Causal inference in epidemiology using target trial principles: *Applications in pregnancy and prostate cancer*

EPFL Statistics Seminar
Louisa Smith, PhD
September 17, 2021

Target trial principles to address a deceptively simple question

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

- 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
  - If we just look at deliveries overall, we may underestimate the effect of COVID-19 on preterm birth

How can we avoid having to think through all this?

- There's a study design we can use to avoid these types of problems…
- Think through how you would design a randomized controlled trial to answer you question instead: a *target trial*

Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *American Journal of Epidemiology*. 2016;183(8):758-764.

Hernán MA, Sauer BC, Hernández-Díaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *Journal of Clinical Epidemiology*. 2016;79:70-75.

Hernán MA. How to estimate the effect of treatment duration on survival outcomes using observational data. *BMJ*. 2018;360:k182.

Cole SR, Li R, Anastos K, et al. Accounting for leadtime in cohort studies: Evaluating when to initiate HIV therapies. *Statistics in Medicine*. 2004;23(21):3351-3363.

Hernán MA, Alonso A, Logan R, et al. Observational studies analyzed like randomized experiments: An application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology*. 2008;19(6):766-779.

Example: How do we know if a vaccine works?

- 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
- Who is eligible?
  - Excluded groups due to worries about effectiveness (immunosuppressed), protected groups (pregnancy), etc.
- How is treatment decided?
  - Flip of a coin (50%?)
- How is the outcome measured?
  - Symptomatic, test positive
  - Hospitalization/death

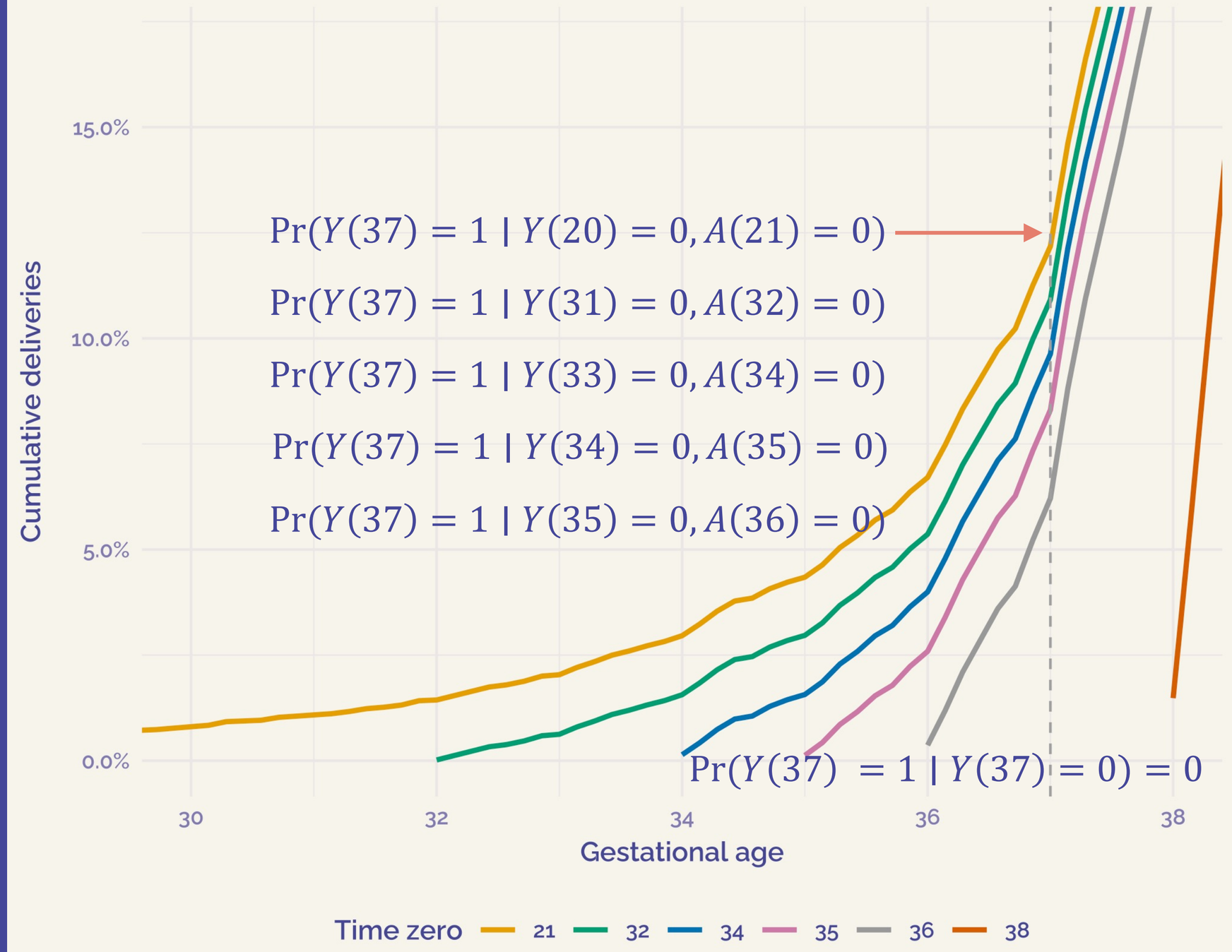
Apply this thinking to our question

- What are the treatment strategies?
  - Get COVID-19 at a specific week (up to week 36) in pregnancy (even assign severity!)
  - Don’t get COVID-19
- Who is eligible?
  - Ongoing pregnancy at that week, never had COVID-19
  - Different eligibility group for every week of gestation (time zero)
- 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

Some notation

- $A(x) = 1$ if assigned COVID−19 in week x
- $A(x) = 0$ if assigned no COVID−19 in week x
- $Y(x) = 1$ if week of delivery $\leq x$, 0 otherwise

- $\Pr(Y(37) = 1 \mid A(21) = 0, Y(20) = 0)$ is risk of preterm in the trial arm in which no COVID-19 was assigned in week 21
- compare: $\Pr(Y(37) = 1 \mid A(33) = 0, Y(32) = 0)$
  - Risks in placebo arms will naturally differ



Outcome measure: Cumulative deliveries starting at different "time zeros"

Trial is answering a counterfactual question

- If you are assigned COVID-19 in a certain week (time zero) during pregnancy, what is the probability of preterm delivery?
  - How does that compare to if assigned to no COVID-19?
- We will say $Y^{20\pm}(37) = Y(37)$ when assigned $A(20) = a$
  - $Y^{20+}(37)$ for $A(20) = 1$ and $Y^{20-}(37)$ for $A(20) = 0$
  - $Y^{21+}(37)$ for $A(21) = 1$ and $Y^{21-}(37)$ for $A(21) = 0$
  - and so on…
- $Y^{20\pm}(37)$ is the counterfactual delivery status at 37 weeks (so preterm or term), had someone been assigned COVID-19 or not at 20 weeks

Estimate counterfactual quantities by *emulating* trial in observational data

- At each week of gestation (time zero), choose the people who developed COVID-19 that week
  - Observations used to estimate $\Pr(Y^{k+}(37) = 1 \mid Y(k-1) = 0)$ for k = 20, …, 36
- 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"
  - Used to estimate $\Pr(Y^{k-}(37) = 1 \mid Y(k-1) = 0)$

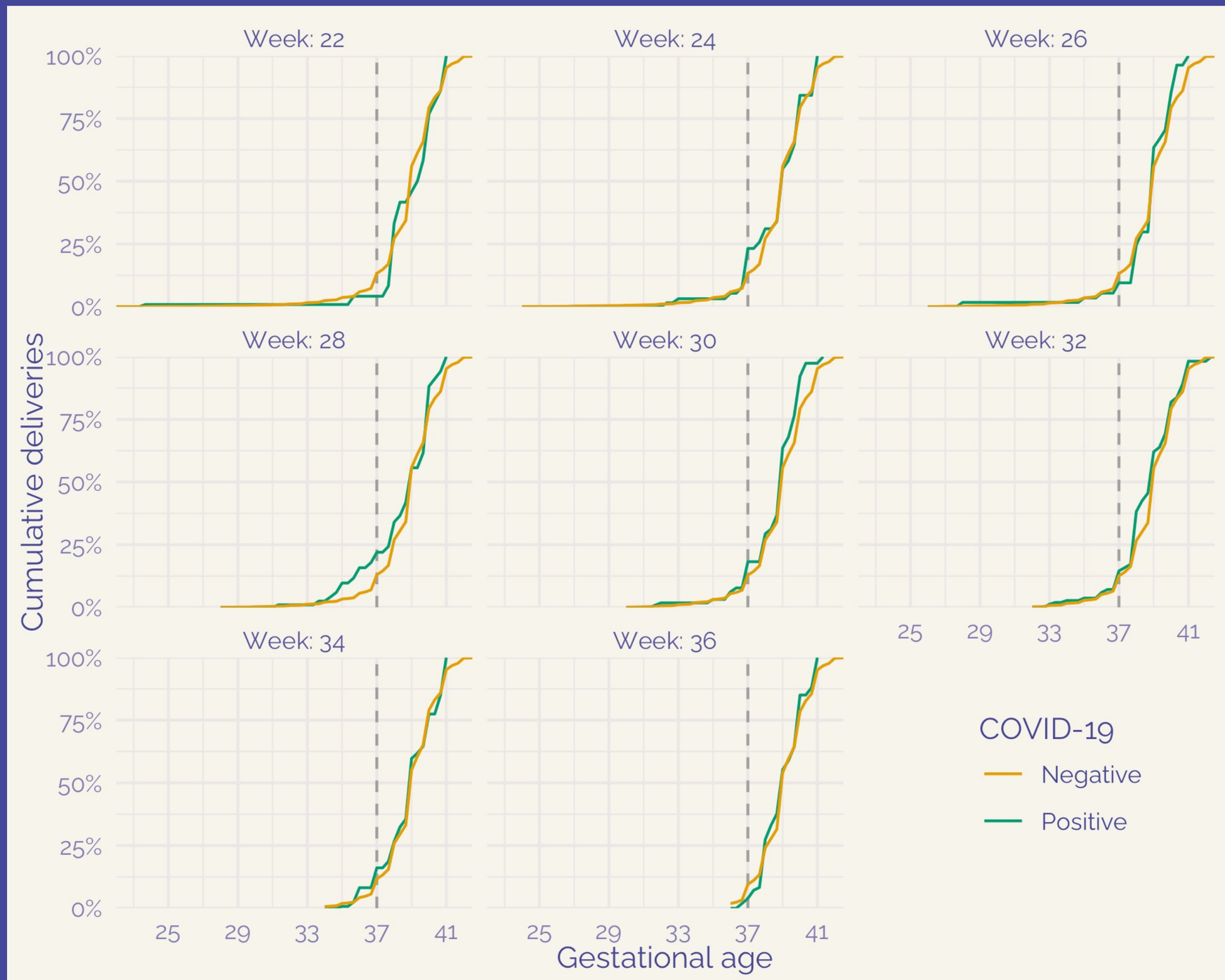
- Confounding, loss to follow-up…

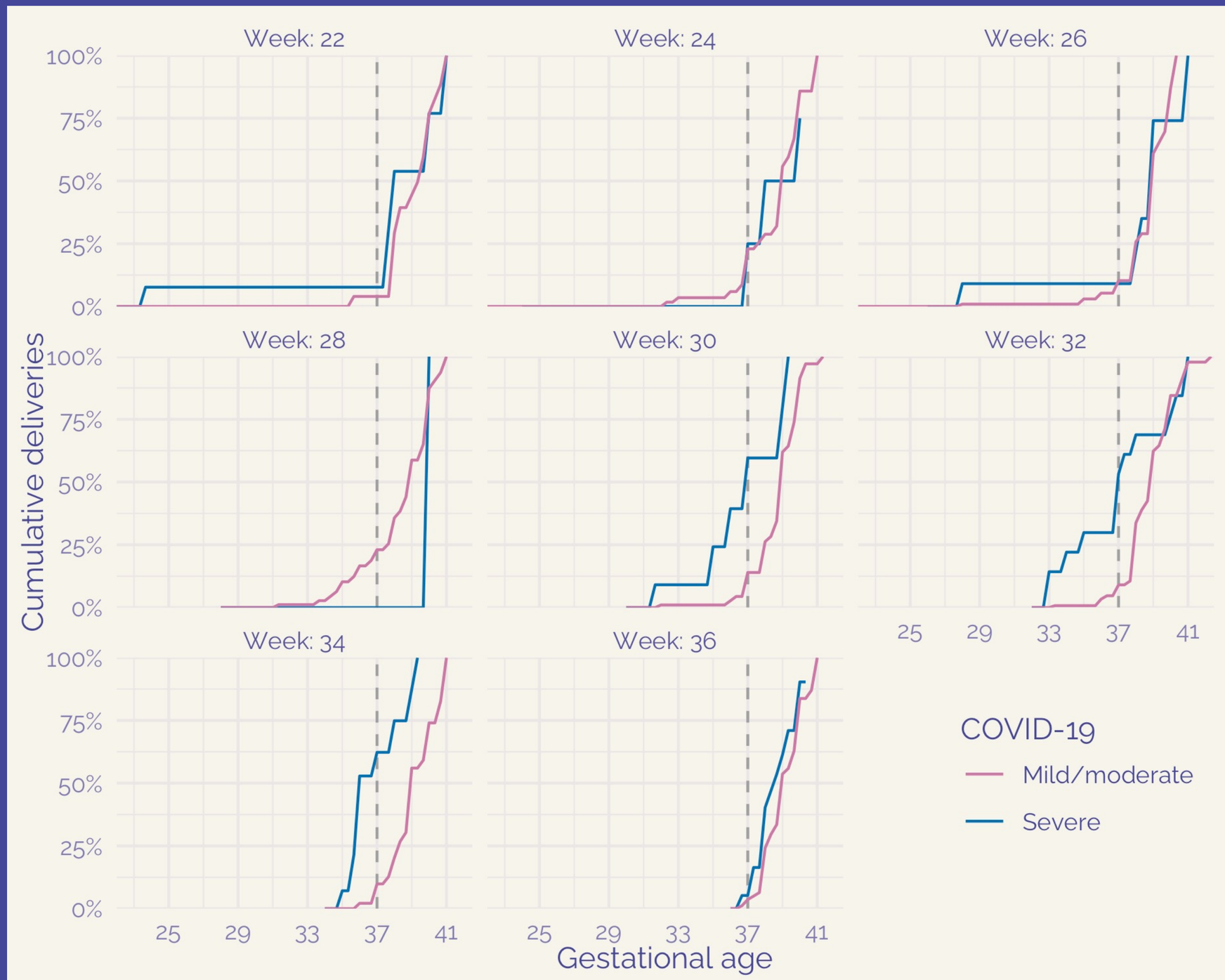
Study design and population

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
- About 14,000 eligible for this analysis

Unadjusted cumulative deliveries





Unadjusted cumulative deliveries

Standardized cumulative delivery curves

- Model weekly (daily) hazard of delivery
  - Pooled logistic regression over person-time data
  - 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 baseline hazard 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+
- Use estimated hazards to compute risks of delivery before 37 weeks, standardized over observed distribution of covariates in test-negative participants still pregnant

e.g., to estimate $\widehat{\Pr}(Y^{k-}(37) = 1 \mid Y(k-1) = 0)$

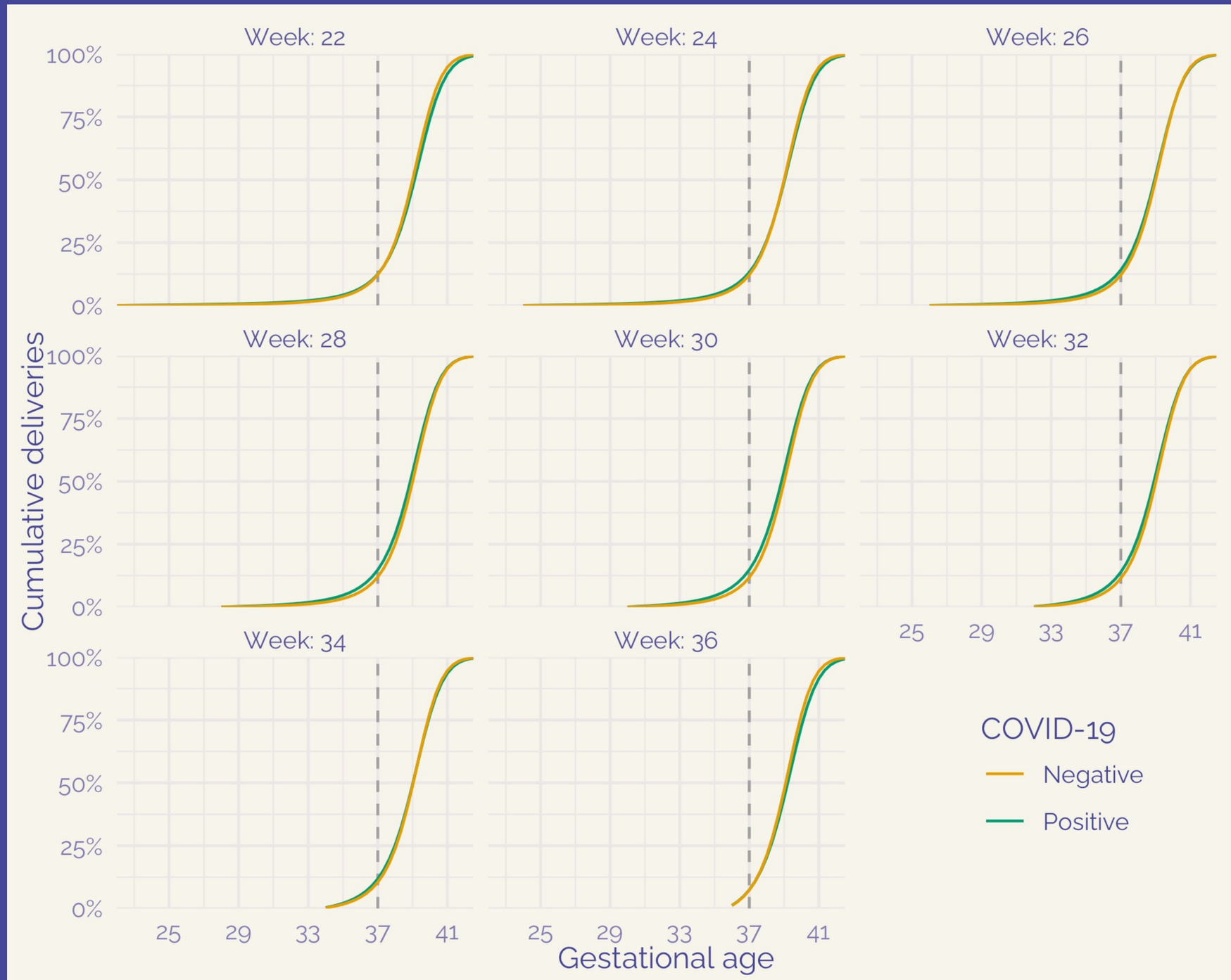
$$1 - \sum_{c} \left\{ \prod_{j=k}^{37} \widehat{\Pr}(Y(j) = 0 \mid Y(j-1) = 0, A(k) = 0, C = c) \right\} \times$$

$$\widehat{\Pr}(C = c \mid Y(k-1) = 0, A(k-1) = 0)$$

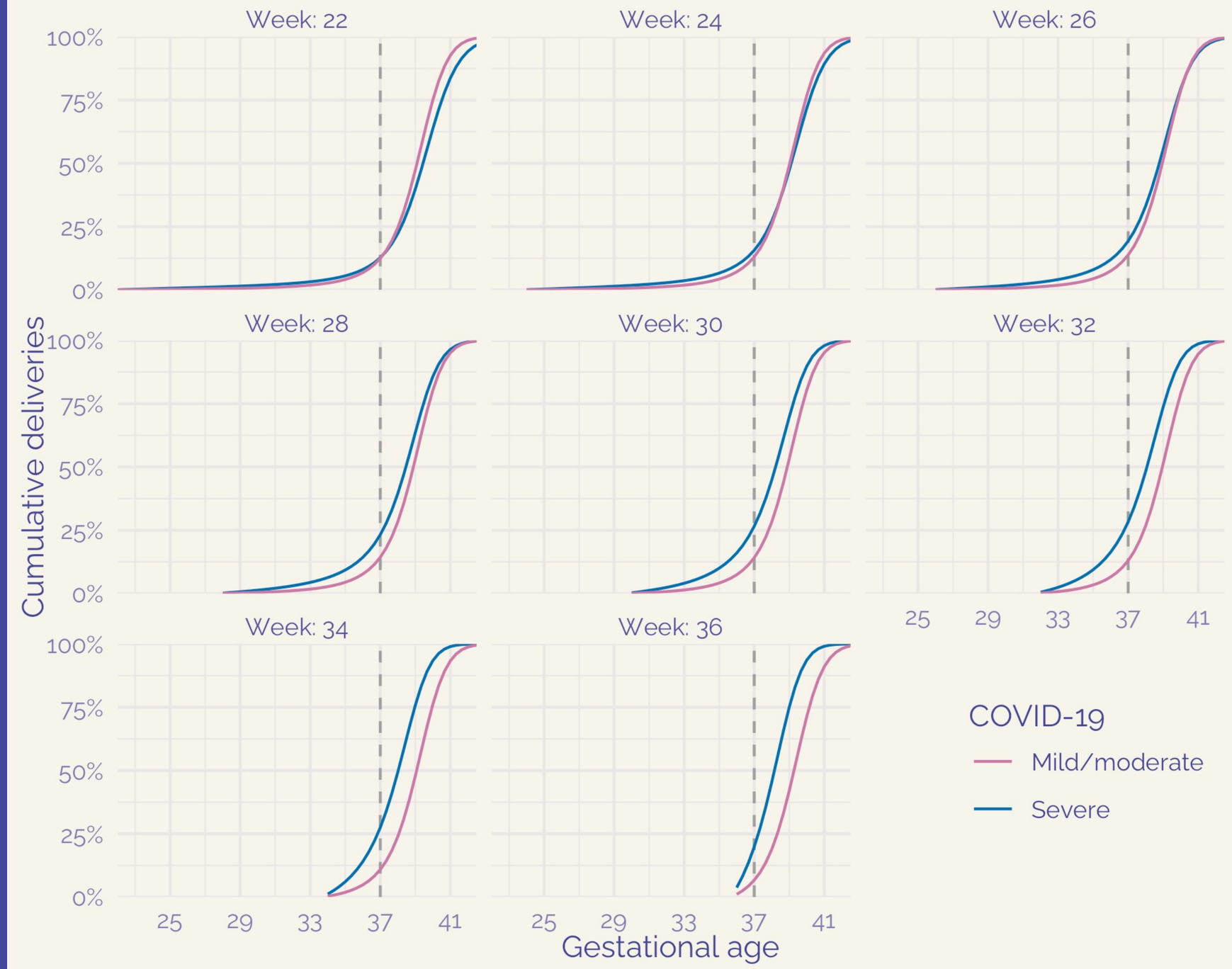
for baseline covariates C

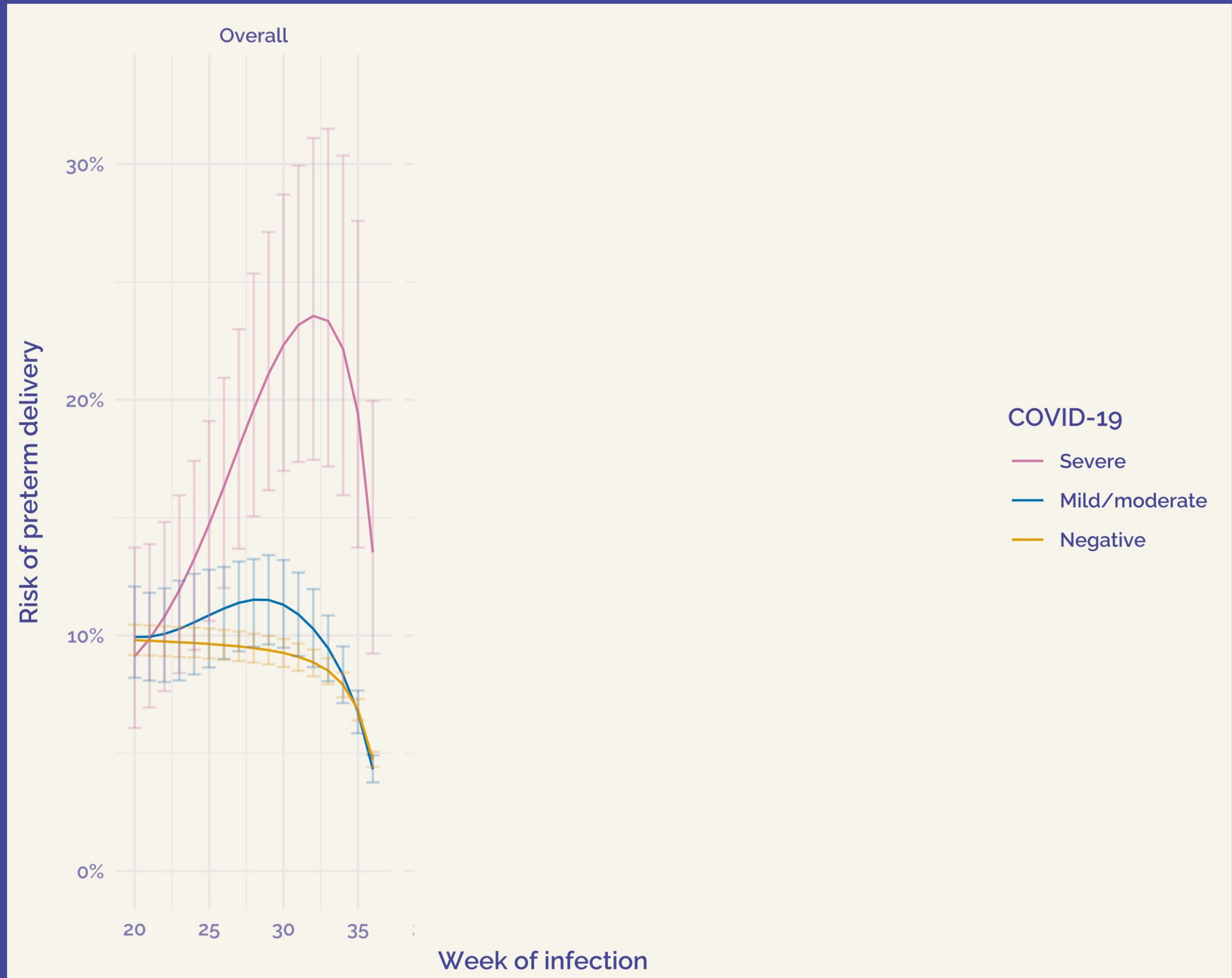
(In a trial, couldn't get access to the drug (i.e., COVID-19) later if not randomized to it at baseline, so intent-to-treat = per-protocol…. due to specifics of our observational study design, made the same assumption – no time-varying confounding.)

Adjusted cumulative deliveries



Adjusted cumulative deliveries





Risks over pregnancy

Risk ratios for preterm delivery

	Positive vs. Negative	Mild/moderate vs. negative	Severe vs. Mild/moderate
Week 0-20	1.0 (0.8, 1.2)	1.0 (0.8, 1.2)	0.9 (0.7, 1.3)
Week 21	1.0 (0.8, 1.2)	1.0 (0.8, 1.2)	1.0 (0.8, 1.4)
Week 22	1.0 (0.8, 1.3)	1.0 (0.8, 1.2)	1.1 (0.8, 1.4)
Week 23	1.1 (0.8, 1.3)	1.1 (0.8, 1.3)	1.2 (0.9, 1.5)
Week 24	1.1 (0.9, 1.3)	1.1 (0.9, 1.3)	1.3 (1.0, 1.6)
Week 25	1.1 (0.9, 1.4)	1.1 (0.9, 1.4)	1.4 (1.1, 1.7)
Week 26	1.2 (1.0, 1.4)	1.2 (1.0, 1.4)	1.5 (1.2, 1.9)
Week 27	1.2 (1.0, 1.4)	1.2 (1.0, 1.4)	1.6 (1.3, 2.0)
Week 28	1.3 (1.1, 1.5)	1.2 (1.0, 1.4)	1.7 (1.4, 2.2)
Week 29	1.3 (1.1, 1.5)	1.2 (1.0, 1.4)	1.8 (1.5, 2.4)
Week 30	1.3 (1.1, 1.5)	1.2 (1.1, 1.4)	2.0 (1.5, 2.6)
Week 31	1.3 (1.1, 1.5)	1.2 (1.0, 1.4)	2.1 (1.6, 2.8)
Week 32	1.2 (1.1, 1.5)	1.2 (1.0, 1.3)	2.3 (1.7, 3.1)
Week 33	1.2 (1.0, 1.4)	1.1 (0.9, 1.3)	2.5 (1.8, 3.4)
Week 34	1.2 (1.0, 1.3)	1.1 (0.9, 1.2)	2.7 (1.9, 3.7)
Week 35	1.1 (1.0, 1.3)	1.0 (0.9, 1.1)	2.9 (2.0, 4.1)
Week 36	1.0 (0.9, 1.2)	0.9 (0.8, 1.0)	3.1 (2.1, 4.7)

Conclusions

Severe disease

Increased risk from COVID-19 is primarily due to severe, rather than mild or moderate, disease, and its effects on induced delivery.

Timing of infection

Third trimester infections, when severe, clearly lead to increases in preterm delivery, but earlier infections have less effect.

Applying target trial principles to a treatment that varies over time

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?

Biochemical recurrence of prostate cancer

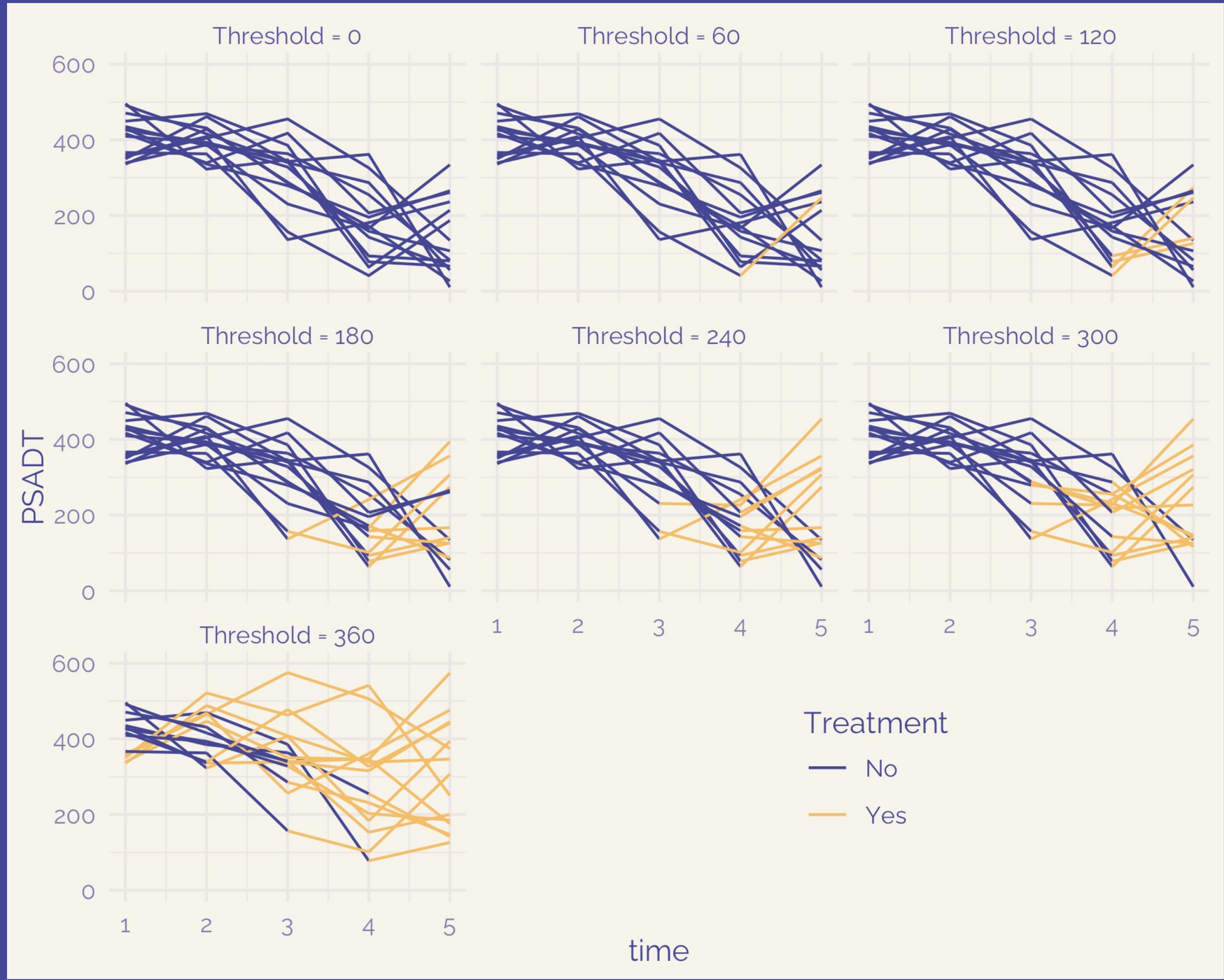
- After successful treatment of early stage cancer, a rise in prostate-specific antigen (PSA)
- Might not lead to death from prostate cancer
  - It can be slow growing
  - It can occur toward the end of the natural lifespan
- Treatment is with androgen deprivation therapy
  - This causes side effects, negative quality of life
  - May be expensive, time-consuming, etc.
- No definite improvement in treating immediately vs. waiting ~2 years
- Goal: treat only if you need it!

More nuanced treatment strategy

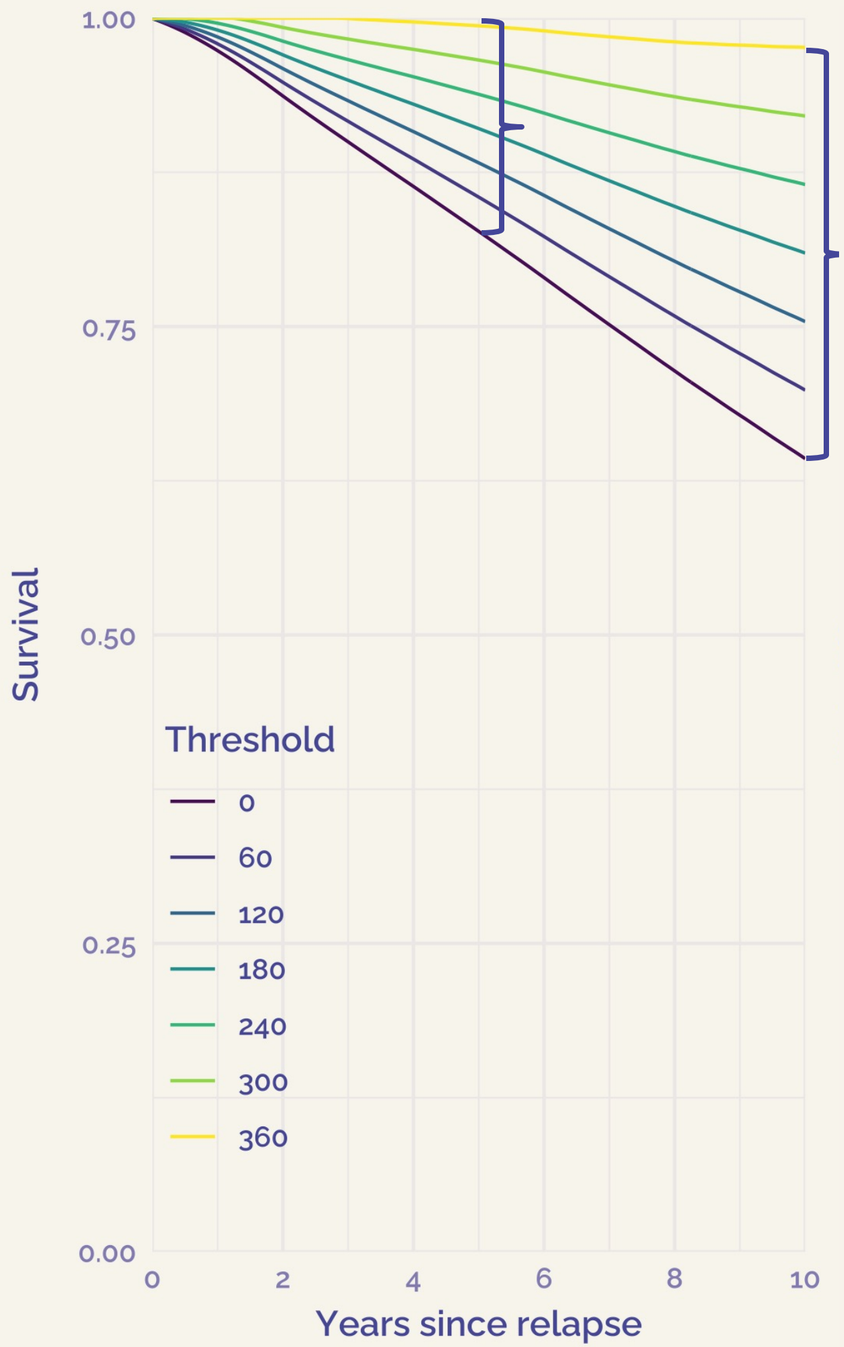
- Start treatment not only when there are clear signs of progression (may be too late in some cases), but also when PSA characteristics indicate that cancer is growing more quickly
- A common measure of growth is PSA doubling time (PSADT)
  - If PSA is rising quickly, it doubles in a short amount of 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
- Trial with 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)
- Counterfactual of interest indexed by x
  - Where $\Pr(Y_{60}^{x} = 1)$ is 60-month (5-year) mortality under treatment strategy x
  - (Note that I'm now indexing time, in months, with subscripts, and all treatment strategies have the same time zero – no conditioning)



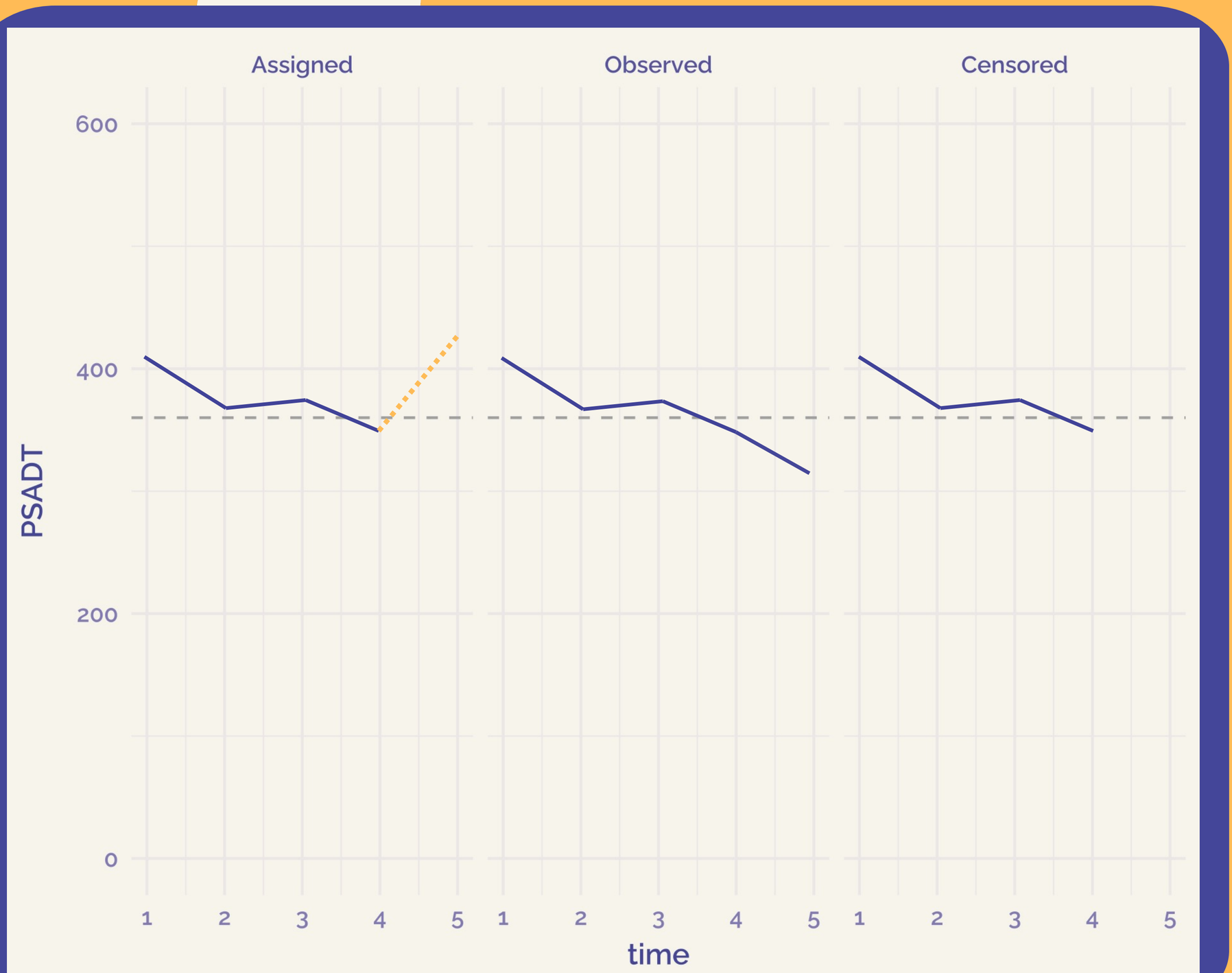
Treatment at PSADT threshold



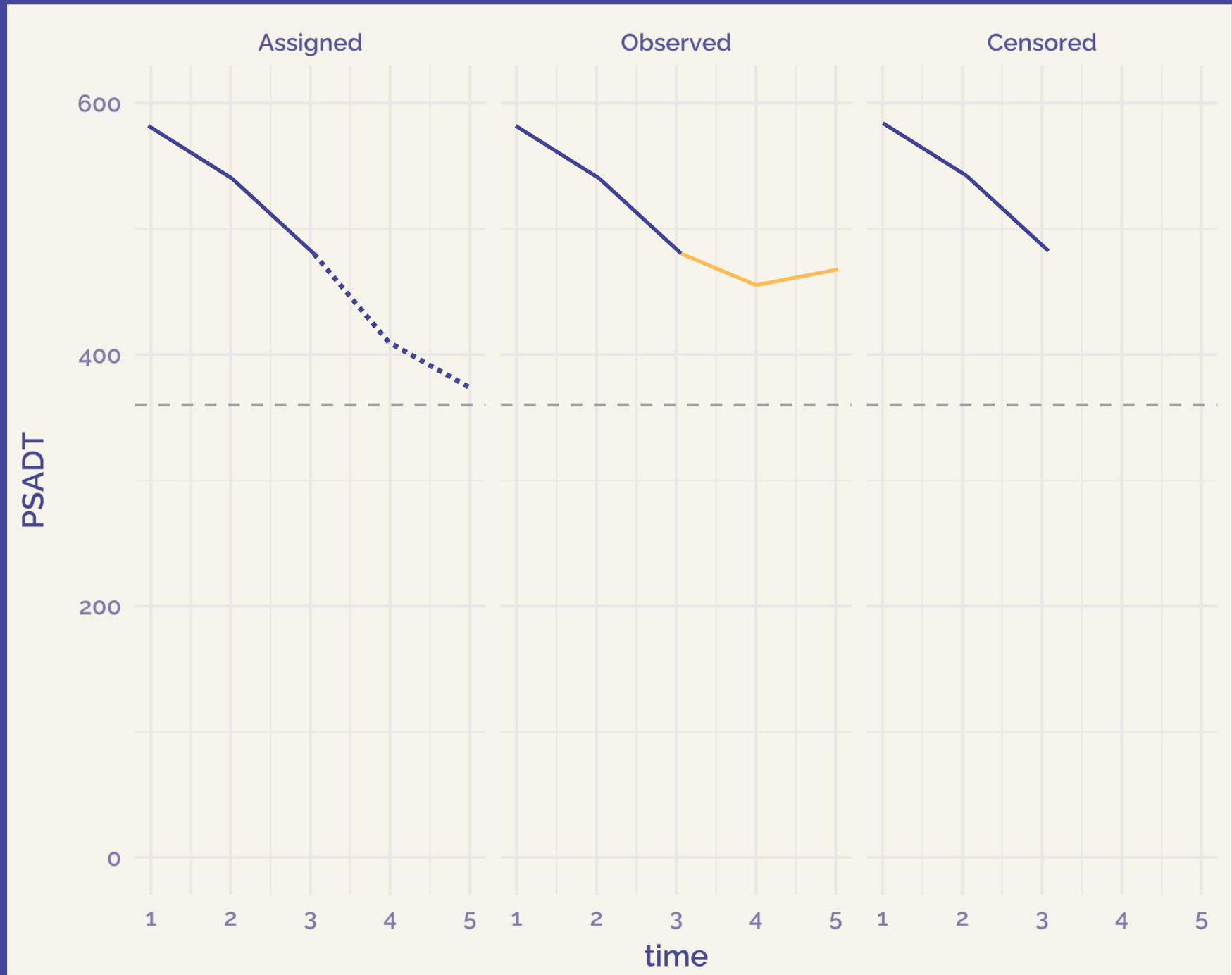
Comparisons using hypothetical trial data

Per-protocol estimand

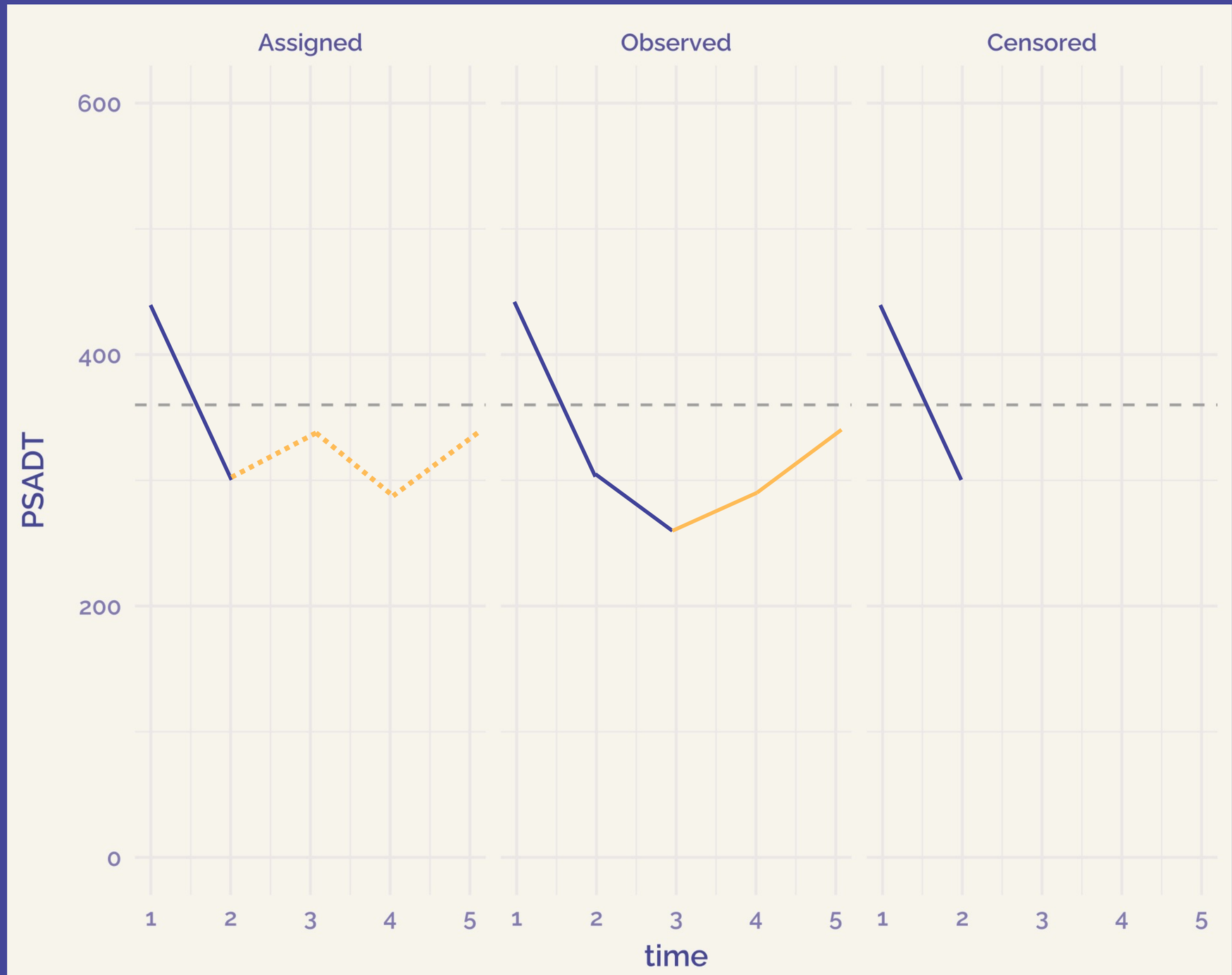
- What if there's non-adherence to the treatment strategy?
- In the COVID-19 trial we didn't really worry about this – we imagined that everyone who was randomized to either get COVID-19 or not get it actually followed through with it
  - In a real trial, depending on the treatment and the trial design, 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 – lack of positivity, more realistic

"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."

Time-varying treatment: affects and depends on time-varying covariates

- 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
- Under assumptions
  - Most importantly, we assume $(Y_k^x, \ldots, Y_K^x) \coprod A_k \mid \overline{L_k} = \overline{l_k}, \overline{A_{k-1}} = \overline{a_{k-1}}, Y_{k-1} = 0$
  - Counterfactual outcomes independent of treatment given the observed history
  - We have measured time-varying confounders L
    - e.g., biomarkers, including PSA, and symptoms that physicians are using to determine whether to start treatment in observed data

G-formula

- We can identify the risk of death by time K under the treatment strategy in which treatment is assigned at threshold x (according to probability density $f_{a_s}^{x}$) with the expression:

$$\Pr\left(Y_K^{x} = 1\right) =$$

$$\int_{\bar{l}_k} \sum_{\bar{a}_k} \sum_{j=0}^{K} \Pr\left(Y_j = 1 \mid \bar{l}_j, \bar{a}_j, Y_{j-1} = 0\right) \times$$

$$\prod_{s=0}^{j} f\left(l_s \mid \bar{l}_{s-1}, \bar{a}_{s-1}, Y_{s-1} = 0\right) \times$$

$$f^{x}(a_s \mid \bar{l}_s, \bar{a}_{s-1}, Y_{s-1} = 0) \times$$

$$\Pr(Y_{s-1} = 0 \mid \bar{l}_{s-1}, \bar{a}_{s-1}, Y_{s-2} = 0)$$

Parametric g-formula

- Fit **models** for clinic visit in a given month, time-varying covariates, and all-cause mortality – given history
- Draw a large number of observations from the baseline covariate distribution
- Use Monte Carlo simulation to progressively assign clinic visits and other time-varying confounders based on models
  - 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
  - $f^x(a_s \mid \bar{l}_s, \bar{a}_{s-1}, Y_{s-1} = 0)$ term that we determine
  - Equal probabilities of initiating during each month of grace period, 0 before eligible, 1 after grace period
- Use predicted probabilities from outcome model to compute survival curves and risk ratios

Inverse-probability weighting

We can also identify the risk of death by time K with the expression

$$\Pr(Y_K^x = 1) = \sum_{k=1}^{K} \mathrm{E}\left[Y_k^x \prod_{s=0}^{k} \frac{\mathrm{I}(A_s = g^x(\bar{L}_s, \bar{A}_{s-1}), Y_{s-1} = 0)}{f_{a_s}(A_s \mid \bar{L}_s, \bar{A}_{s-1}, Y_{s-1} = 0)}\right] \times \prod_{j}^{k-1} \left\{1 - E\left[Y_j^x \prod_{r=0}^{j} \frac{\mathrm{I}\left(A_r = g^x(\bar{L}_r, \bar{A}_{r-1}), Y_{r-1} = 0\right)}{f_{a_r}(A_r \mid \bar{L}_r, \bar{A}_{r-1}, Y_{r-1} = 0)}\right]\right\}$$

where $g^x(\bar{l}_s, \bar{a}_{s-1})$ tells us what treatment A_s would be assigned under our treatment strategy defined by x if someone had treatment and covariate history $\bar{l}_s, \bar{a}_{s-1}$.

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
  - If not, estimates a “representative intervention” (Young, et al. JASA 2019)

When emulating in observational data

- We don’t know treatment assignment. 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

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
  - $\text{logit}\ \Pr(\ Y_k^x = 1 \mid Y_{k-1}^x = 0, L_0\ ) = \beta_k + \beta_x + \beta_2 L_0$
  - Splines for x, time
  - Baseline covariates for precision
- Use hazards estimated from model to compute survival curves, risk ratios comparing different values of x

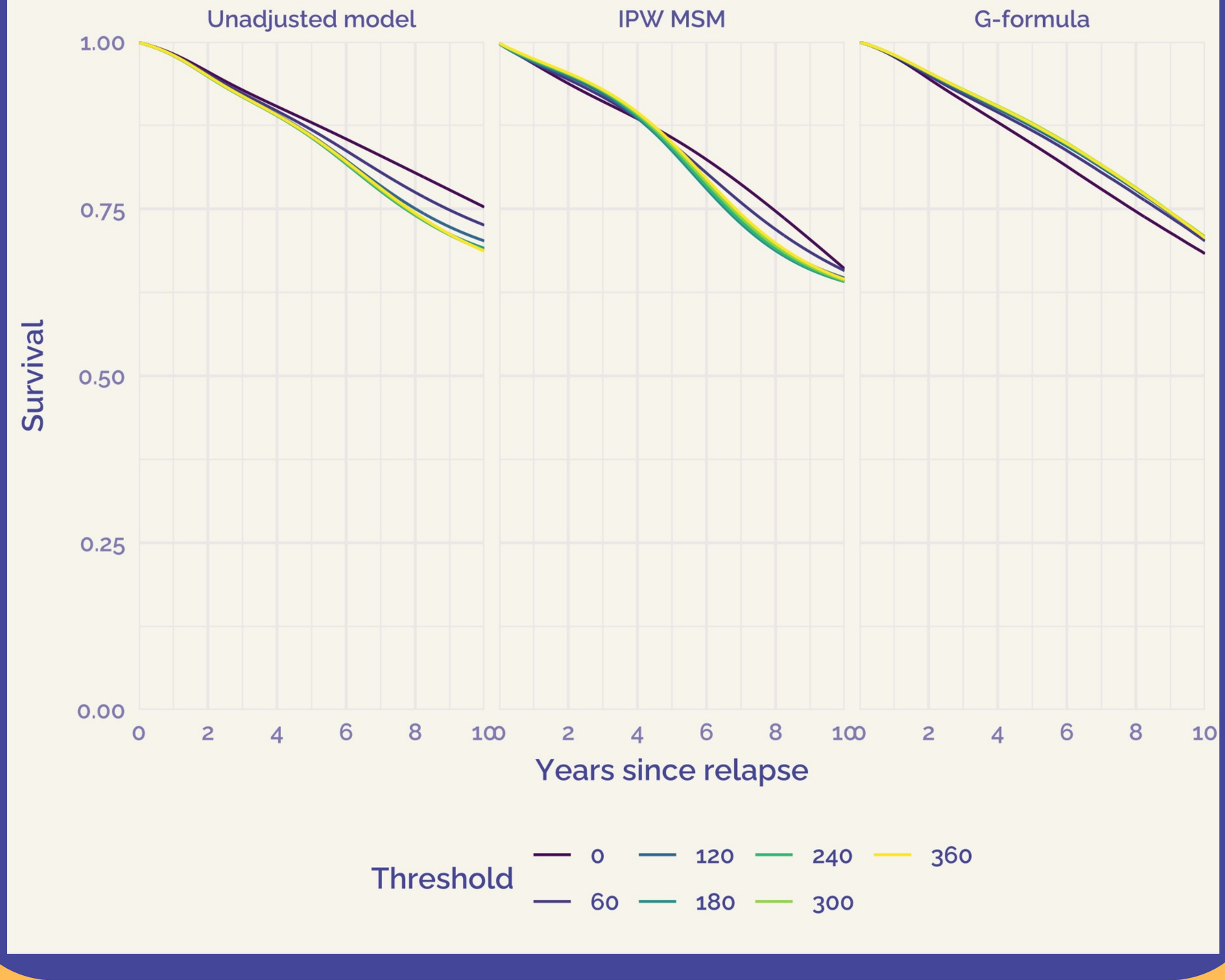
Aligning the protocol makes the methods comparable

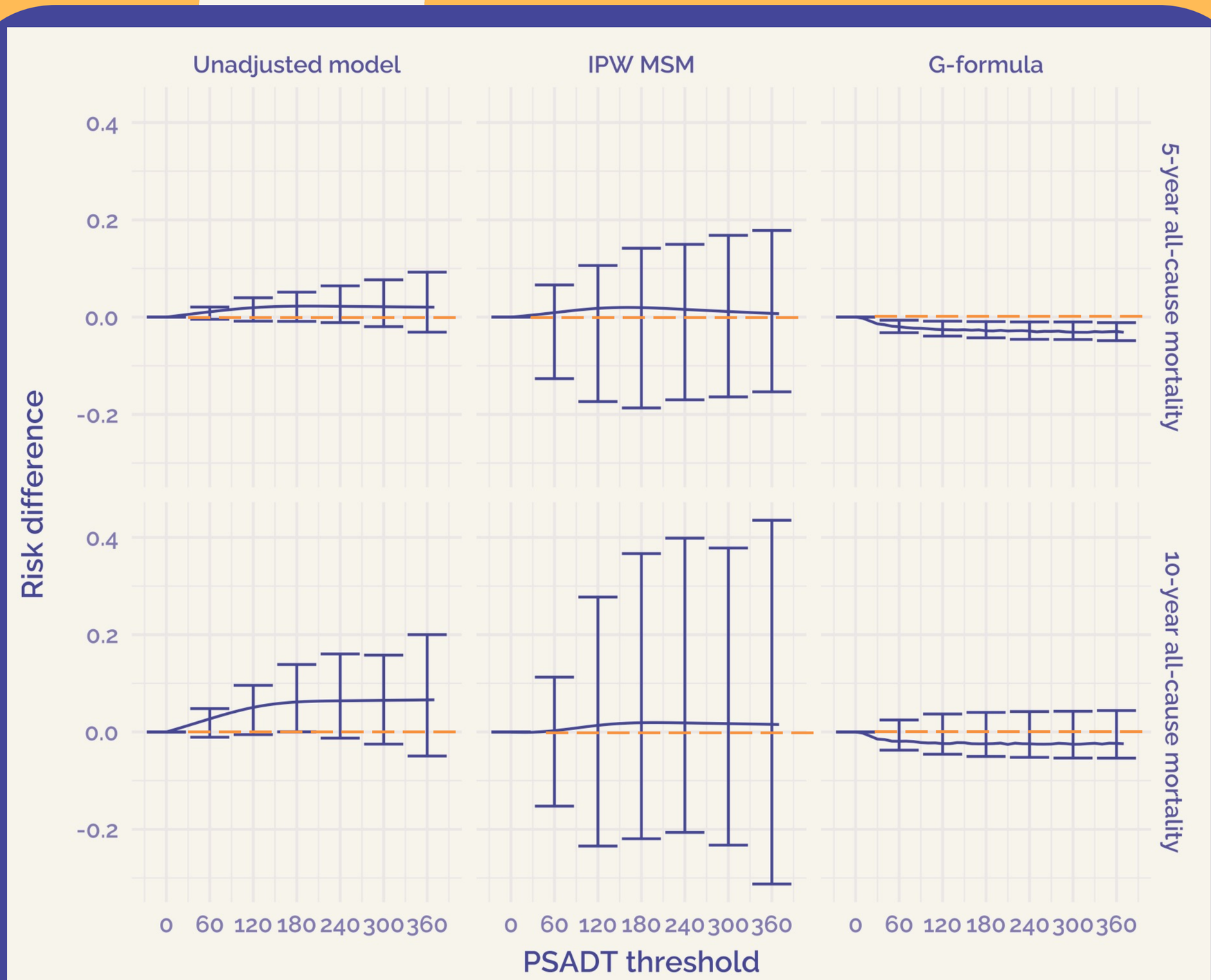
- The two estimators use parametric models for different components of the joint density of the observable data
- The estimands are the same because the treatment strategy we defined is the same
- If we left, e.g., duration of treatment unspecified:
  - IP weighting: Don't censor anyone after treatment 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)

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
- ~1200 eligible for our study

Survival curve





Risk differences

Data contributing to each strategy

Threshold	Person-months	Deaths
0	64247	145
60	57784	134
120	49088	132
180	41271	106
240	33786	85
300	28953	72
360	25205	64

Conclusions

More data needed

We can’t conclusively say whether a strategy based on PSADT would be useful.

Treatment definition

Fully defining a realistic treatment strategy for which there is support in the data is hard!

Thanks!

- Collaborators
  - Pregnancy and COVID-19: Camille Dollinger, Tyler VanderWeele, Diego Wyszynski, Sonia Hernández-Díaz
  - Prostate cancer: Xabier García-Albéniz, June Chan, Miguel Hernán

- Thanks to Mats for the invitation and EPFL for hosting me!

- Contact
  - louisa_h_smith@g.harvard.edu
  - @louisahsmith