FACULTY OF SCIENCES

CAUSAL THINKING IN CLINICAL TRIALS, JULY 2024, THESSALONIKI

TIME-VARYING CONFOUNDING

Alex Ocampo, Novartis Pharma AG, Switzerland

GHENT UNIVERSITY

PROBLEMS OF STANDARD ADJUSTMENT

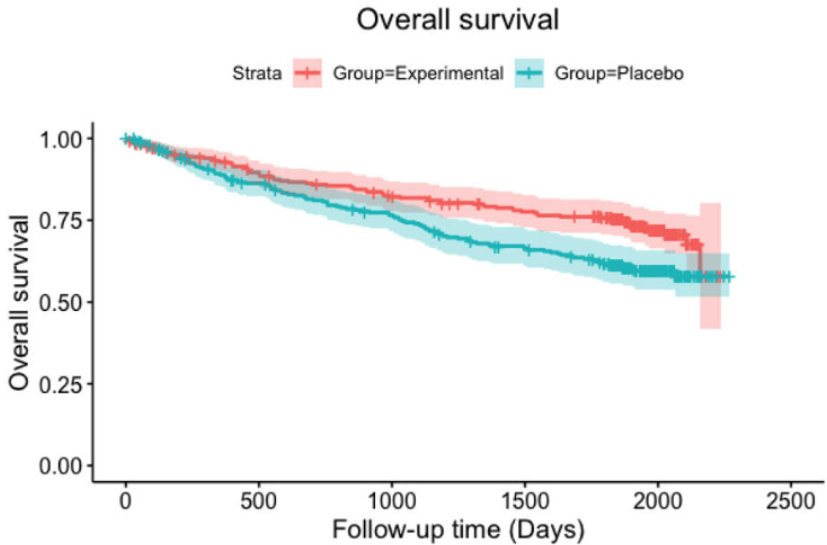
GHENT
UNIVERSITY

THE HELIOS TRIAL (NCT01611090)

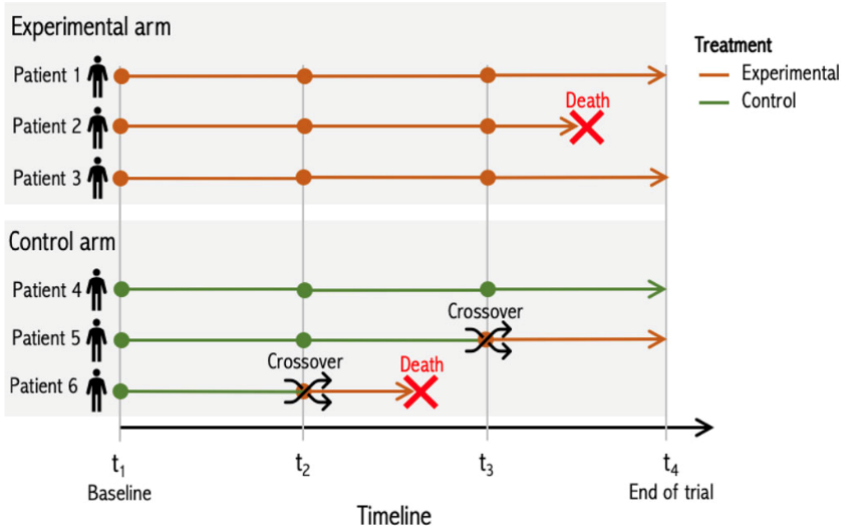
(FRASER ET AL. 2020)

- 578 patients with relapsed/refractory chronic lymphocytic leukemia/small lymphocytic lymphoma without deletion 17p.
- Randomized 1:1 to 420mg daily ibrutinib or placebo
  plus 6 cycles of bendamustine plus rituximab (BR),
  followed by ibrutinib or placebo alone.
- Control patients were allowed to cross over to ibrutinib upon disease progression.
- Median follow-up time 63.7 months.
- 5-year ITT HR 0.61 (95% CI 0.46 to 0.82) for ibrutinib plus BR versus placebo plus BR, despite crossover in 63.3% of control patients.

THE HELIOS TRIAL (NCT01611090)



CROSS-OVER


Experimental arm
Patient 1
Patient 2
Patient 3
Death
Control arm
Patient 4
Patient 5
Patient 6
Crossover
Crossover
Death
Treatment
Experimental
Control
t_1
Baseline
t_2
t_3
t_4
End of trial
Timeline

TRADITIONAL CORRECTIONS FOR CROSS-OVER

- Cross-over dilutes the treatment policy estimand.
- The relative risk of surviving the time of unblinding ($\approx$ 3 years) is 1.08 (95% CI 0.98 to 1.18).

TRADITIONAL CORRECTIONS FOR CROSS-OVER

- Cross-over dilutes the treatment policy estimand.
- The relative risk of surviving the time of unblinding ($\approx$ 3 years) is 1.08 (95% CI 0.98 to 1.18).
- The per-protocol effect (excluding patients who cross over) is 1.87 (95% CI 1.45 to 2.40).

TRADITIONAL CORRECTIONS FOR CROSS-OVER

- Cross-over dilutes the treatment policy estimand.
- The relative risk of surviving the time of unblinding ($\approx$ 3 years) is 1.08 (95% CI 0.98 to 1.18).
- The per-protocol effect (excluding patients who cross over) is 1.87 (95% CI 1.45 to 2.40).
- The as-treated effect (treatment as time-varying covariate) is 1.05 (95% CI 0.94 to 1.17).
- Major concerns about these analyses.

Acknowledge the time of cross-over!

- It is essential to acknowledge the time of cross-over.
- The per-protocol analysis does not do so,
  which makes it vulnerable to immortal time-bias.
  - 'You need to live long enough to cross'.
  - Reverse causality,
    which may make the control treatment look especially bad.

Acknowledge the time of cross-over!

- It is essential to acknowledge the time of cross-over.
- The per-protocol analysis does not do so,
  which makes it vulnerable to immortal time-bias.
  - 'You need to live long enough to cross'.
  - Reverse causality,
    which may make the control treatment look especially bad.
  - Per-protocol analysis also 'over-represents' early deaths.
- Patient exclusions must therefore never be based on intercurrent events.

PRINCIPAL STRATIFICATION

- Principal stratification naturally resolves this problem
  by turning the exclusion criterion into one that is balanced across arms.
- E.g., to measure the effect of treatment on surviving time t
  one may focus on the principal stratum of patients who would not have crossed by time t,
  no matter what treatment assignment.

PRINCIPAL STRATIFICATION

- Principal stratification naturally resolves this problem
  by turning the exclusion criterion into one that is balanced across arms.
- E.g., to measure the effect of treatment on surviving time t
  one may focus on the principal stratum of patients who would not have crossed by time t,
  no matter what treatment assignment.
- This technically resolves the concern, but not practically:
  - the stratum likely contains many patients who die early;
  - *how meaningful is it to learn the treatment effect for those?*
- The stratum is also different at each time t,
  making survival curves difficult to interpret.

TIME-VARYING CONFOUNDING (1)

- When the exposure (crossing-over) varies over time, the confounders are time-varying as well.
  - The decision to cross is based on disease progression.
  - Baseline confounding adjustment can never suffice!

    (shared parameter models can also not work for that reason)

TIME-VARYING CONFOUNDING (2)

- It is not uncommon to aggregate exposure and confounders over time.
  - E.g., Cox regression on treatment, duration of cross-over and their interaction, adjusting for measures of disease progression before cross-over.

TIME-VARYING CONFOUNDING (2)

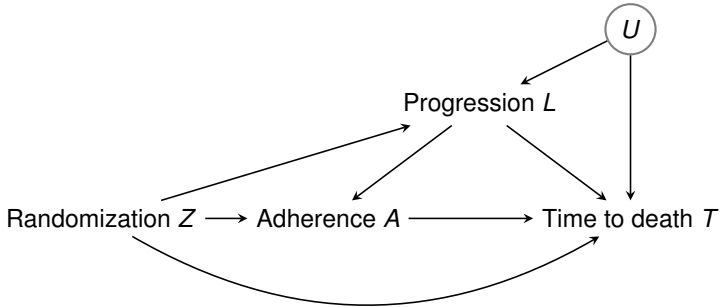
- It is not uncommon to aggregate exposure and confounders over time.
  - E.g., Cox regression on treatment, duration of cross-over and their interaction, adjusting for measures of disease progression before cross-over.
- Regression with time-varying covariates (cross-over status and disease progression) is likewise not uncommon.

TIME-VARYING CONFOUNDING (2)

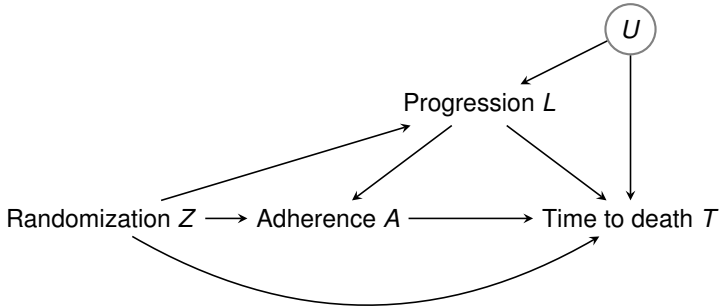
- It is not uncommon to aggregate exposure and confounders over time.
  - E.g., Cox regression on treatment, duration of cross-over and their interaction, adjusting for measures of disease progression before cross-over.
- Regression with time-varying covariates (cross-over status and disease progression) is likewise not uncommon.
- Both strategies do not carefully distinguish what comes before/after exposure (cross-over).

(the first strategy also ignores that the duration of cross-over is influenced by the endpoint)

WHY DOES STANDARD REGRESSION ADJUSTMENT FAIL?

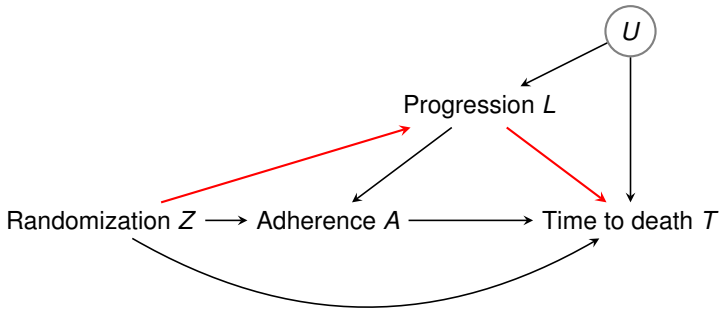


WHY DOES STANDARD REGRESSION ADJUSTMENT FAIL?



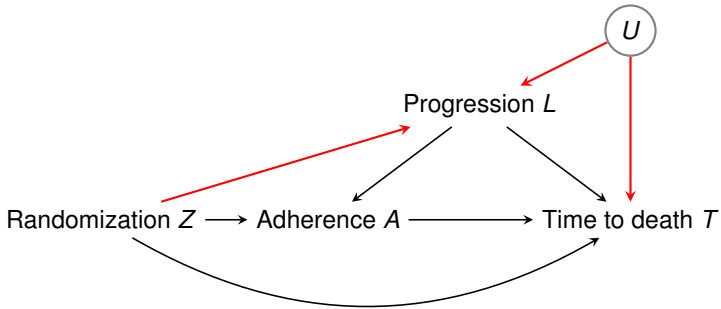
- Adjustment for progression is needed as it confounds
  the association between adherence A (i.e., cross-over) and time to death T.

PROBLEM 1: ELIMINATING INDIRECT EFFECTS

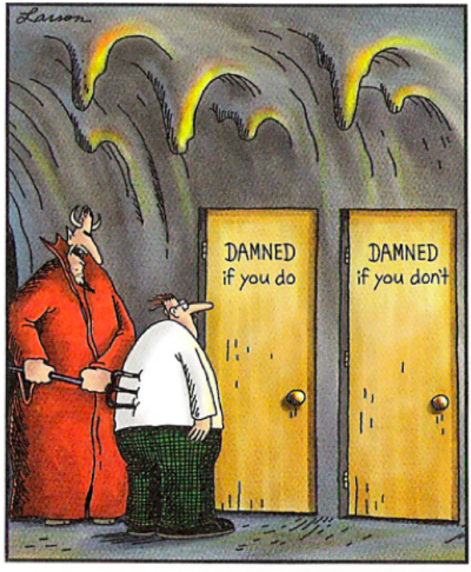


- Adjustment for progression eliminates part of the effect of randomization Z on time to death T.

PROBLEM 2: COLLIDER-STRATIFICATION BIAS



- Adjustment for progression induces selection bias.



"C'mon, c'mon—it's either one or the other."

THE PROBLEM OF TIME-VARYING CONFOUNDING

- Standard methods
  for estimating the effects of time-varying treatments
  are biased when confounders predict subsequent treatments.
- In such cases, they do not distinguish whether
  progression happened before or after cross-over.
- This is common in observational studies, e.g.
  - HAART treatment is re-assessed at regular visits on the basis of CD4 count,
    but also affects future CD4 counts.

THE PROBLEM OF TIME-VARYING CONFOUNDING

- Standard methods
  for estimating the effects of time-varying treatments
  are biased when confounders predict subsequent treatments.
- In such cases, they do not distinguish whether
  progression happened before or after cross-over.
- This is common in observational studies, e.g.
  - HAART treatment is re-assessed at regular visits on the basis of CD4 count,
    but also affects future CD4 counts.
  - Patients with high disease severity are more susceptible to infections,
    which in turn affect disease severity.

G-METHODS

How to adjust for time-varying confounding?

- James Robins diagnosed these problems in the ‘80’s.

How to adjust for time-varying confounding?

- James Robins diagnosed these problems in the ‘80’s.
- He devised 3 strategies, collectively called g-methods, to overcome them:
  - g-computation

(Robins, 1986; Taubman et al., 2009)

How to adjust for time-varying confounding?

- James Robins diagnosed these problems in the ‘80’s.
- He devised 3 strategies, collectively called g-methods, to overcome them:
  - g-computation

    (Robins, 1986; Taubman et al., 2009)
  - g-estimation for Structural Nested Models

    (Robins and Tsiatis, 1991; Robins, 1994; Vansteelandt and Joffe, 2014)

How to adjust for time-varying confounding?

- James Robins diagnosed these problems in the ‘80’s.
- He devised 3 strategies, collectively called g-methods, to overcome them:
  - g-computation

    (Robins, 1986; Taubman et al., 2009)
  - g-estimation for Structural Nested Models

    (Robins and Tsiatis, 1991; Robins, 1994; Vansteelandt and Joffe, 2014)
  - Inverse probability weighting (for Marginal Structural Models)

    (Robins, Hernan and Brumback, 2000; Hernan, Robins and Brumback, 2001)

How to adjust for time-varying confounding?

- James Robins diagnosed these problems in the ‘80’s.
- He devised 3 strategies, collectively called g-methods, to overcome them:
  - g-computation

    (Robins, 1986; Taubman et al., 2009)
  - g-estimation for Structural Nested Models

    (Robins and Tsiatis, 1991; Robins, 1994; Vansteelandt and Joffe, 2014)
  - Inverse probability weighting (for Marginal Structural Models)

    (Robins, Hernan and Brumback, 2000; Hernan, Robins and Brumback, 2001)
- These strategies can make a big difference...

Effect of combined ART on mortality?

- $\approx$ 60000 individuals in Europe and US.

  (Ray et al., 2010)

- Individuals start cART at different times.
- Time of initiation is influenced by CD4 count, viral load, ...

Effect of combined ART on mortality?

- $\approx$ 60000 individuals in Europe and US.

  (Ray et al., 2010)

- Individuals start cART at different times.
- Time of initiation is influenced by CD4 count, viral load, ...
- Unadjusted: HR 1.39 (95% CI 1.25 to 1.53)
- Baseline adjusted: HR 0.79 (95% CI 0.70 to 0.88)
- Standard adjustment for time-varying confounding: HR 0.83 (95% CI 0.74 to 0.94)

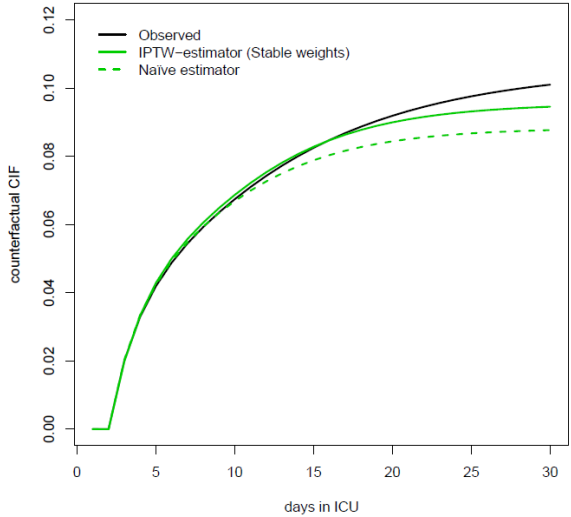
EFFECT OF COMBINED ART ON MORTALITY?

- $\approx$ 60000 individuals in Europe and US.

  (Ray et al., 2010)
- Individuals start cART at different times.
- Time of initiation is influenced by CD4 count, viral load, ...
- Unadjusted: HR 1.39 (95% CI 1.25 to 1.53)
- Baseline adjusted: HR 0.79 (95% CI 0.70 to 0.88)
- Standard adjustment for time-varying confounding: HR 0.83 (95% CI 0.74 to 0.94)
- IPW adjusted: HR 0.48 (95% CI 0.41 to 0.57),
  in line with RCT results.

EFFECT OF HOSPITAL-ACQUIRED INFECTION ON MORTALITY

(Bekaert et al., 2009)



THE CAUSAL ROADMAP

GHENT
UNIVERSITY

ROADMAP STEP 1: HYPOTHETICAL ESTIMAND

- Robins' g-methods follow the causal roadmap: their starting point is an estimand.
- *What would the treatment effect be in the absence of cross-over?* (hypothetical strategy)

(likewise, in the absence of treatment discontinuation or use of rescue medication)

ROADMAP STEP 1: HYPOTHETICAL ESTIMAND

- Robins' g-methods follow the causal roadmap: their starting point is an estimand.
- *What would the treatment effect be in the absence of cross-over?* (hypothetical strategy)

  (likewise, in the absence of treatment discontinuation or use of rescue medication)

- Let $T^{\bar{a}_\tau}$ denote the time to death
  with treatment status $\bar{a}_\tau = (a_0, a_1, ..., a_\tau)$ until the end of study time τ.
  - $\bar{a}_\tau = (0, ..., 0)$: assignment to control ($a_0 = z = 0$) and no cross-over.
  - $\bar{a}_\tau = (0, 1, ..., 1)$: assignment to control and cross-over at first visit.
  - $\bar{a}_\tau = (1, 1, ..., 1)$: assignment to treatment.
- Then we may contrast

$$P(T^{1...1} > t) \quad \text{versus} \quad P(T^{0...0} > t).$$

- *Is this a relevant estimand?*

Cox regression with time-varying covariates

- The first step in a traditional analysis is very different:
  postulation of a Cox model with time-varying covariates:

$$\lambda(t|A_0, A_t, L_t) = \lambda_0(t) \exp(\beta A_0 + \gamma A_t + \lambda L_t)$$

  for all $t > 0$.

Cox regression with time-varying covariates

- The first step in a traditional analysis is very different:
  postulation of a Cox model with time-varying covariates:

$$\lambda(t|A_0, A_t, L_t) = \lambda_0(t) \exp(\beta A_0 + \gamma A_t + \lambda L_t)$$

  for all $t > 0$.

- It is not so clear how to interpret β
  because the model adjusts for different variables at each time
  (and considering non-collapsibility).

COX REGRESSION WITH TIME-VARYING COVARIATES

- The first step in a traditional analysis is very different:
  postulation of a Cox model with time-varying covariates:

$$\lambda(t|A_0, A_t, L_t) = \lambda_0(t) \exp(\beta A_0 + \gamma A_t + \lambda L_t)$$

  for all $t > 0$.
- It is not so clear how to interpret β
  because the model adjusts for different variables at each time
  (and considering non-collapsibility).
- The model also cannot be used to predict survival in the absence of cross-over.

ROADMAP STEP 2: ASSUMPTIONS

- At each visit time t, we will assume that

$$A_t \perp\!\!\!\perp T^{\bar{a}_\tau} | \overline{L}_t, \overline{A}_{t-1} = 0$$

  where $\overline{A}_{t-1} = (A_0, ..., A_{t-1})$ and $\overline{L}_t = (L_0, ..., L_t)$.

- This assumption is known as sequential randomization or no unmeasured confounding.

  (Robins, 1986; Hernán and Robins, 2020)

- *What does it mean?*

ROADMAP STEP 2: ASSUMPTIONS

- At each visit time t, we will assume that

$$A_t \perp\!\!\!\perp T^{\bar{a}_T} | \overline{L}_t, \overline{A}_{t-1} = 0$$

  where $\overline{A}_{t-1} = (A_0, ..., A_{t-1})$ and $\overline{L}_t = (L_0, ..., L_t)$.

- This assumption is known as sequential randomization or no unmeasured confounding.

  (Robins, 1986; Hernán and Robins, 2020)

- *What does it mean?*

- At each visit time t, the decision to cross a control patient (who was not previously crossed) may depend on the history of confounders $\overline{L}_t$,
  but has no residual association with the endpoint **if the same decision were made**.

ROADMAP STEP 2: ASSUMPTIONS

- At each visit time t, we will assume that

$$A_t \perp\!\!\!\perp T^{\bar{a}_\tau} | \overline{L}_t, \overline{A}_{t-1} = 0$$

  where $\overline{A}_{t-1} = (A_0, ..., A_{t-1})$ and $\overline{L}_t = (L_0, ..., L_t)$.

- This assumption is known as sequential randomization or no unmeasured confounding.

  (Robins, 1986; Hernán and Robins, 2020)

- *What does it mean?*
- At each visit time t, the decision to cross a control patient (who was not previously crossed) may depend on the history of confounders $\overline{L}_t$,
  but has no residual association with the endpoint **if the same decision were made**.
- Or, consider two control patients with the same history $\overline{L}_t$,
  and suppose that one crosses at time t and the other does not.
  - Then these patients are exchangeable.

ROADMAP STEP 3: ESTIMATION (1)

- g-computation imputes the counterfactual data that would be seen 'in the hypothetical world'.
- From the time of cross-over onwards until the end-of-study time, it thus repeatedly imputes for control patients
  - whether they will survive the next visit time;
    - if they do, their time-varying covariates at the next visit time,
    - if they don't, their actual event time.

  if no cross-over occurs.

ROADMAP STEP 3: ESTIMATION (1)

- g-computation imputes the counterfactual data that would be seen 'in the hypothetical world'.
- From the time of cross-over onwards until the end-of-study time,
  it thus repeatedly imputes for control patients
  - whether they will survive the next visit time;
    - if they do, their time-varying covariates at the next visit time,
    - if they don't, their actual event time.

  if no cross-over occurs.
- For this, it uses parametric imputation models,
  fitted to patients who have not crossed by the considered time,
  which is demanding when there are many time-varying confounders.

ROADMAP STEP 3: ESTIMATION (1)

- g-computation imputes the counterfactual data that would be seen 'in the hypothetical world'.
- From the time of cross-over onwards until the end-of-study time,
  it thus repeatedly imputes for control patients
  - whether they will survive the next visit time;
    - if they do, their time-varying covariates at the next visit time,
    - if they don't, their actual event time.

  if no cross-over occurs.
- For this, it uses parametric imputation models,
  fitted to patients who have not crossed by the considered time,
  which is demanding when there are many time-varying confounders.
- Be cautious: some papers ignore that also time-varying covariates must be imputed!

  (Qu et al., 2020)

ROADMAP STEP 3: ESTIMATION (2)

- Inverse probability weighting (IPW) is a general, more user-friendly method to adjust for confounding and selection bias based on propensity scores.

ROADMAP STEP 3: ESTIMATION (2)

- Inverse probability weighting (IPW) is a general, more user-friendly method to adjust for confounding and selection bias based on propensity scores.
- The propensity score is the chance of receiving treatment ('crossing over'), based on subject characteristics,
  and plays an important role in confounding adjustment.

  (Rosenbaum and Rubin, 1984)
- Confounding adjustment based on propensity scores
  is less prone to extrapolation (by making it very explicit).

  (Rubin, 1997)

ROADMAP STEP 3: ESTIMATION (2)

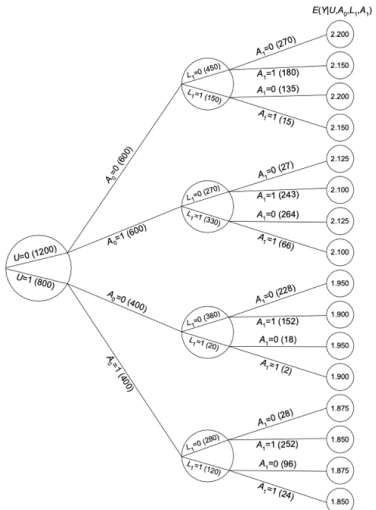
- Inverse probability weighting (IPW) is a general, more user-friendly method to adjust for confounding and selection bias based on propensity scores.
- The propensity score is the chance of receiving treatment ('crossing over'), based on subject characteristics,
  and plays an important role in confounding adjustment.

  (Rosenbaum and Rubin, 1984)
- Confounding adjustment based on propensity scores
  is less prone to extrapolation (by making it very explicit).

  (Rubin, 1997)
- The real benefit here is that they can 'easily' overcome
  the problems of regression adjustment.
- I will illustrate this using a toy example due to Rhian Daniel.

TOY EXAMPLE WITH CONTINUOUS ENDPOINT

(L_0: EMPTY, L_1: 0 MEANS PROGRESSION, A_1: 1 MEANS CROSS-OVER)



$$Pr(U = 1) = 0.4$$

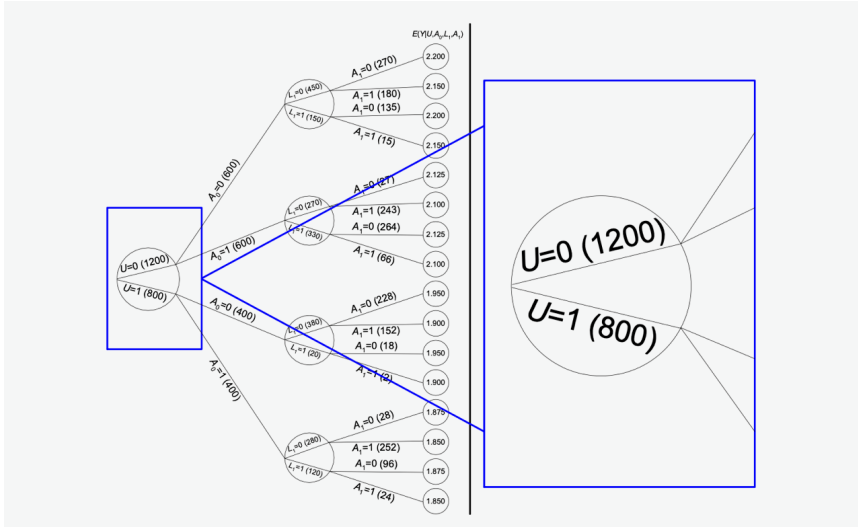
$$Pr(A_0 = 1 \,|\, U) = 0.5$$

$$Pr(L_1 = 1 \,|\, U, A_0) = 0.25 + 0.3A_0 - 0.2U - 0.05A_0U$$

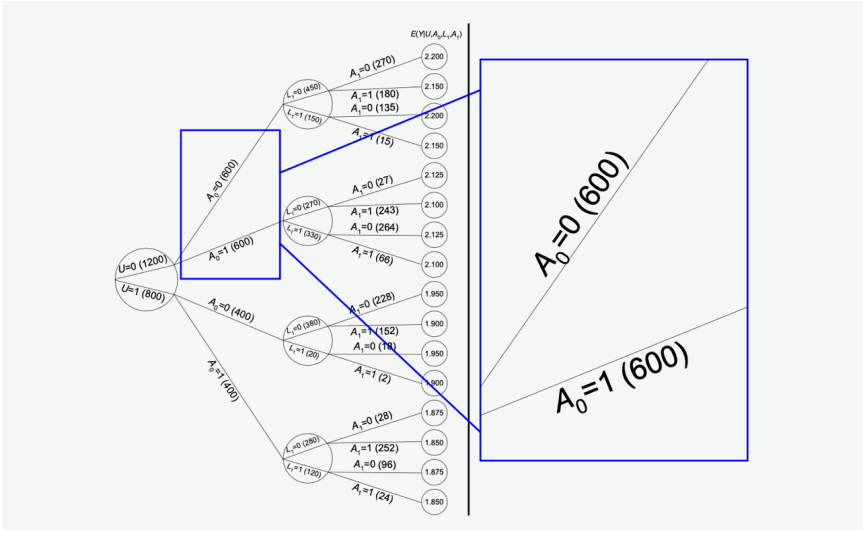
$$Pr(A_1 = 1 \,|\, U, A_0, L_1) = 0.4 + 0.5A_0 - 0.3L_1 - 0.4A_0L_1$$

$$Y \sim N(\mu_Y, 0.12^2) \text{ where } \mu_Y = 2.2 - 0.075A_0 - 0.05A_1 - 0.25U + 0.025A_0A_1$$

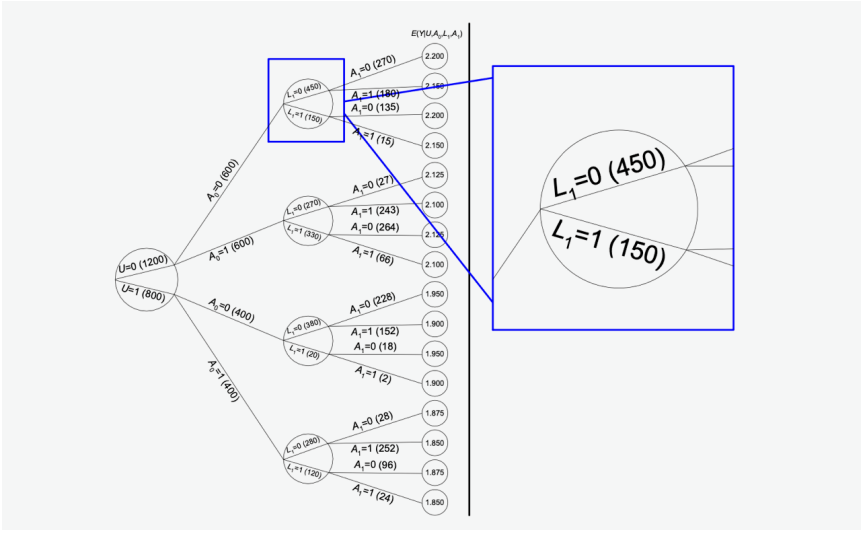
THE DATA-GENERATING MECHANISM (1)



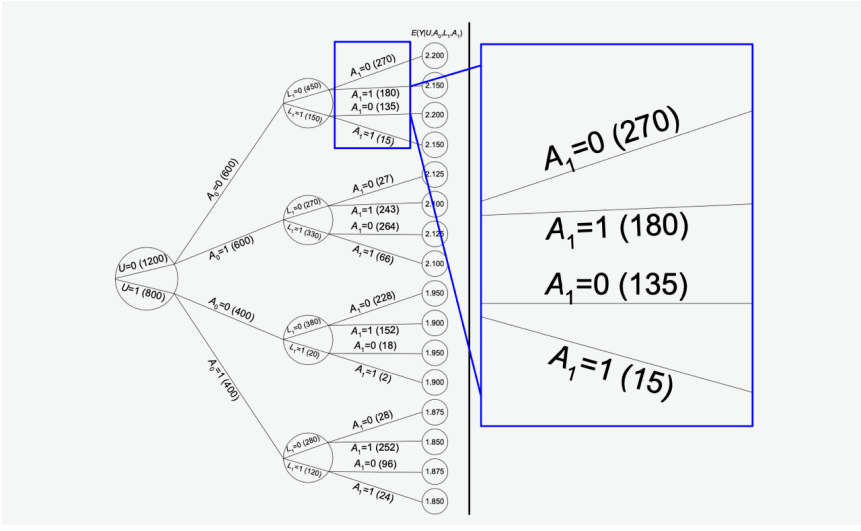
THE DATA-GENERATING MECHANISM (2)



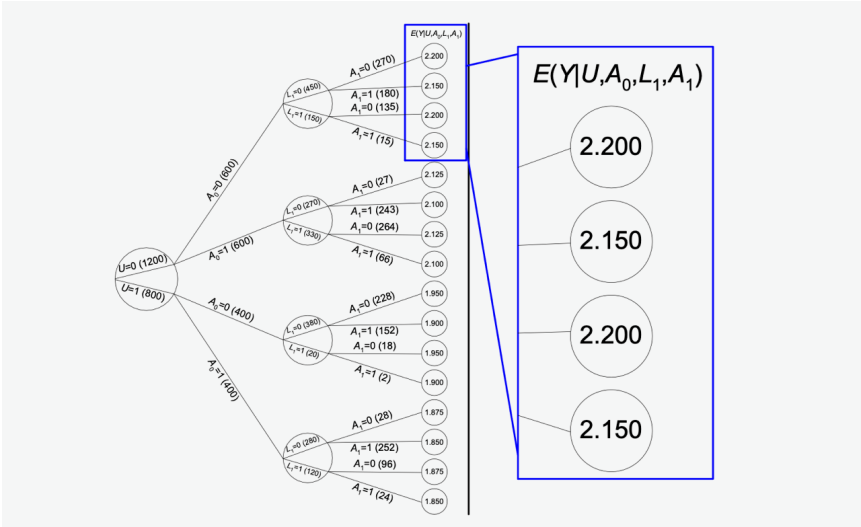
THE DATA-GENERATING MECHANISM (3)



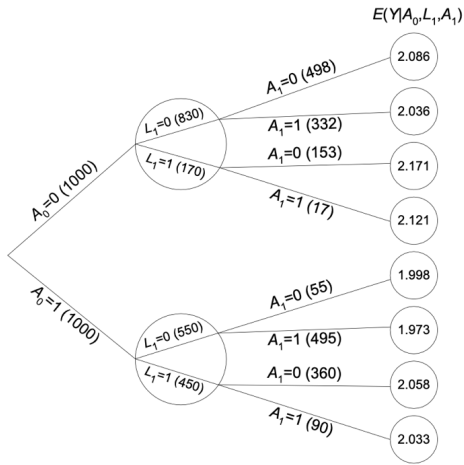
THE DATA-GENERATING MECHANISM (4)



THE DATA-GENERATING MECHANISM (5)



THE OBSERVED DATA



A CLOSER LOOK (1)

- Consider the stratum with $A_0 = 0$ and $L_1 = 0$.
- The 332 patients who crossed are exchangeable with the 498 patients who did not cross.
  - So if none had crossed, we would still see an expected outcome of 2.086.
  - So if all had crossed, we would still see an expected outcome of 2.036.
- We can mimic this situation by weighing each patient's data by

$$\frac{1}{P(A_1|A_0, L_1, L_0)}$$

A CLOSER LOOK (1)

- Consider the stratum with $A_0 = 0$ and $L_1 = 0$.
- The 332 patients who crossed are exchangeable with the 498 patients who did not cross.
  - So if none had crossed, we would still see an expected outcome of 2.086.
  - So if all had crossed, we would still see an expected outcome of 2.036.
- We can mimic this situation by weighing each patient's data by

$$\frac{1}{P(A_1|A_0, L_1, L_0)}$$

- This means
  - weighing each patient who did not cross by $830/498 = 5/3$, so the 498 patients now represent 830.
  - weighing each patient who did cross by $830/332 = 5/2$, so the 332 patients now represent 830.

A CLOSER LOOK (1)

- Consider the stratum with $A_0 = 0$ and $L_1 = 0$.
- The 332 patients who crossed are exchangeable with the 498 patients who did not cross.
  - So if none had crossed, we would still see an expected outcome of 2.086.
  - So if all had crossed, we would still see an expected outcome of 2.036.
- We can mimic this situation by weighing each patient's data by

$$\frac{1}{P(A_1|A_0, L_1, L_0)}$$

- This means
  - weighing each patient who did not cross by $830/498 = 5/3$, so the 498 patients now represent 830.
  - weighing each patient who did cross by $830/332 = 5/2$, so the 332 patients now represent 830.
- *What would you do with the stratum $A_0 = 0, L_1 = 1$?*

A CLOSER LOOK (2)

- For all control patients (A_0), we have now worked out what we expect to see if all cross, or no one crosses.
- The 1000 control patients are exchangeable with the 1000 treated patients.
- So if all 2000 patients received control and were not crossed, we would still see the same.
  - *How many would have disease progression?*

A CLOSER LOOK (2)

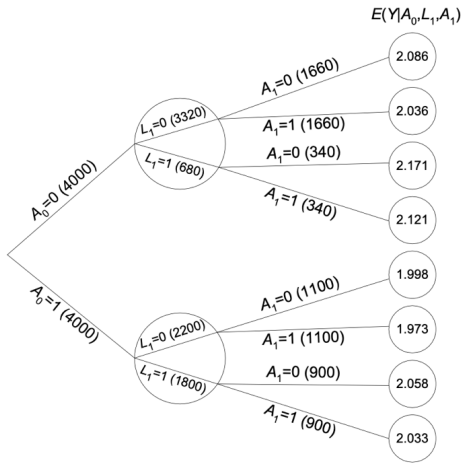
- For all control patients (A_0), we have now worked out what we expect to see if all cross, or no one crosses.
- The 1000 control patients are exchangeable with the 1000 treated patients.
- So if all 2000 patients received control and were not crossed, we would still see the same.
  - *How many would have disease progression?*
  - 1660 would progress and 340 would not.
  - Those 1660 would still have an expected outcome of 2.086, and the 340 an expected outcome of 2.171.

A CLOSER LOOK (2)

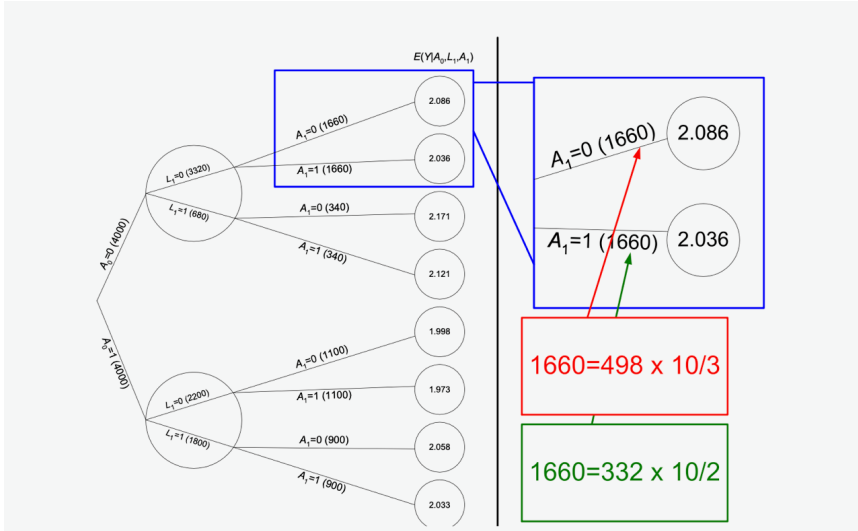
- For all control patients (A_0), we have now worked out what we expect to see if all cross, or no one crosses.
- The 1000 control patients are exchangeable with the 1000 treated patients.
- So if all 2000 patients received control and were not crossed, we would still see the same.
  - *How many would have disease progression?*
  - 1660 would progress and 340 would not.
  - Those 1660 would still have an expected outcome of 2.086, and the 340 an expected outcome of 2.171.
- We can mimic this situation by weighing each patient's data by

$$\frac{1}{P(A_0|L_0)P(A_1|A_0, L_1, L_0)}$$

THE INVERSE PROBABILITY WEIGHTED TREE



ZOOMING IN


$E(Y|A_0,L_1,A_1)$
A_0=0 (4000)
A_0=1 (4000)
L_1=0 (3320)
L_1=1 (680)
L_1=0 (2200)
L_1=1 (1800)
A_1=0 (1660)
A_1=1 (1660)
A_1=0 (340)
A_1=1 (340)
A_1=0 (1100)
A_1=1 (1100)
A_1=0 (900)
A_1=1 (900)
2.086
2.036
2.171
2.121
1.998
1.973
2.058
2.033
A_1=0 (1660) 2.086
A_1=1 (1660) 2.036
1660=498 x 10/3
1660=332 x 10/2

How to do the analysis?

- *Why does it seem like we are randomizing 8000 patients?*

How to do the analysis?

- *Why does it seem like we are randomizing 8000 patients?*
- Because each patient now follows each treatment strategy,
  it is now valid to select the patients with $A_0 = A_1$
  and analyze their outcomes in the usual way, so long as we weigh them.

How to do the analysis?

- *Why does it seem like we are randomizing 8000 patients?*
- Because each patient now follows each treatment strategy,
  it is now valid to select the patients with $A_0 = A_1$
  and analyze their outcomes in the usual way, so long as we weigh them.
- Since only treatment strategies before the endpoint assessment matter,
  there is no need for weighting by propensity scores for later time points.
- We will thus
  - censor control patient's data at the time of cross-over,
  - weigh the risk set at time t by 1 over the product of probabilities of not crossing until t or cross-over, whichever comes first.

How to do the analysis?

- *Why does it seem like we are randomizing 8000 patients?*
- Because each patient now follows each treatment strategy,
  it is now valid to select the patients with $A_0 = A_1$
  and analyze their outcomes in the usual way, so long as we weigh them.
- Since only treatment strategies before the endpoint assessment matter,
  there is no need for weighting by propensity scores for later time points.
- We will thus
  - censor control patient's data at the time of cross-over,
  - weigh the risk set at time t by 1 over the product of probabilities of not crossing until t or cross-over, whichever comes first.
- *Why only weigh until the time of cross-over?*
- *What weights to apply in the treatment arm?*

Illustration in R

GHENT
UNIVERSITY

Illustration on the SHIVA01 Trial

(Graffeo, Latouche, Le Toruneau and Chevret, 2019)

- First randomized clinical trial comparing Molecularly Targeted Therapy (MTA) based on tumor molecular profiling versus Conventional Therapy (CT) for advanced cancer.
- Treatment arm switches proposed at disease progression for both groups.
- Primary endpoint: Progression-Free Survival (time from randomization to death or progression).
- 100 patients in MTA arm, 97 in CT arm.
- 167 disease progressions, 27 deaths before progression.
- 95 patients switched arms after progression (25 MTA to CT, 70 CT to MTA).

Long data format

```
> install.packages("ipcwswitch")
> library("ipcwswitch")
> head(SHIlong)
```

	id	tstart	tstop	event	agerand	bras.f	debttCO	ps	ttc	tran
1	1	0 days	28 days	0	76.62628	CT	31 days	1	0	0
2	1	28 days	133 days	0	76.62628	CT	31 days	1	1	0
3	1	133 days	145 days	1	76.62628	CT	31 days	2	1	0
4	2	0 days	34 days	0	65.10541	MTA	NA days	1	0	0
5	2	34 days	64 days	1	65.10541	MTA	NA days	3	1	0
6	3	0 days	20 days	0	36.82341	CT	127 days	1	1	0
7	3	20 days	48 days	0	36.82341	CT	127 days	1	0	0
8	3	48 days	70 days	0	36.82341	CT	127 days	0	1	0
9	3	70 days	127 days	0	36.82341	CT	127 days	1	1	0
10	3	127 days	155 days	0	36.82341	CT	127 days	2	1	1
11	3	155 days	184 days	0	36.82341	CT	127 days	2	1	0
12	3	184 days	287 days	1	36.82341	CT	127 days	1	1	0
13	4	0 days	30 days	0	64.41547	MTA	30 days	1	0	0
14	4	30 days	59 days	0	64.41547	MTA	30 days	0	0	0
15	4	59 days	156 days	1	64.41547	MTA	30 days	1	0	0

Update long format to include all jump times and censor at the time of cross-over

```
> head(SHIrep)

     id bras.f debttCO ps ttc tran cens tstart tstop event
5193  1     CT 31 days  1   0    0    0      0    19     0
5194  1     CT 31 days  1   0    0    0     19    20     0
5195  1     CT 31 days  1   0    0    0     20    24     0
5196  1     CT 31 days  1   0    0    0     24    25     0
5197  1     CT 31 days  1   0    0    0     25    26     0
5198  1     CT 31 days  1   0    0    0     26    27     0
5199  1     CT 31 days  1   0    0    0     27    28     0
5200  1     CT 31 days  1   1    0    0     28    30     0
5201  1     CT 31 days  1   1    0    1     30    31     0
```

CALCULATE INVERSE PROBABILITY OF CENSORING WEIGHTS

```
> install.packages("ipw")
> library(ipw)
> temp <- ipwtm(
      exposure = cens,
      family = "survival",
      denominator = ~ bras.f + agerand + sex.f + tt_Lnum + rmh_alea.c + pathway.f + ps + ttc + tran,
      id = id,
      tstart = tstart,
      timevar = tstop,
      type = "cens",
      data = SHIrep)
```

With A the time to cross-over, weights at time t will be calculated as:

$$\prod_{t_s \leq t} \frac{1}{P(A > t_s | A > t_{s-1}, \overline{L}_{t_s})}$$

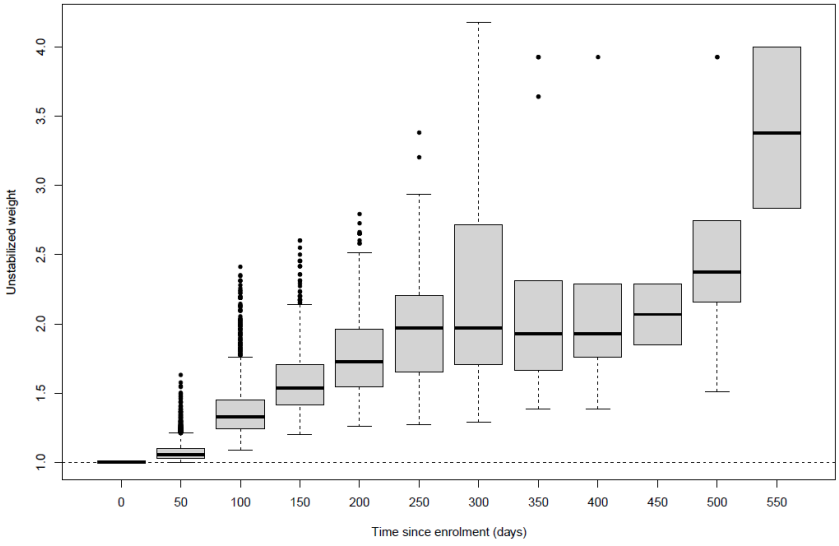
where the product runs over all times at which events, cross-overs or visit times occur.

PLOT INVERSE PROBABILITY OF CENSORING WEIGHTS

```
     id bras.f debttCO ps ttc tran cens tstart tstop event  weights
5193  1     CT 31 days  1   0    0    0      0    19     0 1.000000
5194  1     CT 31 days  1   0    0    0     19    20     0 1.000000
5195  1     CT 31 days  1   0    0    0     20    24     0 1.000000
5196  1     CT 31 days  1   0    0    0     24    25     0 1.000000
5197  1     CT 31 days  1   0    0    0     25    26     0 1.016789
5198  1     CT 31 days  1   0    0    0     26    27     0 1.016789
5199  1     CT 31 days  1   0    0    0     27    28     0 1.033878
5200  1     CT 31 days  1   1    0    0     28    30     0 1.045380
5201  1     CT 31 days  1   1    0    1     30    31     0 1.068904
```

```
> ipwplot(weights = temp$ipw.weights, timevar = SHIres$tstop,binwidth = 50,
     logscale = F, xlab = "Time since enrolment (days)", ylab = "Unstabilized weight")
```

UNTRUNCATED WEIGHTS



ITT Cox analysis

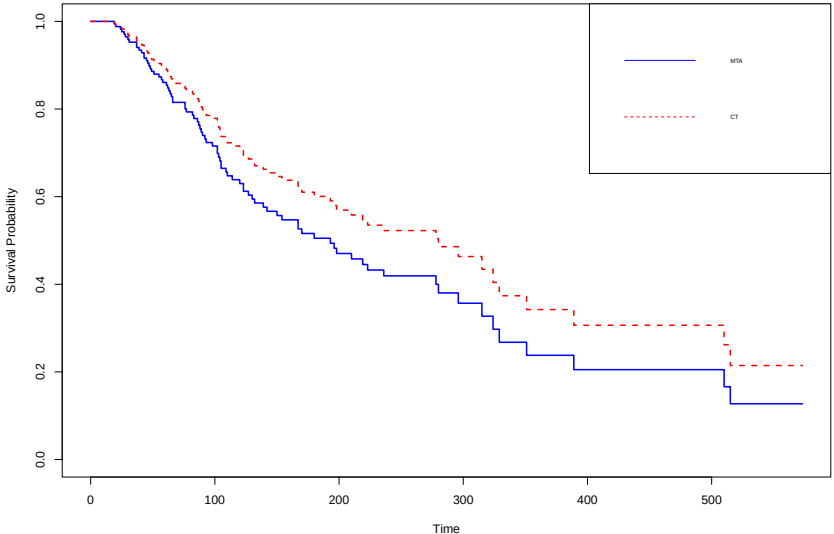
```
> itt_model <- coxph(Surv(tstart, tstop, event) ~ bras.f + cluster(id), data = SHIlong2)
> itt_model

            coef exp(coef) se(coef) robust se      z     p
bras.fCT -0.2923    0.7465   0.2367    0.2294 -1.274 0.203

Likelihood ratio test=1.57  on 1 df, p=0.21
n= 458, number of events= 83


> survival_curves <- survfit(itt_model,newdata=data.frame(bras.f=c("MTA","CT")))
> plot(survival_curves, col = c("blue", "red"), lty = 1:2, xlab = "Time",
     ylab = "Survival Probability")
> legend("topright", legend = c("MTA", "CT"), cex = 0.5, col = c("blue", "red"), lty = 1:2)
```

SURVIVAL CURVES: ITT ANALYSIS



IPCW Cox analysis

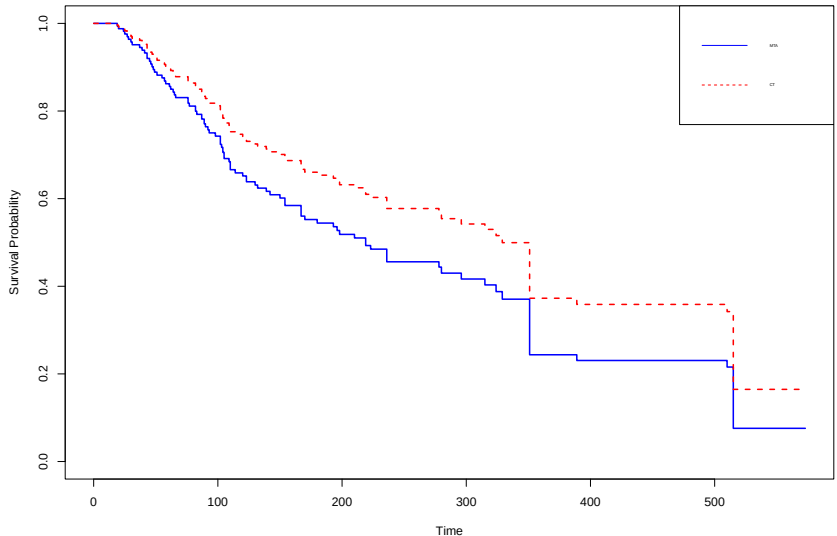
```
> fit.unstab.w <- coxph(Surv(tstart, tstop, event) ~ bras.f + cluster(id),
              data = SHIres, weights = temp$ipw.weights)
> fit.unstab.w

             coef exp(coef) se(coef) robust se      z     p
bras.fCT -0.3579    0.6991   0.1839    0.2540 -1.409 0.159

Likelihood ratio test=3.76  on 1 df, p=0.05264
n= 9745, number of events= 76
```

Only robust standard errors are valid (but may be conservative)!

SURVIVAL CURVES: HYPOTHETICAL ESTIMAND



CALCULATE STABILIZED INVERSE PROBABILITY WEIGHTS

```
> temp <- ipwtm(exposure = cens, family = "survival",
   numerator = ~ bras.f + agerand + sex.f + tt_Lnum + rmh_alea.c + pathway.f,
   denominator = ~ bras.f + agerand + sex.f + tt_Lnum + rmh_alea.c + pathway.f + ps + ttc + tran,
   id = id, tstart = tstart, timevar = tstop, type = "cens", data = SHIrep)
> ipwplot(weights = temp$ipw.weights, timevar = SHIres$tstop,binwidth = 50, logscale = F,
   xlab = "Time since enrolment (days)", ylab = "Stabilized weight")
```

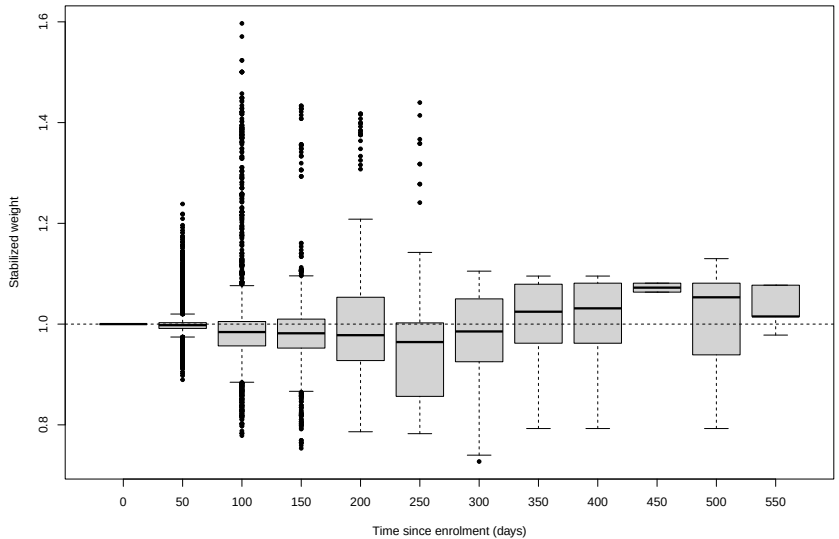
- With A the time to cross-over, the weight calculation stabilizes the weights at time t as follows:

$$\prod_{t_s \leq t} \frac{P(A > t_s | A > t_{s-1}, L_0)}{P(A > t_s | A > t_{s-1}, \overline{L}_{t_s})}$$

  where the product runs over all times at which events, cross-overs or visit times occur.

- These weights reduce to 1 if there is only confounding by baseline covariates, suggesting that adjustment for baseline covariates is now crucial (but can be done via standard regression).

UNTRUNCATED STABILIZED WEIGHTS



IPW Cox analysis with stabilized weights

```
> fit.stab.w <- coxph(Surv(tstart, tstop, event) ~ bras.f + agerand + sex.f + tt_Lnum
   + rmh_alea.c + pathway.f + cluster(id),
   data = SHIrep, weights = temp$ipw.weights)
> fit.stab.w

                           coef exp(coef)  se(coef) robust se      z        p
bras.fCT                -0.338454  0.712872  0.254708  0.255541 -1.324 0.185350
agerand                 -0.007028  0.992997  0.009823  0.010263 -0.685 0.493496
sex.fFemale             -0.492635  0.611014  0.239527  0.244836 -2.012 0.044209
tt_Lnum                  0.010644  1.010700  0.046586  0.041585  0.256 0.797989
rmh_alea.c               0.941328  2.563384  0.251063  0.253190  3.718 0.000201
pathway.fHR             -0.137047  0.871930  0.342980  0.357161 -0.384 0.701193
pathway.fPI3K/AKT/mTOR -0.177219  0.837597  0.340084  0.349221 -0.507 0.611826

Likelihood ratio test=23.93  on 7 df, p=0.001171
n= 9745, number of events= 76
```

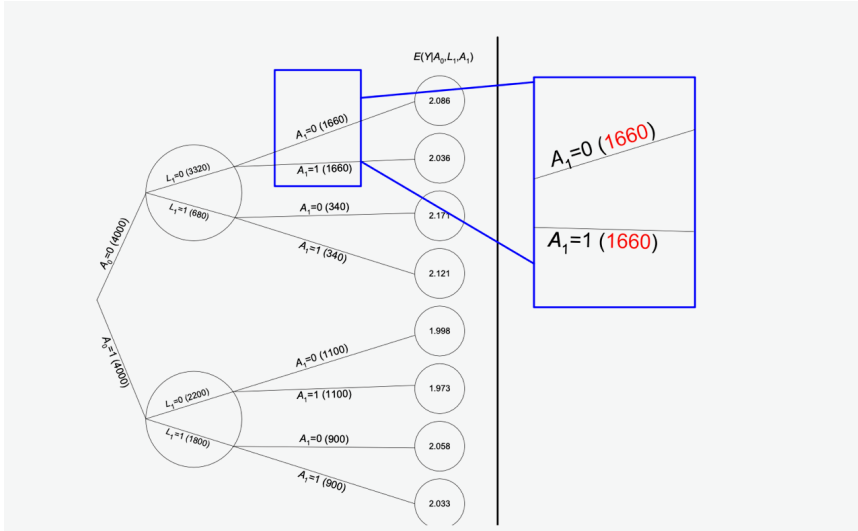
Note that population survival curves are now harder to obtain,

unless the numerator weights only include treatment (which we anyway adjust for)!

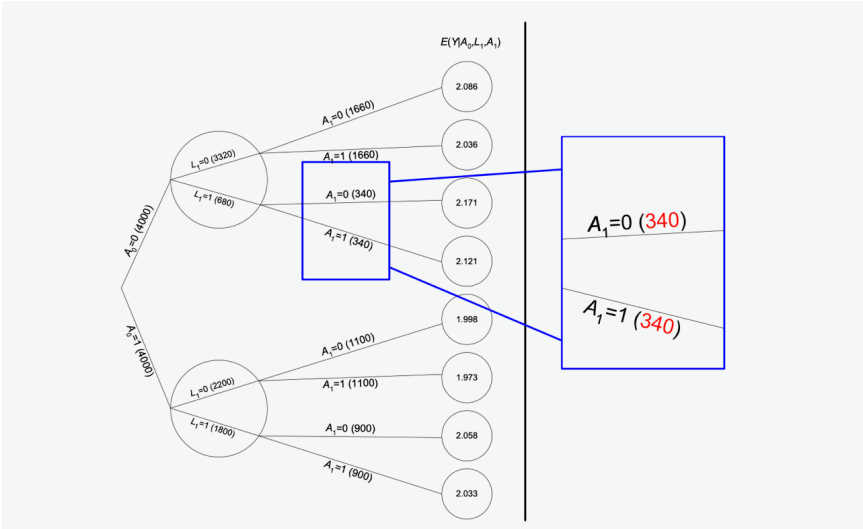
IPW ELIMINATES CONFOUNDING

GHENT
UNIVERSITY

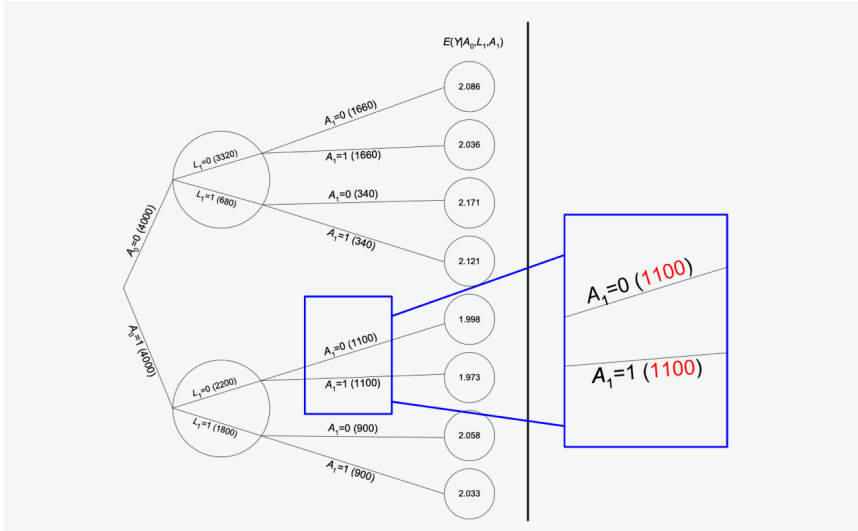
IPW ELIMINATES CONFOUNDING (1)



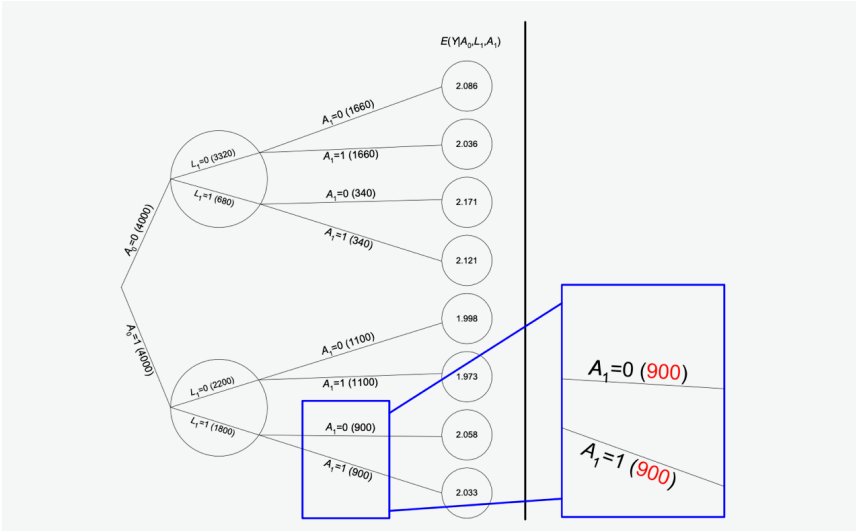
IPW ELIMINATES CONFOUNDING (2)



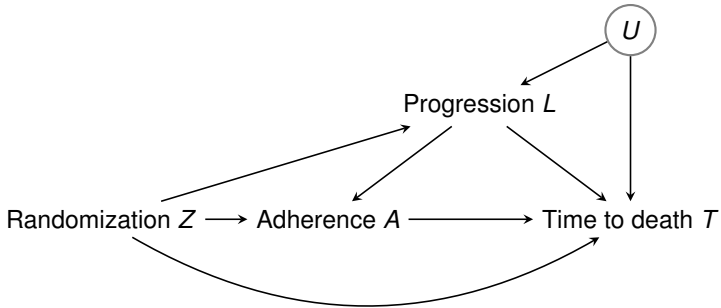
IPW ELIMINATES CONFOUNDING (3)



IPW ELIMINATES CONFOUNDING (4)

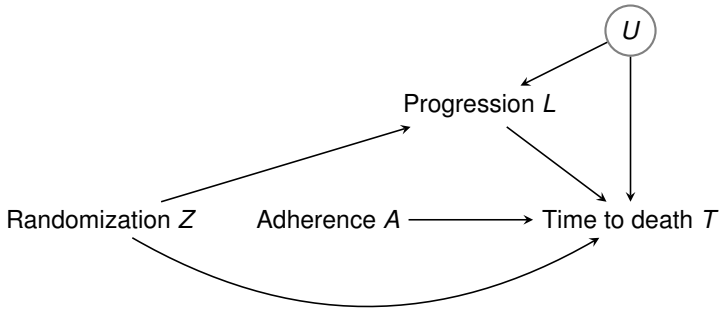


THE ORIGINAL CAUSAL DIAGRAM



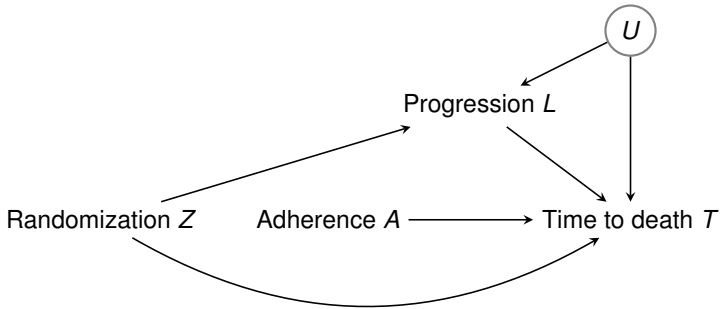
What combination of paths corresponds with the hypothetical effect of interest?

THE INVERSE PROBABILITY WEIGHTED CAUSAL DIAGRAM



What analysis would you propose in this diagram?

THE INVERSE PROBABILITY WEIGHTED CAUSAL DIAGRAM



What analysis would you propose in this diagram?

As if cross-over were randomized.

IPW FOR THE EFFECT OF A TIME-VARYING EXPOSURE

- Suppose we are interested in a collection of regimes $\overline{a}_t$,
  e.g. $\overline{a}_t = (0, ..., 0), \overline{a}_t = (0, ..., 0, 1), ..., \overline{a}_t = (0, 1, ..., 1), \overline{a}_t = (1, ..., 1)$.
- It then follows that one may fit model

$$\lambda(t|\overline{A}_t) = \lambda_0(t) \exp(\beta A_t)$$

  using weighted Cox regression on a large dataset
  by pooling datasets that censor patients when they deviate from a given regime $\overline{a}_t$.

IPW FOR THE EFFECT OF A TIME-VARYING EXPOSURE

- Suppose we are interested in a collection of regimes $\overline{a}_t$,
  e.g. $\overline{a}_t = (0, ..., 0), \overline{a}_t = (0, ..., 0, 1), ..., \overline{a}_t = (0, 1, ..., 1), \overline{a}_t = (1, ..., 1)$.
- It then follows that one may fit model

$$\lambda(t|\overline{A}_t) = \lambda_0(t) \exp(\beta A_t)$$

  using weighted Cox regression on a large dataset
  by pooling datasets that censor patients when they deviate from a given regime $\overline{a}_t$.
- Since the weighting 'reproduces data' for all these regimes,
  the results can be interpreted as for the marginal structural model

$$\lambda_{T^{\overline{a}_t}}(t) = \lambda_0(t) \exp(\beta a_t)$$

(Robins, Hernan and Brumback, 2000; Hernan, Robins and Brumback, 2001)

- These express how the results would have looked like
  if the exposure was randomised at each time.

Discussion

GHENT
UNIVERSITY

Summary

- Conventional methods are fallible
  for the analysis of time-varying exposures.
- This is because time-varying confounders
  are both causes and effects of the exposure.
- Like cross-sectional analyses,
  conventional analyses cannot distinguish these.

Summary

- Conventional methods are fallible
  for the analysis of time-varying exposures.
- This is because time-varying confounders
  are both causes and effects of the exposure.
- Like cross-sectional analyses,
  conventional analyses cannot distinguish these.
- Inverse probability weighting offers a way out.
- After reweighting the data,
  there is no further need to adjust for time-varying confounders.

Summary

- Conventional methods are fallible
  for the analysis of time-varying exposures.
- This is because time-varying confounders
  are both causes and effects of the exposure.
- Like cross-sectional analyses,
  conventional analyses cannot distinguish these.
- Inverse probability weighting offers a way out.
- After reweighting the data,
  there is no further need to adjust for time-varying confounders.
- Results can be interpreted
  as if they originated from a sequentially randomised trial,
  provided that all time-varying confounders were available,
  and correctly modelled via propensity scores.

POSITIVITY VIOLATIONS

- A major drawback of IPW methods is that are very inefficient and can be unstable.
- Augmented IPW methods are much more efficient, but computationally difficult.
  - Software is developing, e.g. `ltmle`.

POSITIVITY VIOLATIONS

- A major drawback of IPW methods is that are very inefficient and can be unstable.
- Augmented IPW methods are much more efficient, but computationally difficult.
  - Software is developing, e.g. `ltmle`.
- The non-stability of IPW can also be a sign of positivity violations.
  - This is not uncommon with intercurrent events!
  - Other estimators or estimands can provide some remedy.

(Michiels, Vandebosch and Vansteelandt, 2021, 2023)

POSITIVITY VIOLATIONS

- A major drawback of IPW methods is that are very inefficient and can be unstable.
- Augmented IPW methods are much more efficient, but computationally difficult.
  - Software is developing, e.g. `ltmle`.
- The non-stability of IPW can also be a sign of positivity violations.
  - This is not uncommon with intercurrent events!
  - Other estimators or estimands can provide some remedy.

    (Michiels, Vandebosch and Vansteelandt, 2021, 2023)
  - *How would you handle IPW in the Helios trial, considering that all cross after 3 years? What estimand would you target?*

POSITIVITY VIOLATIONS

- A major drawback of IPW methods is that are very inefficient and can be unstable.
- Augmented IPW methods are much more efficient, but computationally difficult.
  - Software is developing, e.g. `ltmle`.
- The non-stability of IPW can also be a sign of positivity violations.
  - This is not uncommon with intercurrent events!
  - Other estimators or estimands can provide some remedy.

    (Michiels, Vandebosch and Vansteelandt, 2021, 2023)
  - *How would you handle IPW in the Helios trial, considering that all cross after 3 years? What estimand would you target?*
- The 'standard' use of weight truncation can be problematic and hide positivity violations.
  - Sensitivity analyses are essential and must be reported.

SELECTED REFERENCES

Bekaert, M., Timsit, J.-F., Vansteelandt, S., Depuydt, P., Vosin, A., Garrouste-Orgeas, M., Decruyenaere, J., Clec'h, C., Azoulay, E., Benoit, D. Attributable Mortality of Ventilator Associated Pneumonia: A Reappraisal Using Causal Analysis. American Journal of Respiratory and Critical Care Medicine 2011, 184, 1133-1139.

Daniel R, Cousens S, De Stavola B, Kenward M, Sterne J. Methods for dealing with time-dependent confounding. Statistics in Medicine 2013; 32:1584-1618.

Hernan M, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. Epidemiology 2000, 11(5):561-570.

Robins JM, Hernan M, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology 2000, 11(5):550-560.

Robins J, Hernan M (2009). Longitudinal Data Analysis, chap. Estimation of the causal effects of time-varying exposures. Chapman and Hall/CRC Press: New York; 553-599.

Vansteelandt S, Joffe M. Structural nested models and G-estimation: the partially realized promise. Statistical Science 2014; 29:707-731.