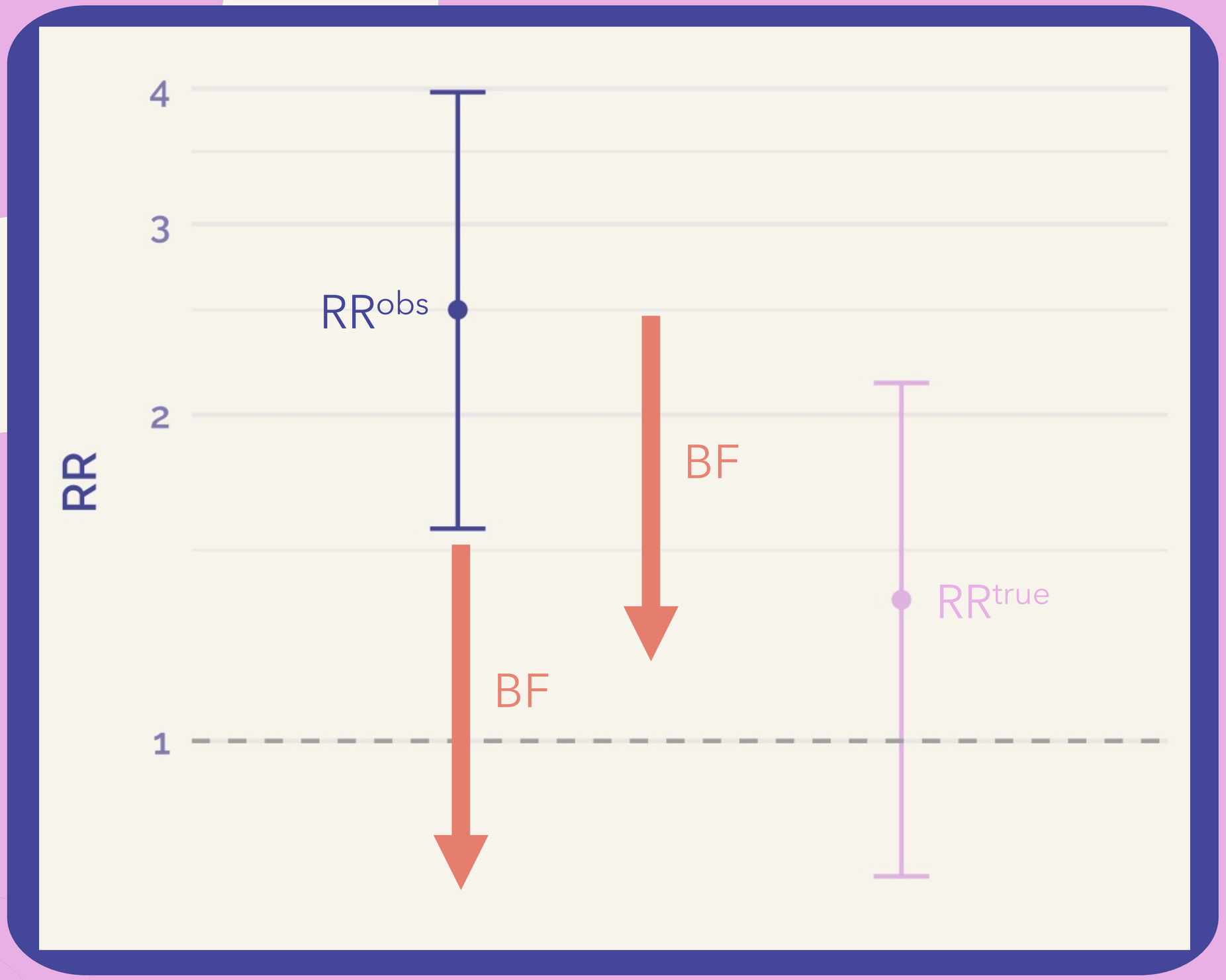
- Idea is to understand the relationship between an observed risk ratio ($RR^{obs}$) and the underlying causal risk ratio ($RR^{true}$)
- In particular, to bound the bias on the multiplicative scale:

$$bias = RR^{obs} / RR^{true}$$

Can we find a bound (BF) such that

$$BF \geq bias$$

so that we can guarantee that $RR^{true} \geq RR^{obs} / BF$?




The bound is *at least* the (relative) difference between the observed and true RR

# What are we estimating with multiple biases at once?

With outcome misclassification:

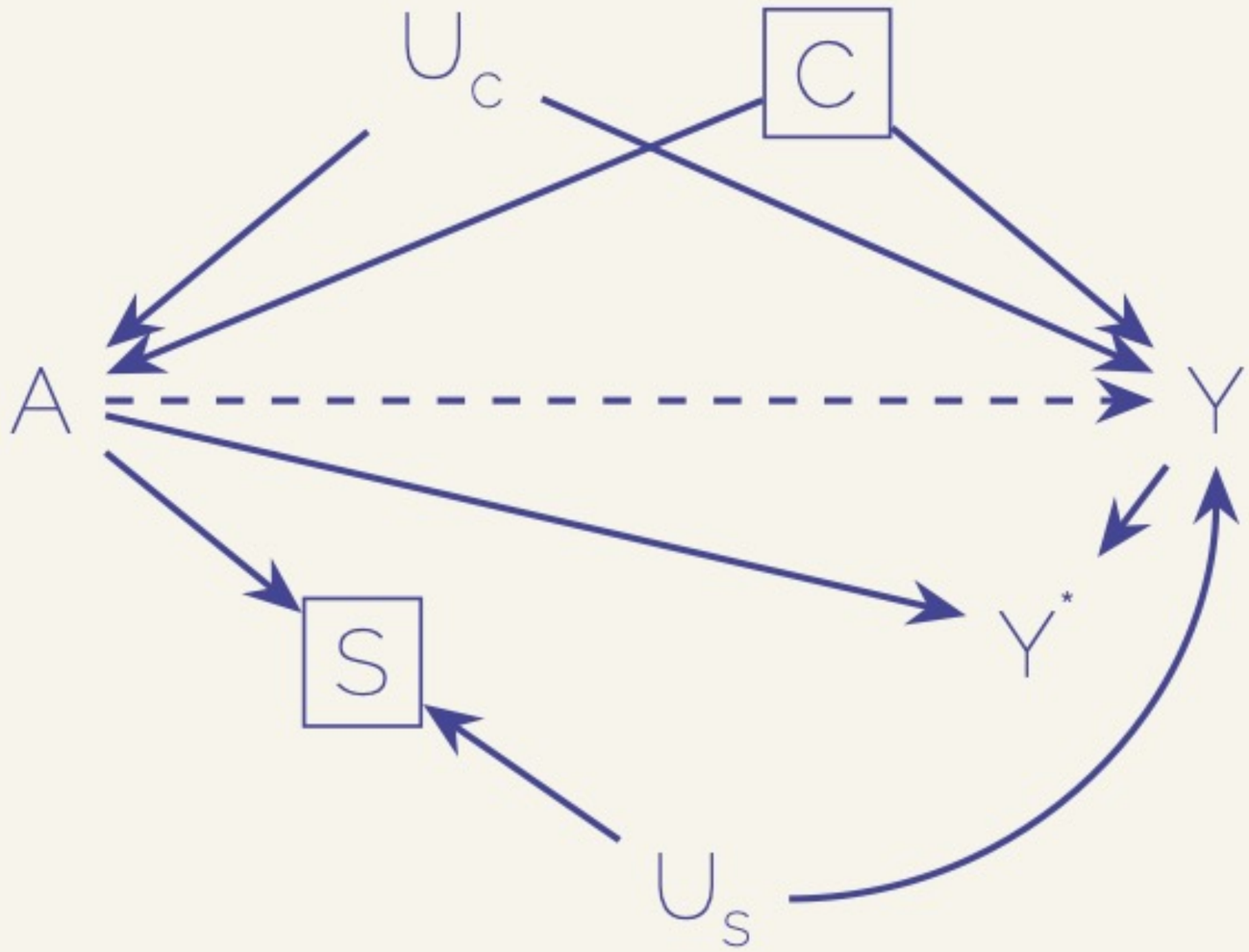
$$\text{RR}_{AY^*}^{\text{obs}} = \frac{\Pr(Y^* = 1 \mid A = 1, S = 1, c)}{\Pr(Y^* = 1 \mid A = 0, S = 1, c)}$$

$$\text{RR}_{AY}^{\text{true}} = \frac{\Pr(Y_1 = 1 \mid c)}{\Pr(Y_0 = 1 \mid c)}$$

$$\text{bias} = \text{RR}_{AY^*}^{\text{obs}} / \text{RR}_{AY}^{\text{true}}$$

where $Y^*$ is a misclassified binary outcome, $A$ the exposure of interest, $S$ an indicator of selection into the study, and $c$ values of covariates conditioned on in the analysis.

# DAG: Differential outcome misclassification




$Y \coprod S \mid A, C, U_S$

$Y_a \coprod A \mid C, U_C$

# Result for outcome misclassification

$$Let\ g(a,b) = \frac{a \times b}{a+b-1}.$$

If $Y_a \coprod A \mid C, U_c$ and $Y \coprod S \mid A, C, U_s$, then:

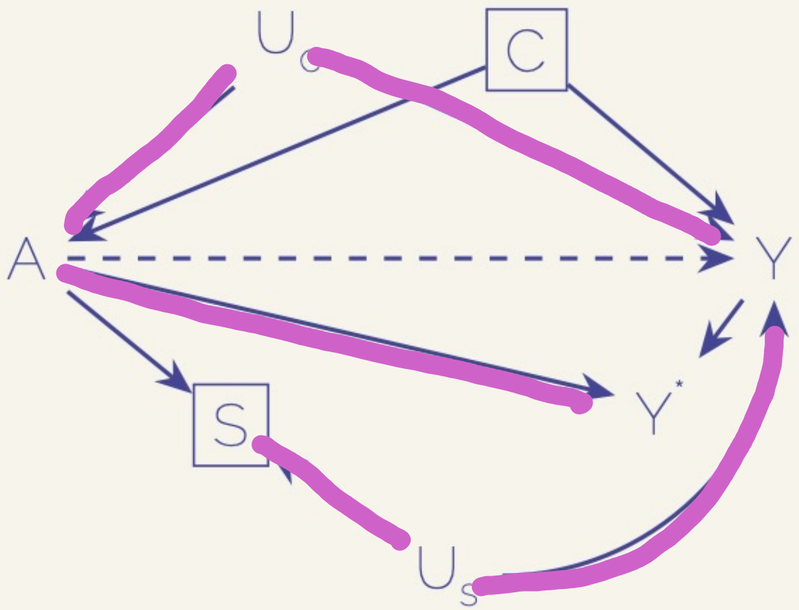
$$\mathrm{RR}_{AY^*}^{\mathrm{obs}} / \mathrm{RR}_{AY}^{\mathrm{true}} \leq \mathrm{BF}_m \times \mathrm{BF}_s \times \mathrm{BF}_c$$

*where:*

$$\mathrm{BF}_m = \mathrm{RR}_{AY^*|y,S=1},$$

$$\mathrm{BF}_s = g\left(\mathrm{RR}_{U_sY|A=1}, \mathrm{RR}_{SU_s|A=1}\right) \times g\left(\mathrm{RR}_{U_sY|A=0}, \mathrm{RR}_{SU_s|A=0}\right)$$

$$\mathrm{BF}_c = g\left(\mathrm{RR}_{AU_c}, \mathrm{RR}_{U_cY}\right).$$




$$\mathrm{RR}_{AY^*|y,S=1} = \max_y \frac{\Pr(Y^* = 1 \mid Y = y, A = 1, S = 1, c)}{\Pr(Y^* = 1 \mid Y = y, A = 0, S = 1, c)}$$

$$\mathrm{RR}_{U_s Y|A=a} = \frac{\max_u \Pr(Y = 1 \mid A = a, c, U_s = u)}{\min_u \Pr(Y = 1 \mid A = a, c, U_s = u)}$$

$$\mathrm{RR}_{SU_s|A=a} = \max_u \frac{\Pr(U_s = u \mid A = a, S = a, c)}{\Pr(U_s = u \mid A = a, S = 1 - a, c)}$$

$$\mathrm{RR}_{U_c Y} = \max_a \frac{\max_u \Pr(Y = 1 \mid A = a, c, U_c = u)}{\min_u \Pr(Y = 1 \mid A = a, c, U_c = u)}$$

$$\mathrm{RR}_{AU_c} = \max_u \frac{\Pr(U_c = u \mid A = 1, c)}{\Pr(U_c = u \mid A = 0, c)}$$

# Interpretation of the parameters

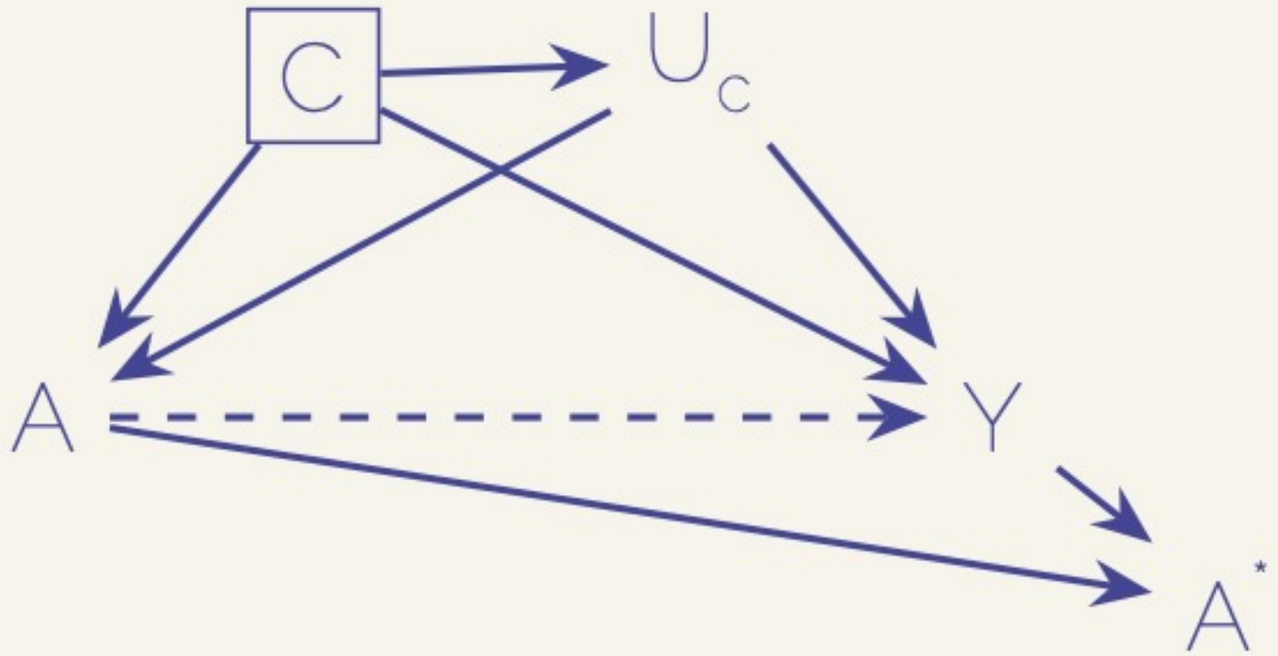
- $RR_{AY^*|y,S=1}$: Among people in IRCEP, how much more likely is someone with severe disease to be falsely categorized as preterm?
- $RR_{U_sY|A=a}$ : How much more likely are people with a history of preterm delivery to have a preterm delivery (among exposed and unexposed)?
- $RR_{SU_s|A=a}$ : How much more likely are the participants than non-participants to be health conscious (among exposed and unexposed)?
- $RR_{U_cY}$ : How much more likely are people who have to work in person to have a preterm delivery?
- $RR_{AU_c}$ : How much more likely are people with COVID-19 to have to work in person?

# Variations

- Exposure misclassification
  - Pertinent parameter would refer to, e.g., how much more likely people with preterm deliveries were to over-report the severity of disease
- Target of inference for selection bias
- “Ordering” of selection bias and misclassification
  - Is selection based on the misclassified data? Or did the misclassification happen within the sample?
  - In IRCEP, reporting happened only among participants
  - Would be the reverse if we selected people with (possibly misclassified) severe disease / preterm from a hospital

# Example: vitamins

- Do prenatal vitamins decrease risk of childhood leukemia?
- Ross et al. found their consumption associated with 50% lower risk of leukemia, conditional on maternal age, race, education.
  - Concern that parents whose children had leukemia overreport *not* taking a vitamin – differential exposure misclassification?
  - Concern that those who take vitamins differ in other ways, e.g., more likely to breastfeed (proxy for confounding by income/other correlates of vitamin use)?
- A probabilistic bias analysis (Jurek et al.) investigated exposure misclassification but ignored confounders




$Y_a \coprod A \mid C, U_C$

# Example: vitamins

# Parameters

- Worried about overreporting of non-vitamin use in families with leukemia vs. not
  - Propose $\Pr(A^* = 0 \mid Y = 1, A = 1) = 0.15$ and $\Pr(A^* = 0 \mid Y = 0, A = 1) = 0.1$
  - Failing to remember taking vitamins when you really did
  - Exposure misclassification parameter is 1.59 (odds ratio scale)
- Worried about unmeasured confounding
  - Not breastfeeding associated with increased risk of leukemia by a factor of 1.22
  - Perhaps mothers who take vitamins 2x as likely to breastfeed as those who do not
  - Conditional on age, race, education

# R package

```r
library(EValue)

leuk_biases <- multi_bias(
   confounding(),
   misclassification("exposure",
      rare_outcome = TRUE,
      rare_exposure = FALSE)
)
multi_bound(
   leuk_biases,
   RRAUc = 2, RRUcY = 1.22, ORYAa = 1.59
)
## [1] 1.747568
```

# Conclusion of example

- BF = 1.75 if our proposed parameters accurately describe the bias
- Not taking vitamins was associated with 2x the risk of leukemia ($RR^{obs} = 1.96$)
- $RR^{true} \geq RR^{obs} / BF$ implies $RR^{true} \geq 1.96 / 1.75 = 1.12$
- Barely above 1, and confidence interval crosses the null
- Result seems sensitive to this combination of biases – explore more fully

# Allow for easy computation of the bound

```
multi_bound(biases,
    RRAUc = 2, RRUcY = 3, RRUsYA1 = 1.5,
    RRSUsA1 = 1.25, RRUsYA0 = 3, RRSUsA0 = 1,
    RRAYyS = 2.5)
## [1] 4.017857
```

```
sapply(seq(1.25, 3, by = .25),
    function(RRAUc) {
      multi_bound(biases, RRAUc = RRAUc,
          RRUcY = 3, RRUsYA1 = 1.5,
          RRSUsA1 = 1.25, RRUsYA0 = 3,
          RRSUsA0 = 1, RRAYyS = 2.5)
  })
## [1] 3.090659 3.443878 3.750000 4.017857
4.254202 4.464286 4.652256 4.821429
```

# What parameters are necessary to specify?

```
summary(biases)
```

```
##                       bias     output argument
## 1             confounding      RR_AUc    RRAUc
## 2             confounding      RR_UcY    RRUcY
## 3               selection RR_UsY|A=1   RRUsYA1
## 4               selection RR_SUs|A=1   RRSUsA1
## 5               selection RR_UsY|A=0   RRUsYA0
## 6               selection RR_SUs|A=0   RRSUsA0
## 7 outcome misclassification RR_AY*|y,S    RRAYyS
```

```
biases
```

```
## The following arguments can be copied and pasted into the
multi_bound() function: RRAUc = , RRUcY = , RRUsYA1 = ,
RRSUsA1 = , RRUsYA0 = , RRSUsA0 = , RRAYyS =
```

# Strengths and Limitations

## Conservative

Even if parameters are correctly specified, bound represents a “worst-case scenario”

## Exposure misclassification

Bound does not hold unless outcome is rare; parameter on OR scale.

## Interpretation

Interpretation and specification of the parameters forces thought.

## Ease of use

Simple calculation with easy-to-use software.