Session 2.1

Target trials

Louisa Smith

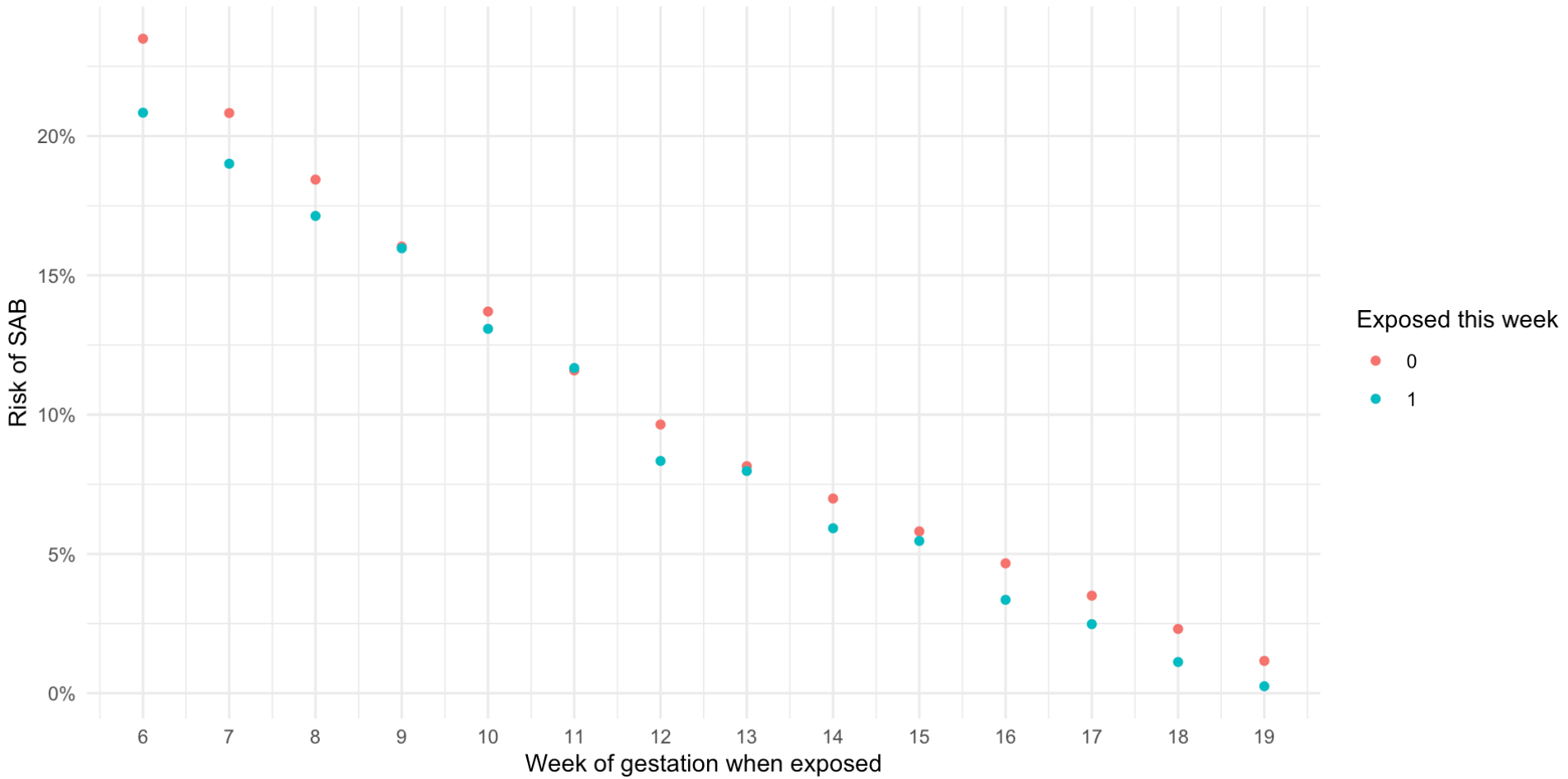
Recap

We saw that we could avoid bias from immortal time due to selection and to misclassification if we

- define people as unexposed or unexposed at a specific point in time
- restrict to those who we are defining as exposed or unexposed at that time

We will refer to this as “aligning time zero” and this is one of the main benefits of the target trial approach that we will discuss

But there was still a little bias left in our design (n = 100,000)



These risk ratios should be 1

week_comparison	exp_0	exp_1	risk_difference	r
6	0.235	0.208	-0.027	
7	0.208	0.190	-0.018	
8	0.184	0.171	-0.013	
9	0.160	0.160	-0.001	
10	0.137	0.131	-0.006	
11	0.116	0.117	0.001	
12	0.096	0.083	-0.013	
13	0.082	0.080	-0.002	

week_comparison	exp_0	exp_1	risk_difference	r
14	0.070	0.059	-0.011	
15	0.058	0.055	-0.003	
16	0.047	0.034	-0.013	
17	0.035	0.025	-0.010	
18	0.023	0.011	-0.012	
19	0.012	0.003	-0.009	

Why is there still bias?

We made our comparison groups on the basis of weeks.

But note that we defined:

```
dat <- dat |>
    mutate(...,
           week_exposed = floor(time_exposed),
           ...)
```

That is, we rounded down the time of exposure to the nearest week.

If we had rounded up, some people wouldn't be labeled "exposed" until after they

Why is there still bias

That means even though we are labeling someone exposed or unexposed at the beginning of the week, we are including some people who were exposed later in that week

Day-by-day comparisons?

Does this mean we need to make comparisons day-by-day? Hour-by-hour? Minute-by-minute?

- Not by hour or minute, luckily! We can only go as granular as our data – if people are only defined as exposed or unexposed, and only have an event on the day level, we can't (and don't need to) go finer than that
  - If we only had data at the week level, we would consider people exposed or unexposed for the whole week, and events wouldn't be recorded until the next week

We always have to think carefully about the order in which events are defined and

Day-by-day comparisons?

But we have daily data!

- If we wanted to ask a question about whether a *first-trimester* exposure causes an outcome, will we be unable to do this because we need to ask whether an exposure at xx day causes an outcome?

Luckily, no!

Grace periods

What we really need to do is acknowledge that some people are taking a few days to be exposed after the week starts

- Until then, we don't know if they will be exposed or unexposed that week
- They should actually be considered *both* exposed and unexposed until we know for sure

We can understand why we need this "grace period", and what to do about it, by thinking about how a randomized controlled trial works

Target trial emulation

A framework for designing observational studies that

- makes the counterfactual contrasts explicit
- aligns the “time zero” to avoid immortal time bias
- clarifies assumptions about target population

First we need to think about the design and specification of randomized controlled trials

Randomized controlled trials

NOT YET RECRUITING

NEW

NCT05944926

Improving Outcomes in Depression in Primary Care in a Low Resource Setting

CONDITIONS

Depression | Depressive Disorder

LOCATIONS

Location not provided

RECRUITING

NEW

NCT05944120

Decreasing On-Shift Stress With a Crisis Intervention Cart

CONDITIONS

Decreasing On-Shift Stress With a Crisis Intervention ...

LOCATIONS

Dallas, Texas, United States

NOT YET RECRUITING

NEW

NCT05944016

Phase 3 Clinical Trial With Dapagliflozin in Chronic Kidney Disease in Adolescents and Young Adult Patients

A randomized controlled trial for depression

The research question boils down to: Does an activity program or an antidepressant work better against depression?

Eligibility Criteria

DESCRIPTION

Inclusion Criteria:

- Participants will be adults aged 18 or over of any gender attending one of eight Primary Health Care Centers with a "diagnosis" of moderate to severe depression based on scores of 10 or above on the Patient Health Questionnaire-9 (PHQ-9).

Exclusion Criteria:

- Women who are pregnant or are breastfeeding or lactating
- Patients with a history of psychosis including schizophrenia spectrum disorders or bipolar disorder.
- Participants planning to move out of the study area during the follow-up period.
- Patients over 65 years of age with evidence of cognitive impairment - Patients who do not speak the study or local language (English or Hindi)
- Patients who are undergoing treatment for depression at the time of recruitment

AGES ELIGIBLE FOR STUDY

18 Years to 99 Years (Adult, Older Adult)

SEXES ELIGIBLE FOR STUDY

All

ACCEPTS HEALTHY VOLUNTEERS

No

ARMS AND INTERVENTIONS

Participant Group/Arm	Intervention/Treatment
Experimental: Healthy Activity Program (HAP) HAP is a brief psychological treatment adapted from behavioral activation therapy, an empirically supported psychological treatment recommended by WHO.	Behavioral: Healthy Activity Program (HAP) • HAP, delivered over 6-8 sessions by non-specialist healthcare workers, has behavioural activation as the core psychological strategy along with other strategies such as problem-solving and activation of social networks.
Experimental: Antidepressant medication (fluoxetine) Fluoxetine is a selective serotonin reuptake inhibitors (SSRIs) and one of the safest medications used to treat depression. It is a routinely used medication and part of the Essential Drug List (EDL) in India.	Drug: Antidepressant medication (fluoxetine) • Patients assigned to antidepressant medication will start on fluoxetine 20 mg/day and can be raised to 40 mg/day (the maximum mandated by treatment guidelines for primary care in India) at week 3 or 6 for patients who have yet to remit.

PRIMARY OUTCOME MEASURES

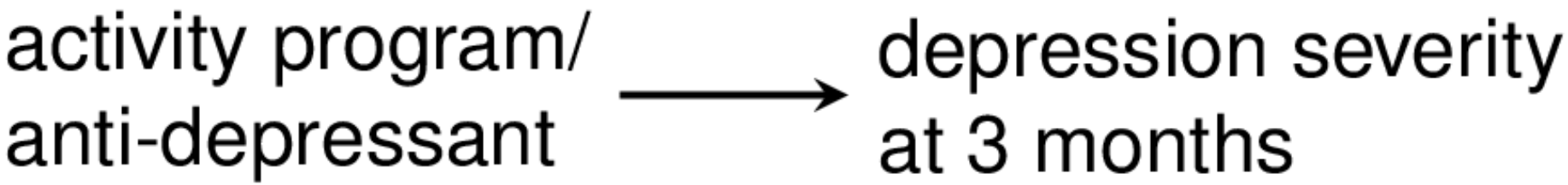
Outcome Measure	Measure Description	Time Frame
Depression severity, as measured by the Patient Health Questionnaire-9 (PHQ-9)	The PHQ-9 is a 9-item self-report scale to screen for symptoms of depression. Items are rated on a 4-point Likert scale from 0 (not at all) to 3 (nearly every day), with total scores ranging from 0 to 27, where higher scores indicate more severe depressive symptoms.	3 months post recruitment

SECONDARY OUTCOME MEASURES

Outcome Measure	Measure Description	Time Frame
Cost-effectiveness of optimization	Cost-effectiveness analysis by comparing costs and effectiveness between those who were randomly allocated to their optimal treatment vs. those who were randomly allocated to a non-optimal treatment. Effectiveness will be measures by (1) likelihood of remission and (2) Quality Adjusted Life Years (QALYs). Costs will be measured using the Client Service Receipt Inventory (CSRI) and system-level costs (see sections below for a description of these measures).	3-, 6-, 9-, 12-months post recruitment

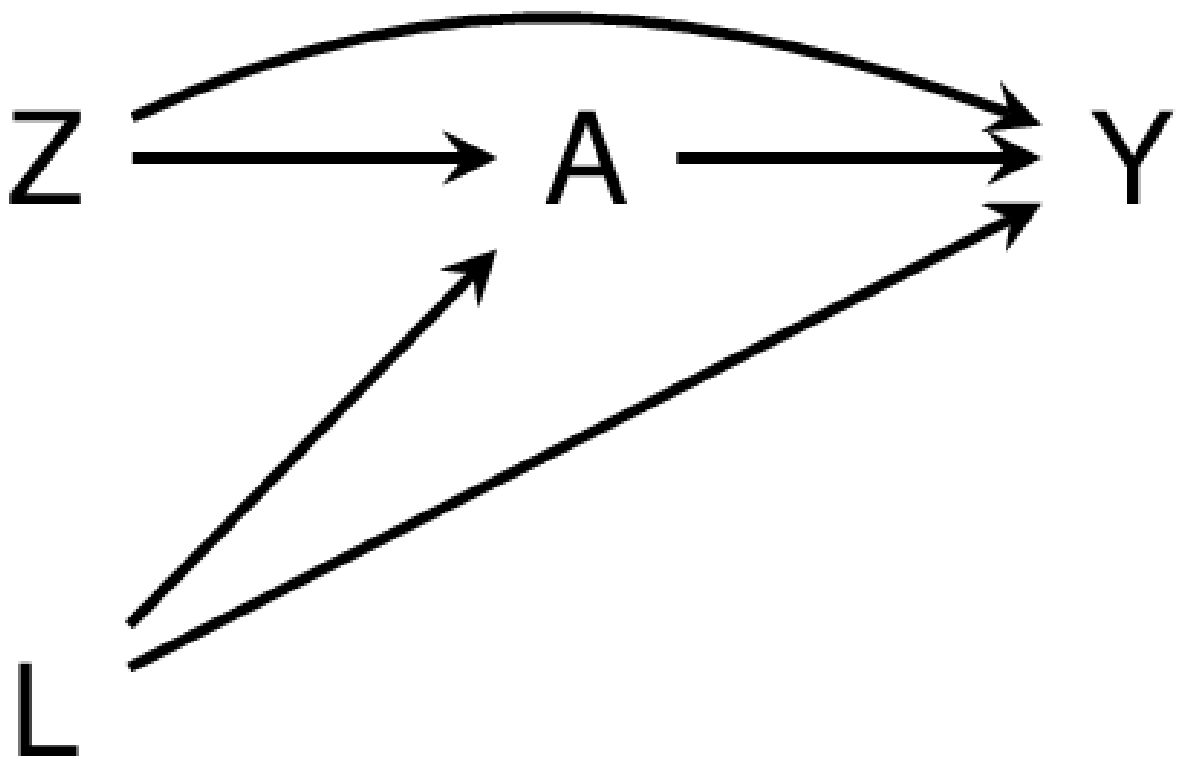
DAG for simple, perfect randomized trial

Depression treatment trial



- 20mg of anti-depressant to be taken daily can be raised to 40mg/day at week 3 or 6
- The activity program takes 8 weekly sessions
  - Do you think everyone will complete all 8 weeks of the activity program?

DAG for more realistic randomized trial



Intention-to-treat effect

$$\E[Y^{z = 1}] - \E[Y^{z = 0}]$$

What is the effect of being randomized to one treatment vs. another?

- When is this useful? When is it less useful?

Per-protocol effect

$$\E[Y^{z = 1, a = 1}] - \E[Y^{z = 0, a = 0}]$$

What is the effect of *actually taking* the treatment you were assigned to?

- When is this the same as the ITT effect?

Per-protocol effect for a sustained treatment

$$\E[Y^{z = 1, a_1 = 1, a_2 = 1, ..., a_8 = 1}] - \E[Y^{z = 0, a_1 = 0, a_2 = 0, ..., a_8 = 0}]$$

What is the effect of taking the assigned treatment for *all 8 weeks* of the intervention?

- When do we care about this in experimental studies? In observational studies?

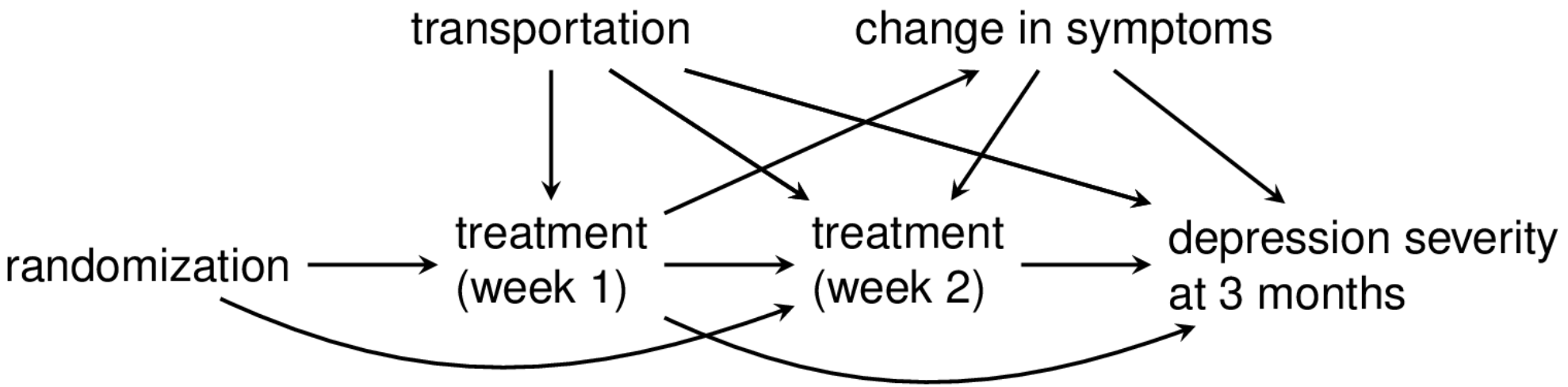
When do we have a time-varying (sustained) treatment (strategy/regimen)?

Easier question: when do we *not*?

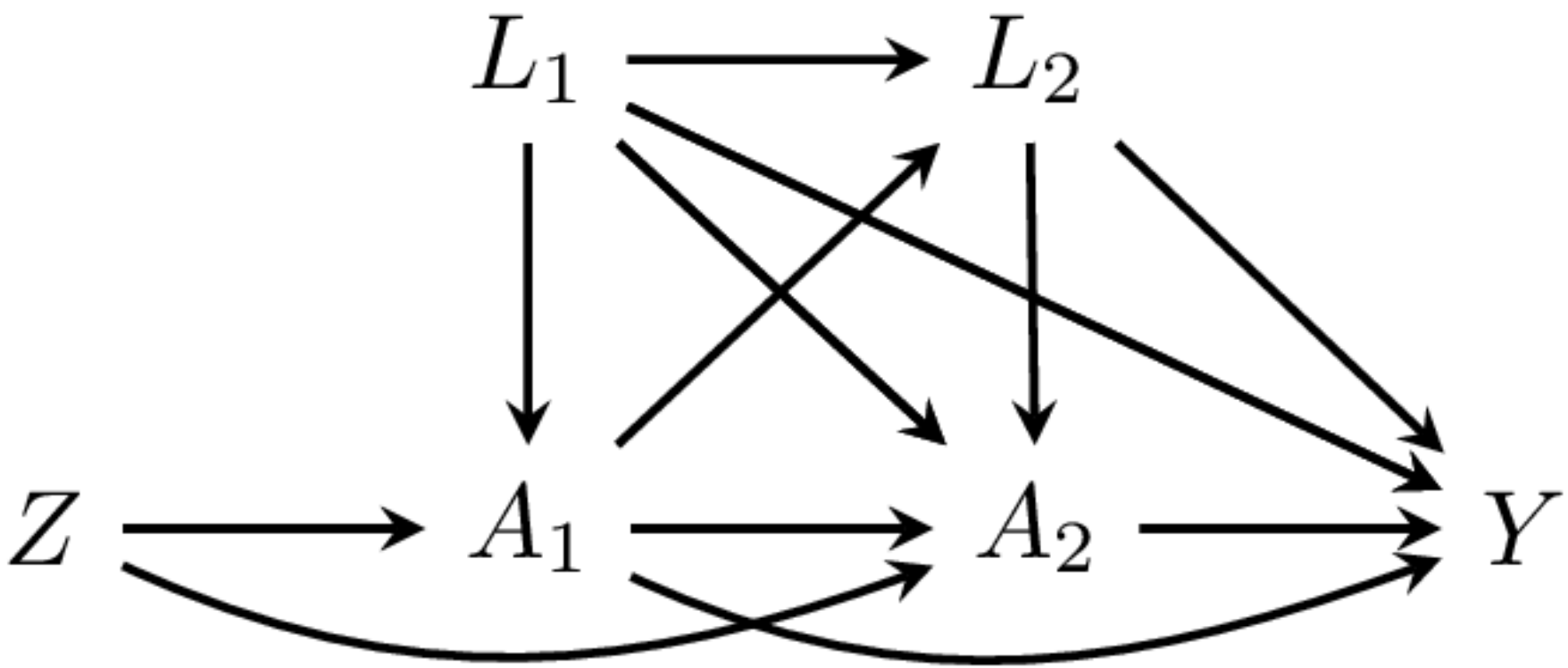
- One-time event
  - surgery, vaccine (sometimes), infection, genetics
- We really only care about initiation
  - prescribing a new drug
  - real-world effectiveness

Your causal questions: are they time-varying?

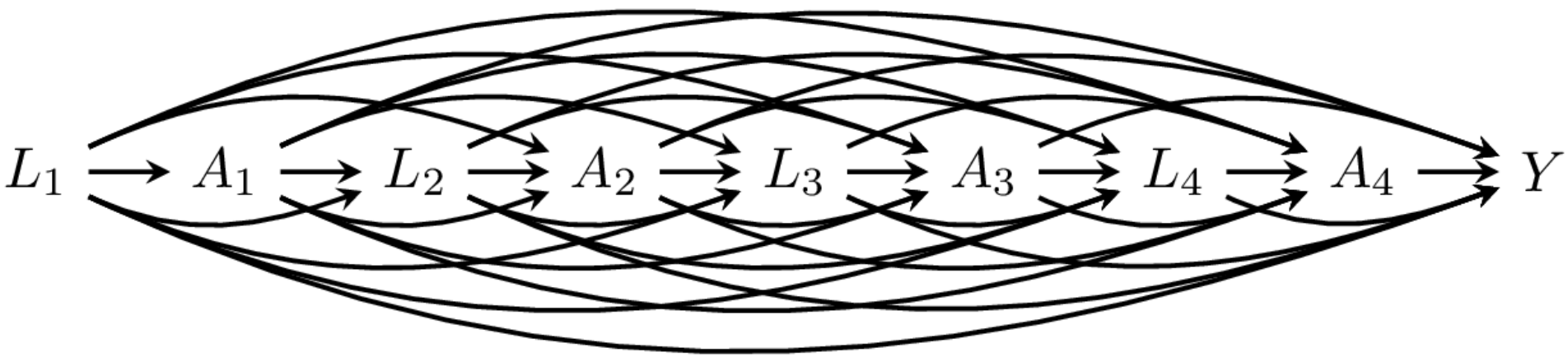
DAG for time-varying depression treatment (over just 2 weeks)



DAG for time-varying treatment, more generally



DAG for time-varying treatment, expanded…

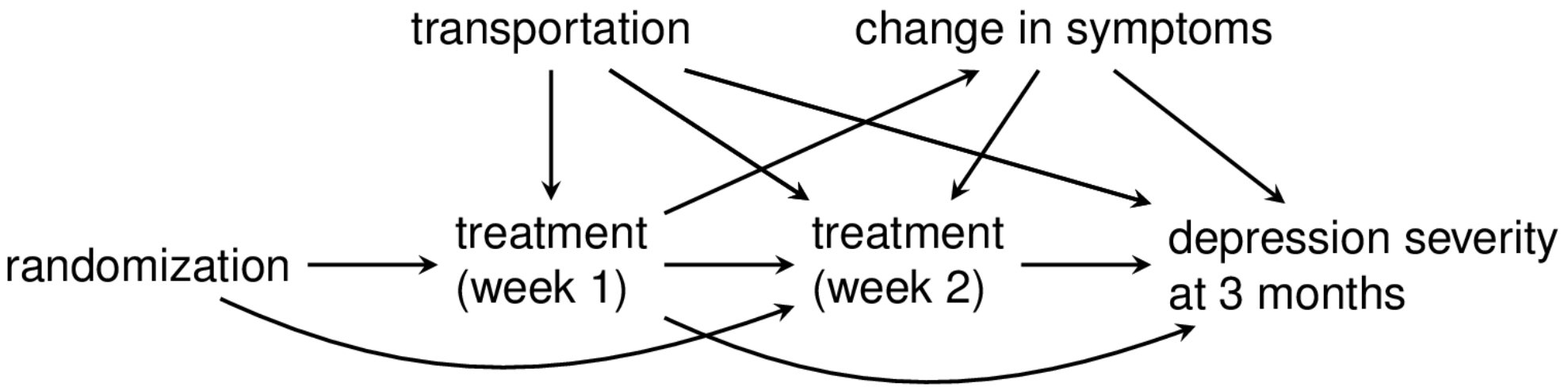


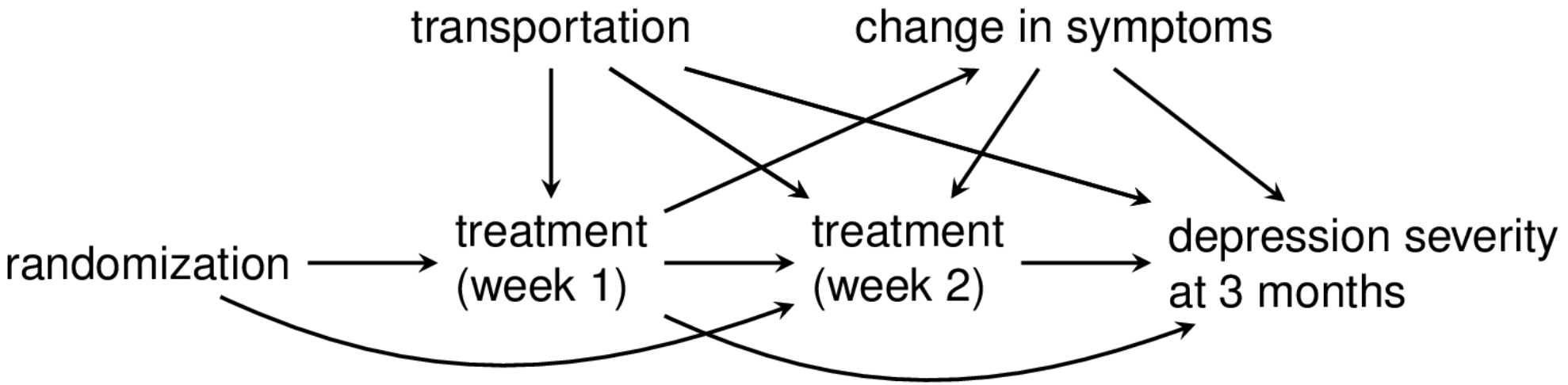
And this is with only one confounder, over 4 points in time…

How to make DAGs easier to work with when you have a time-varying treatment

- Two treatment nodes: initial treatment and continued treatment
  - Often you can assume the same pattern of confounding
- Group confounders into baseline confounders and post-baseline confounders
  - Post-baseline confounders are any that could plausibly change after/ be affected by the initiation of treatment
- For survival outcomes, you actually have multiple \(Y\)s as well…
  - You can often also think of this as just two
  - In order to continue treatment, you need to have survived
  - *(We will come back to this idea)*

In other words, this DAG is sufficient for our purposes!





To estimate the per-protocol effect of depression treatment (i.e., activity program for 8 weeks vs. anti-depressant):

- We need to adjust for transportation
- Do we need to adjust for change in symptoms?
  - What if there's a common cause of symptoms and depression severity (e.g., serotonin levels)?

Time-varying confounding

When variables:

- are affected by previous treatment
- confound future treatment

We somehow need to both *adjust* and ***not*** *adjust* for them

These are called time-varying confounders, and they require special methods (that we will get to!)

Wait, we're still in a randomized trial!

Randomized trials need causal inference methods for observational data in order to estimate anything but the intention-to-treat effect!

(Besides the fact that you can estimate an unbiased ITT effect assuming blinding, good randomization, etc.)

RCTs offer one huge benefit compared to typical observational studies:

- *They automatically align **time zero***

Consider an observational study

We use electronic health records to compare people with moderate depression:

- those who completed 8 weeks of a depression activity program
- those who were prescribed anti-depressants

and calculate risk of depression-related hospitalization (or some kind of severity symptom score if we have that data)

- We’re back to our old immortal time problem!

Solution: design your study like a randomized trial and start counting time when participants are “randomized”

- People are “randomized” as soon as they meet eligibility criteria
- In some designs, people can be “randomized” to “placebo” multiple times (in our previous week-specific comparisons, many people served in the comparison group many times)
- People might be randomized to a treatment but allowed a grace period to actually start it (in our week-specific comparison, they could be treated any time that week, though we haven’t accounted for that yet)

Steps

1. Specify the protocol of a target trial that answers your causal question – this doesn't need to be ethical or practically feasible, but you should not make compromises for the sake of your observational data
    - (You can base it on the observational data you know you have in some respects, e.g., your population of interest)
2. Emulate the target trial using your observational data
    - This may require compromises, but you should be explicit about them
    - If the compromises are too severe, you may need to reconsider your causal question/redefine your target trial

Components of a target trial

- Eligibility criteria
- Treatment strategies
- Assignment procedures
- Follow-up
- Outcome(s)
- Causal contrast(s)
- Assumptions for emulation
- Statistical analysis plan

Recently published reporting guidelines

Cashin et al. (2025)

Item 3: Summarize the causal question.

Item 6: Specify the components of the target trial protocol that would answer the causal question.

Item 7: Describe how the components of the target trial protocol were emulated with the observational data, including how all variables were measured or ascertained.

Component	Target Trial Specification (Item 6)	Target Trial Emulation (Item 7)
Eligibility Criteria	Describe the eligibility criteria.	Describe how the eligibility criteria were operationalized with the data.
Treatment Strategies	Describe the treatment strategies that would be compared.	Describe how the treatment strategies were operationalized with the data.

Component	Target Trial Specification (Item 6)	Target Trial Emulation (Item 7)
Assignment Procedures	Report that eligible individuals would be randomly assigned to treatment strategies and may be aware of their treatment allocation.	Describe how assignment to treatment strategies was operationalized with the data.
Follow-up	Clarify that follow-up would start at time of assignment to the treatment strategies. Specify when follow-up would end.	Clarify that follow-up starts at the time individuals were assigned to the treatment strategies. Describe how the end of follow-up was operationalized with the data.
Outcomes	Describe the outcomes.	Describe how the outcomes were operationalized with the data.
Causal Contrasts	Describe the causal contrasts of interest, including effect measures.	Describe how the causal contrasts were operationalized with the data, including effect measures.

Pharmaceutical example from Hernán and Robins (2016)

Component	Target Trial Specification
Causal question	What is the effect of postmenopausal hormone therapy on breast cancer.
Eligibility criteria	Postmenopausal women within 5 years of menopause and with no history of cancer and no use of hormone therapy in the past 2 years
Treatment strategies	1. Refrain from taking hormone therapy during follow-up 2. Initiate estrogen plus progestin hormone therapy at baseline and remain on it during the follow-up, unless diagnosed with deep vein thrombosis, etc.
Assignment procedures	Participants will be randomly assigned to either strategy at baseline and will be aware of the strategy to which they have been assigned
Follow-up	Patients are followed from enrollment (time zero) until breast cancer diagnosis, loss to follow-up, or administrative end of

	follow-up (5 years from baseline)
Outcome	Breast cancer diagnosed by an oncologist
Contrast	Intention-to-treat effect, per-protocol effect

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

Components	Target trial
Causal question	What is the effect of COVID-19 infection 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.

Components	Target trial
Outcome	Preterm delivery, defined as delivery before 37 completed weeks of gestation.
Causal contrast	Intention-to-treat effect on the risk ratio and risk difference scales.

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	• Based only on **pre-baseline** characteristics • Generally requires pre-baseline observation window
Treatment strategies	• Can't assign actual placebo or blinding (can assign no treatment if realistic) • Some people must have "adhered" to the treatment strategy
Assignment procedures	• Randomization (within levels of confounders) is always an assumption • Assignment "happens" as soon as someone meets eligibility criteria

Components	Target trial
Follow-up	• Monitoring for the outcome throughout follow-up (e.g., regular mammograms) may need to be part of the treatment strategy
Outcome	• Outcome ascertainment can't be blinded
Causal contrast	• Intention-to-treat effect makes sense when "most" of the treatment happens immediately upon randomization • Per-protocol useful when you don't know right away who starts what treatment

Choose one

Chiu et al. (2024)

Yland et al. (2022)

Caniglia et al. (2023)

Caniglia et al. (2018)

Wong et al. (2024)

Grandi et al. (2024)

Or any other of your choice!

Your target trial

Components	Target trial
Eligibility criteria	
Treatment strategies	
Assignment procedures	
Follow-up	
Outcome	
Causal contrast	

References

Caniglia, Ellen C., Rebecca Zash, Christina Fennell, Modiegi Diseko, Gloria Mayondi, Jonathan Heintz, Mompati Mmalane, et al. 2023. “Emulating Target Trials to Avoid Immortal Time Bias – an Application to Antibiotic Initiation and Preterm Delivery.” *Epidemiology* 34 (3): 430. https://doi.org/10.1097/EDE.0000000000001601.

Caniglia, Ellen C., Rebecca Zash, Denise L. Jacobson, Modiegi Diseko, Gloria Mayondi, Shahin Lockman, Jennifer Y. Chen, et al. 2018. “Emulating a Target Trial of Antiretroviral Therapy Regimens Started Before Conception and Risk of Adverse Birth Outcomes.” *AIDS* 32 (1): 113–20. https://doi.org/10.1097/qad.0000000000001673.

Cashin, Aidan G., Harrison J. Hansford, Miguel A. Hernán, Sonja A. Swanson, Hopin Lee, Matthew D. Jones, Issa J. Dahabreh, 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, Krista F. Huybrechts, Elisabetta Patorno, Jennifer J. Yland, Carolyn E. Cesta, Brian T. Bateman, Ellen W. Seely, Miguel A. Hernán, and Sonia

Hernández-Díaz. 2024. “Metformin Use in the First Trimester of Pregnancy and Risk for Nonlive Birth and Congenital Malformations: Emulating a Target Trial Using Real-World Data.” *Annals of Internal Medicine* 177 (7): 862–70. https://doi.org/10.7326/M23-2038.

Grandi, Sonia M., Ya-Hui Yu, Pauline Reynier, Robert W. Platt, Oriana H. Y. Yu, and Kristian B. Filion. 2024. “Levothyroxine Initiation and the Risk of Pregnancy Loss Among Pregnant Women with Subclinical Hypothyroidism: An Observational Study Emulating a Target Trial.” *Paediatric and Perinatal Epidemiology* 38 (6): 470–81. https://doi.org/10.1111/ppe.13015.

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.

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.

Wong, Carlos K. H., Kristy T. K. Lau, Matthew S. H. Chung, Ivan C. H. Au, Ka Wang Cheung, Eric H. Y. Lau, Yasmin Daoud, Benjamin J. Cowling, and Gabriel M. Leung. 2024. “Nirmatrelvir/Ritonavir Use in Pregnant Women with SARS-CoV-2 Omicron Infection: A Target Trial Emulation.” *Nature Medicine*