University of Sheffield

Emulating Target Trials to compare the effectiveness of treatment sequences or pathways

ISPOR Europe 2024, Workshop 201

Jen-Yu Amy Chang

Research Associate, Sheffield Centre for Health and Related Research(SCHARR), University of Sheffield

Health Economist, Lumanity

jy.a.chang@sheffield.ac.uk

19th November 2024

A World TOP 100 University

WHAT UNI? STUDENT CHOICE Awards 2024 UNIVERSITY OF THE YEAR

Acknowledgement & Disclosure

Funders

This work was conducted during Dr. Chang's time as a PhD candidate and employee at the University of Sheffield.

Collaborators

Prof Nicholas Latimer
Prof Jim Chilcott

Dr Carmel Pezaro (Swansea Bay University Health Board, NHS Wales for Swansea and Neath)

Professor Derek Rosario (Sheffield Teaching Hospitals NHS Foundation Trust)

Dr Blythe Adamson
Dr Philani Mpofu

And also, sincere appreciation for the insights gained from courses at Harvard University's CAUSALab, led by Prof. Miguel Hernán.

Estimating the (causal) comparative effectiveness of treatment sequences in RWD for HTA

- Advantages
  - Capture treatment sequences not compared in trials
  - Relaxed the exchangeability assumption required for estimating line-of-treatment effects.
- Challenges
  - *Baseline and time-varying confounding* not addressable with simple multivariate regression/survival models

    ➔ Need causal inference guided advanced statistical methods
  - *Immortal time bias and attrition issues* cannot be addressed by statistical methods alone

    ➔ Need careful study design e.g. Target Trial Emulation (TTE)[1]

Ref: [1]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.

Target Trial Emulation for comparing treatment sequences

1. Prevents immortal time bias by explicitly designing and emulating a trial comparing two treatment sequences. (i.e., avoiding selecting patients based on post-baseline characteristics like subsequent treatments)
2. Guides the use of advanced statistical methods to emulate hypothetical randomisation to address baseline & time-varying confounding.

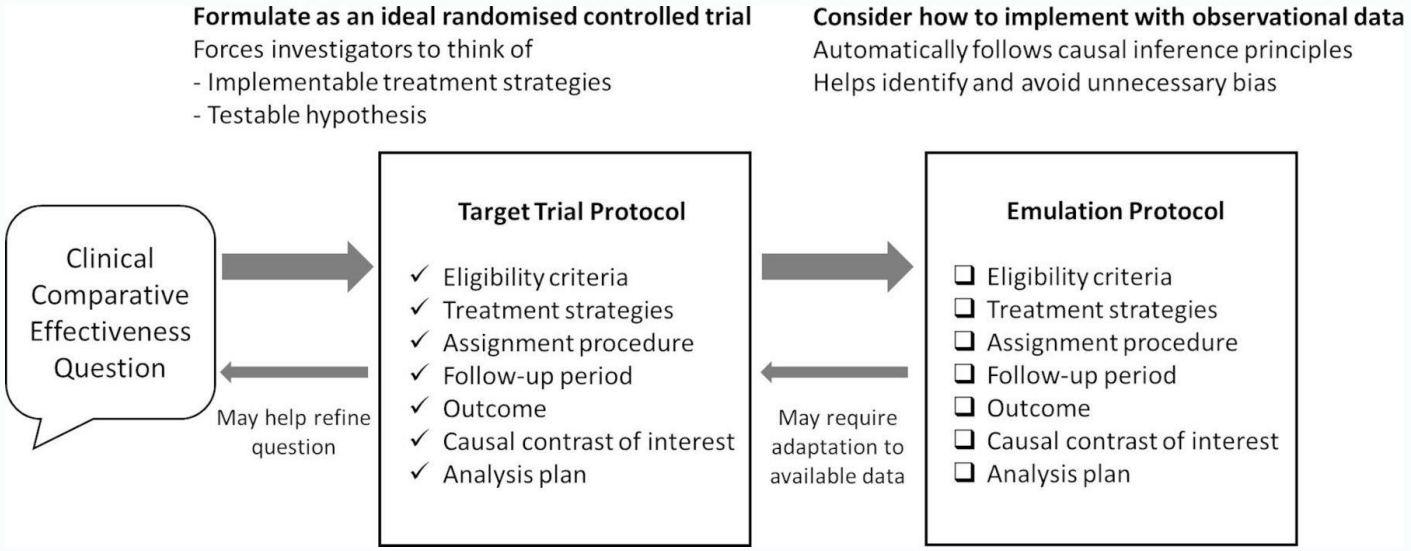


Fig ref: Zhao, et al. Improving rheumatoid arthritis comparative effectiveness research through causal inference principles: systematic review using a target trial emulation framework. *Annals of the rheumatic diseases*, *79*(7), 883-890.

Proof-of-concept case studies: Benchmarking Target Trial Emulations
in US Flatiron database and UK Cancer Registry

A systematic review[1]
informed benchmark RCT

The benchmark RCT[2]

Reference: GUTG-001 trial OS

Population: GUTG-001 Trial (mCRPC patients)

Treatments: Abiraterone → Enzalutamide vs Enzalutamide → Abiraterone

⇔

Method validation TTE study[3]

Data: US Flatrion

Population: GUTG-001 Analogue

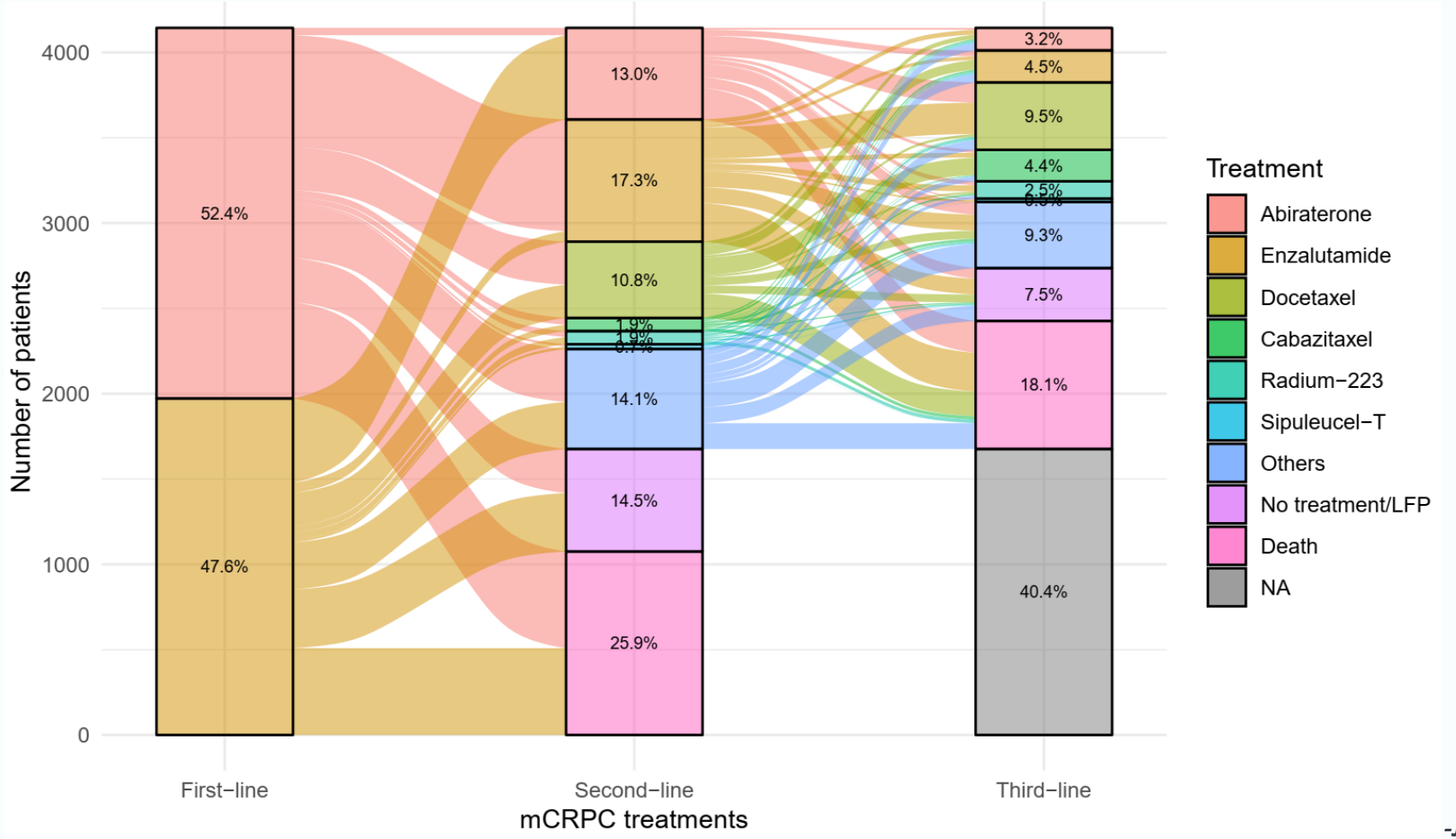
Treatments: Abiraterone → Enzalutamide vs Enzalutamide → Abiraterone

Benchmarking: Comparing results

Keys: GUTG-001: A phase II RCT of sequencing abiraterone acetate and enzalutamide in mCRPC; mCRPC: metastatic castration-resistant prostate cancer; OS, overall survival; RCT, randomised controlled trial

Ref: [1]Chang JYA. Investigating the Application of Causal Inference Methods for Modelling the Impact of Treatment Sequences in Health Economic Evaluations 2024. (PhD Thesis, University of Sheffield).
[2]Khalaf et al. Optimal sequencing of enzalutamide and abiraterone acetate plus prednisone in metastatic castration-resistant prostate cancer: a multicentre, randomised, open-label, phase 2, crossover trial. Lancet Oncol. 2019; 20(12):1730-1739.
[3]Chang JYA, Chilcott JB, Latimer NR. Leveraging real-world data to assess treatment sequences in health economic evaluations: a study protocol for emulating target trials using the English Cancer Registry and US Electronic Health Records-Derived Database. HEDS Discussion Paper, 2024 (1).

Real-world treatment patterns of TTE GUTG-001 Analogue patients
in Flatiron database



-Journal publication in progress-

Keys: NA, not applicable.

Ref: Chang JYA. Investigating the Application of Causal Inference Methods for Modelling the Impact of Treatment Sequences in Health Economic Evaluations 2024. (PhD Thesis, University of Sheffield).

Challenges in estimating survival of treatment sequences in RWD
(1) immortal time bias and attrition issues (a toy example)

A hypothetical real-world treatment pattern of A as 1L treatment				
1L	2L	N	Percentage	OS (years)
A	-	40	40%	1
A	E	40	40%	4
A	D	20	20%	2

In real-world, patients receiving 1L A (abiraterone) can receive different 2L: E (enzalutamide), D (docetaxel)?

Challenges in estimating survival of treatment sequences in RWD

(1) immortal time bias and attrition issues (a toy example)

A hypothetical real-world treatment pattern of A as 1L treatment				
1L	2L	N	Percentage	OS (years)
A	-	40	40%	1
A	E	40	40%	4
A	**D**	20	20%	2

An approach common in studies: as-treated (AT) analysis limited to patients completing the full treatment sequence.				
1L	2L	N	Percentage	OS (years)
A	E	40	100%	4

In a counterfactual scenario, what is the effect on OS if all patients receive A → E?

Immortal time bias:
Over-estimating A→ E treatment sequence benefits

Challenges in estimating survival of treatment sequences in RWD

(1) immortal time bias and attrition issues (a toy example)

A hypothetical real-world treatment pattern of A as 1L treatment				
1L	2L	N	Percentage	OS (years)
A	-	40	40%	1
A	E	40	40%	4
A	**D**	20	20%	2

An approach common in studies: as-treated (AT) analysis limited to patients completing the full treatment sequence.				
1L	2L	N	Percentage	OS (years)
A	E	40	100%	4

Another approach common in studies: AT analysis				
1L	2L	N	Percentage	OS (years)
A	-	40	50%	1
A	E	40	50%	4

In a counterfactual scenario, what is the effect on OS if all patients receive A → E?

Immortal time bias:
Over-estimating A→ E treatment sequence benefits

Attrition issues:
Over-representing patients who did not receive 2L treatment misinterprets the effect of 1L A on OS.

Challenges in estimating survival of treatment sequences in RWD
(2a) Baseline confounding

A hypothetical RW treatment mix of A as 1L treatment					
1L	2L	N	Percentage	OS (years)	Average OS
A	-	40	40%	1	2.4
A	E	40	40%	4	
A	D	20	20%	2	

A hypothetical RW treatment mix of E as 1L treatment					
1L	2L	N	Percentage	OS (years)	Average OS
E	-	60	60%	1	2.1
E	A	30	30%	4	
E	D	10	10%	3	

Challenges in estimating survival of treatment sequences in RWD
(2a) Baseline confounding

A hypothetical RW treatment mix of A as 1L treatment					
1L	2L	N	Percentage	OS (years)	Average OS
A	-	40	40%	1	?
A	E	40	40%	4	
A	D	20	20%	2	

A hypothetical RW treatment mix of E as 1L treatment					
1L	2L	N	Percentage	OS (years)	Average OS
E	-	60	60%	1	?
E	A	30	30%	4	
E	D	10	10%	3	

G_0: 1L mCRPC treatment (time 0)

L_0: Confounders at baseline (time 0) (e.g. diabetes)

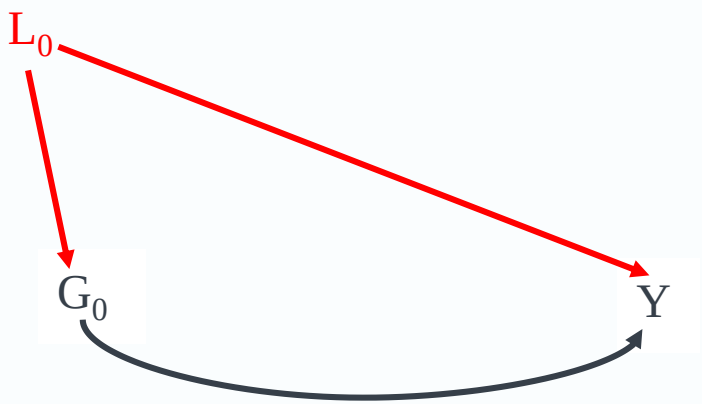
Y: Outcome (e.g. Death)

What would OS have been if all patients received Treatment A or E as 1L treatment?

Challenges in estimating survival of treatment sequences in RWD
(2a) Baseline confounding

A hypothetical RW treatment mix of A as 1L treatment					
1L	2L	N	Percentage	OS (years)	Average OS
A	-	40	40%	1	?
A	E	40	40%	4	
A	D	20	20%	2	

A hypothetical RW treatment mix of E as 1L treatment					
1L	2L	N	Percentage	OS (years)	Average OS
E	-	60	60%	1	?
E	A	30	30%	4	
E	D	10	10%	3	



G_0: 1L mCRPC treatment (time 0)

L_0: Confounders at baseline (time 0) (e.g. diabetes)

Y: Outcome (e.g. Death)

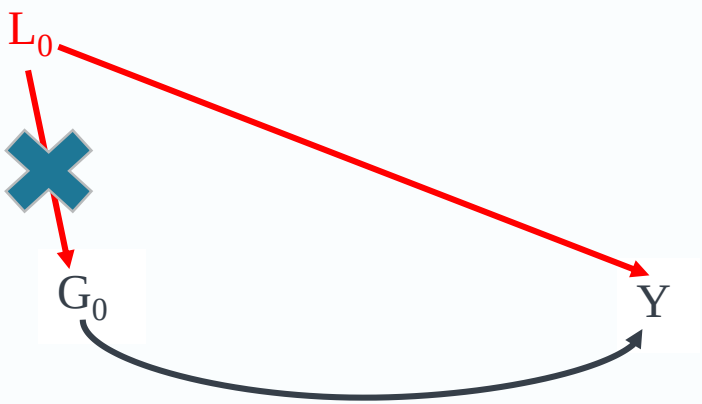
What would OS have been if all patients received Treatment A or E as 1L treatment?

Baseline confounding

Challenges in estimating survival of treatment sequences in RWD
(2a) Baseline confounding

A hypothetical RW treatment mix of A as 1L treatment					
1L	2L	N	Percentage	OS (years)	Average OS
A	-	40	40%	1	?
A	E	40	40%	4	
A	D	20	20%	2	

A hypothetical RW treatment mix of E as 1L treatment					
1L	2L	N	Percentage	OS (years)	Average OS
E	-	60	60%	1	?
E	A	30	30%	4	
E	D	10	10%	3	



G_0: 1L mCRPC treatment (time 0)

L_0: Confounders at baseline (time 0) (e.g. diabetes)

Y: Outcome (e.g. Death)

What would OS have been if all patients received Treatment A or E as 1L treatment?

Baseline confounding

Inverse probability of treatment weighting (IPTW) can be used to address this.

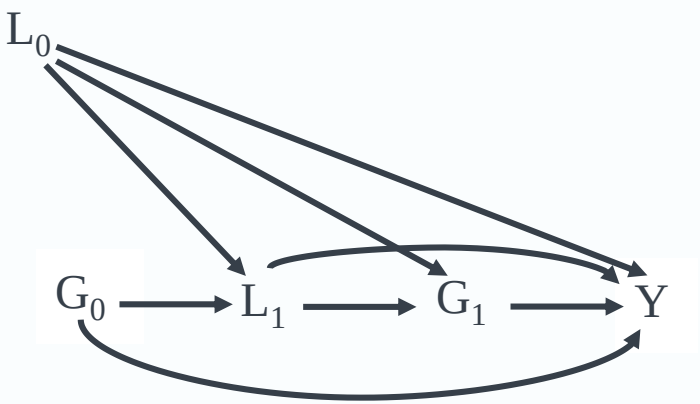
$$W^T = \begin{cases} \frac{1}{\Pr[G_0 = 1 \mid L_0]} & if\ G_0 = 1 \\ \frac{1}{1 - \Pr[G_0 = 1 \mid L_0]} & if\ G_0 = 0 \end{cases}$$

Challenges in estimating survival of treatment sequences in RWD

(2b) Time-varying confounding (within 1L-A arm)

A hypothetical RW treatment mix of A as 1L treatment					
1L	2L	N	Percentage	OS (years)	Average OS
A	-	40	40%	1	?
A	E	40	40%	4	
A	D	20	20%	2	

What would OS have been if all patients received Treatment A followed E?



G_0: 1L mCRPC treatment (time 0)

L_0: Confounders at baseline (time 0) (e.g. diabetes)

G_1: 2L mCRPC treatment at the time of progression/treatment intolerability

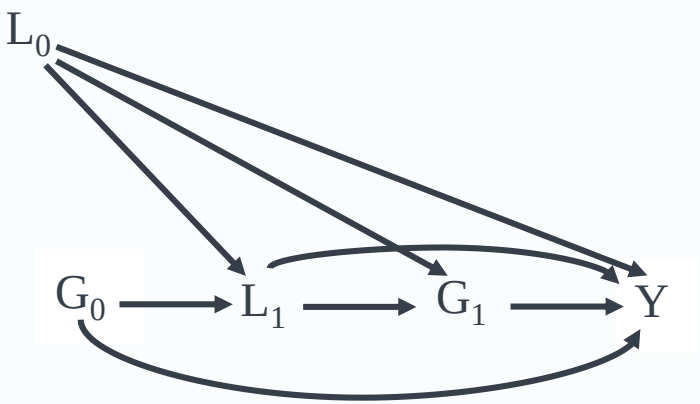
L_1: Confounders at the time of treatment-switching due to progression/treatment intolerability (e.g. performance status)

Y: Outcome (e.g. Death)

Challenges in estimating survival of treatment sequences in RWD
(2b) Time-varying confounding (within 1L-A arm)

A hypothetical RW treatment mix of A as 1L treatment					
1L	2L	N	Percentage	OS (years)	Average OS
A	-	40	40%	1	??
A	E	40	40%	4	
A	**E**	20	20%	?	

What would OS have been if all patients received Treatment A followed E?



G_0: 1L mCRPC treatment (time 0)

L_0: Confounders at baseline (time 0) (e.g. diabetes)

G_1: 2L mCRPC treatment at the time of progression/treatment intolerability

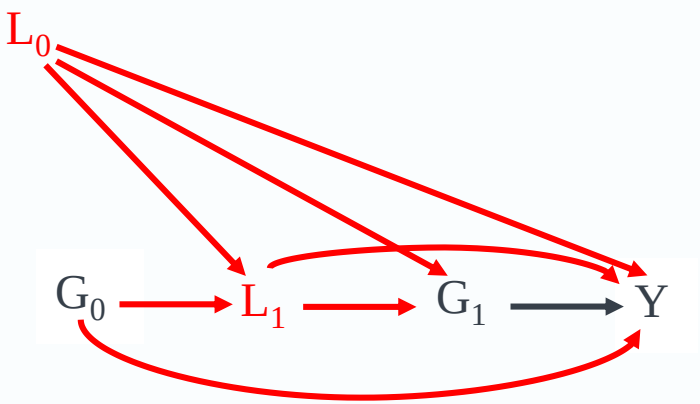
L_1: Confounders at the time of treatment-switching due to progression/treatment intolerability (e.g. performance status)

Y: Outcome (e.g. Death)

Challenges in estimating survival of treatment sequences in RWD

(2b) Time-varying confounding (within 1L-A arm)

A hypothetical RW treatment mix of A as 1L treatment					
1L	2L	N	Percentage	OS (years)	Average OS
A	-	40	40%	1	??
A	E	40	40%	4	
A	**E**	20	20%	?	



G_0: 1L mCRPC treatment (time 0)

L_0: Confounders at baseline (time 0) (e.g. diabetes)

G_1: 2L mCRPC treatment at the time of progression/treatment intolerability

L_1: Confounders at the time of treatment-switching due to progression/treatment intolerability (e.g. performance status)

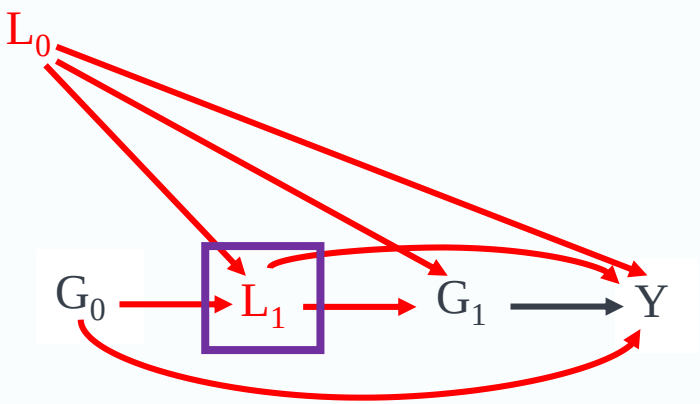
Y: Outcome (e.g. Death)

What would OS have been if all patients received Treatment A followed E?

Time-varying confounding

Challenges in estimating survival of treatment sequences in RWD
(2b) Time-varying confounding (within 1L-A arm)

A hypothetical RW treatment mix of A as 1L treatment					
1L	2L	N	Percentage	OS (years)	Average OS
A	-	40	40%	1	??
A	E	40	40%	4	
A	**E**	20	20%	?	



G_0: 1L mCRPC treatment (time 0)

L_0: Confounders at baseline (time 0) (e.g. diabetes)

G_1: 2L mCRPC treatment at the time of progression/treatment intolerability

L_1: Confounders at the time of treatment-switching due to progression/treatment intolerability (e.g. performance status)

Y: Outcome (e.g. Death)

What would OS have been if all patients received Treatment A followed E?

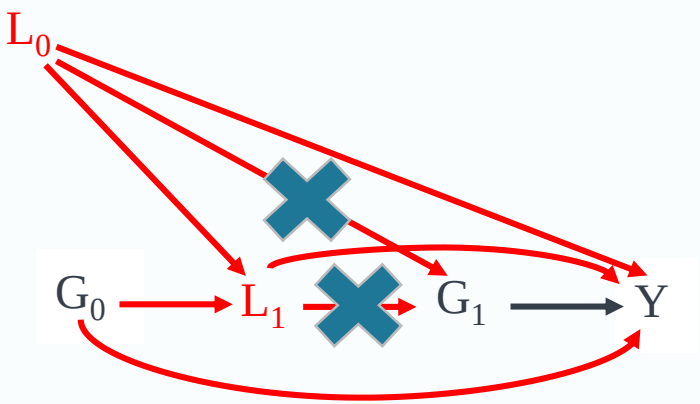
Time-varying confounding

Including time-varying confounders in survival analysis (i.e., conditioning on L_1) can bias the understanding of a treatment sequence's causal effect, while including only baseline confounders (L_0) overlooks confounding by L_1.

Challenges in estimating survival of treatment sequences in RWD

(2b) Time-varying confounding (within 1L-A arm)

A hypothetical RW treatment mix of A as 1L treatment					
1L	2L	N	Percentage	OS (years)	Average OS
A	-	40	40%	1	??
A	E	40	40%	4	
A	**E**	20	20%	?	



G_0: 1L mCRPC treatment (time 0)

L_0: Confounders at baseline (time 0) (e.g. diabetes)

G_1: 2L mCRPC treatment at the time of progression/treatment intolerability

L_1: Confounders at the time of treatment-switching due to progression/treatment intolerability (e.g. performance status)

Y: Outcome (e.g. Death)

What would OS have been if all patients received Treatment A followed E?

Time-varying confounding

Including time-varying confounders in survival analysis (i.e., conditioning on L_1) can bias the understanding of a treatment sequence's causal effect, while including only baseline confounders (L_0) overlooks confounding by L_1.

Inverse probability of censoring weighting due to treatment deviation ($IPCW_{txdev}$) can be used to address this by up-weighting patients who followed the A → E sequence to represent those who didn't.

$$W_t^D = \begin{cases} \prod_{k=0}^{t} \frac{1}{\Pr[D_k = 0 \mid G_{k-1}, L_0, L_k, C_k = 0, D_{k-1} = 0, Y_{k-1} = 0]} \\ 0 \end{cases}$$

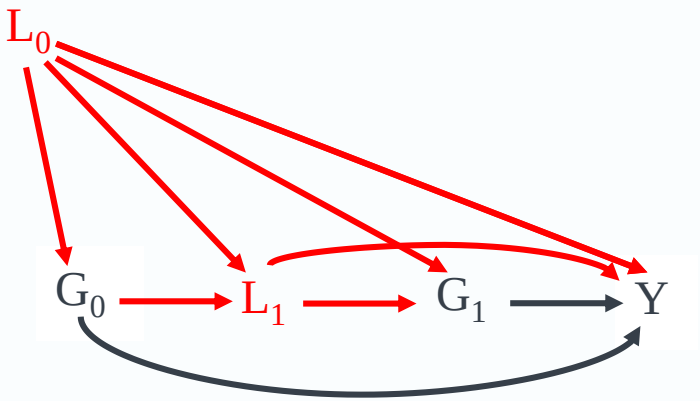
$$= \prod_{k=0}^{t} \frac{1}{1 - \Pr[D_k = 1 \mid G_{k-1}, L_0, L_k, C_k = 0, D_{k-1} = 0, Y_{k-1} = 0]} \quad \begin{matrix} if\ D_k = 0 \\ if\ D_k = 1 \end{matrix} \Big\}$$

$D_k = 0$ means keep receiving the treatment sequence of interest

Challenges in estimating survival of treatment sequences in RWD
(2c) Baseline & time-varying confounding

A hypothetical RW treatment mix of A as 1L treatment					
1L	2L	N	Percentage	OS (years)	Average OS
A	-	40	40%	1	??
A	E	40	40%	4	
A	**E**	20	20%	?	

A hypothetical RW treatment mix of E as 1L treatment					
1L	2L	N	Percentage	OS (years)	Average OS
E	-	60	60%	1	??
E	A	30	30%	4	
E	**A**	10	10%	?	



G_0: 1L mCRPC treatment (time 0)

L_0: Confounders at baseline (time 0) (e.g. diabetes)

Y: Outcome (e.g. Death)

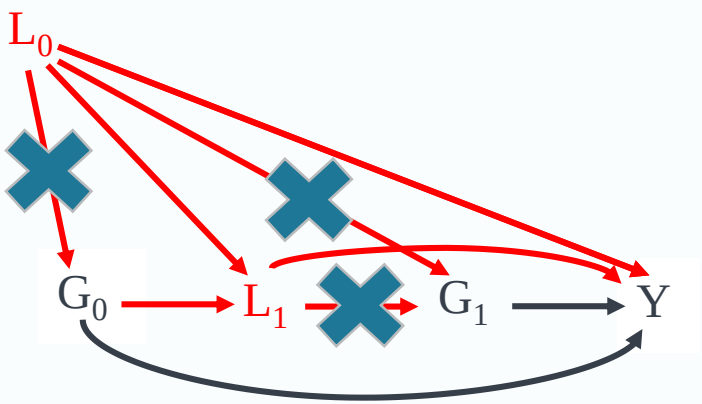
What is the causal effect on OS if all patients received A → E versus E → A?

Baseline & time-varying confounding

Challenges in estimating survival of treatment sequences in RWD
(2c) Baseline & time-varying confounding

A hypothetical RW treatment mix of A as 1L treatment					
1L	2L	N	Percentage	OS (years)	Average OS
A	-	40	40%	1	??
A	E	40	40%	4	
A	**E**	20	20%	?	

A hypothetical RW treatment mix of E as 1L treatment					
1L	2L	N	Percentage	OS (years)	Average OS
E	-	60	60%	1	??
E	A	30	30%	4	
E	**A**	10	10%	?	



G_0: 1L mCRPC treatment (time 0)

L_0: Confounders at baseline (time 0) (e.g. diabetes)

Y: Outcome (e.g. Death)

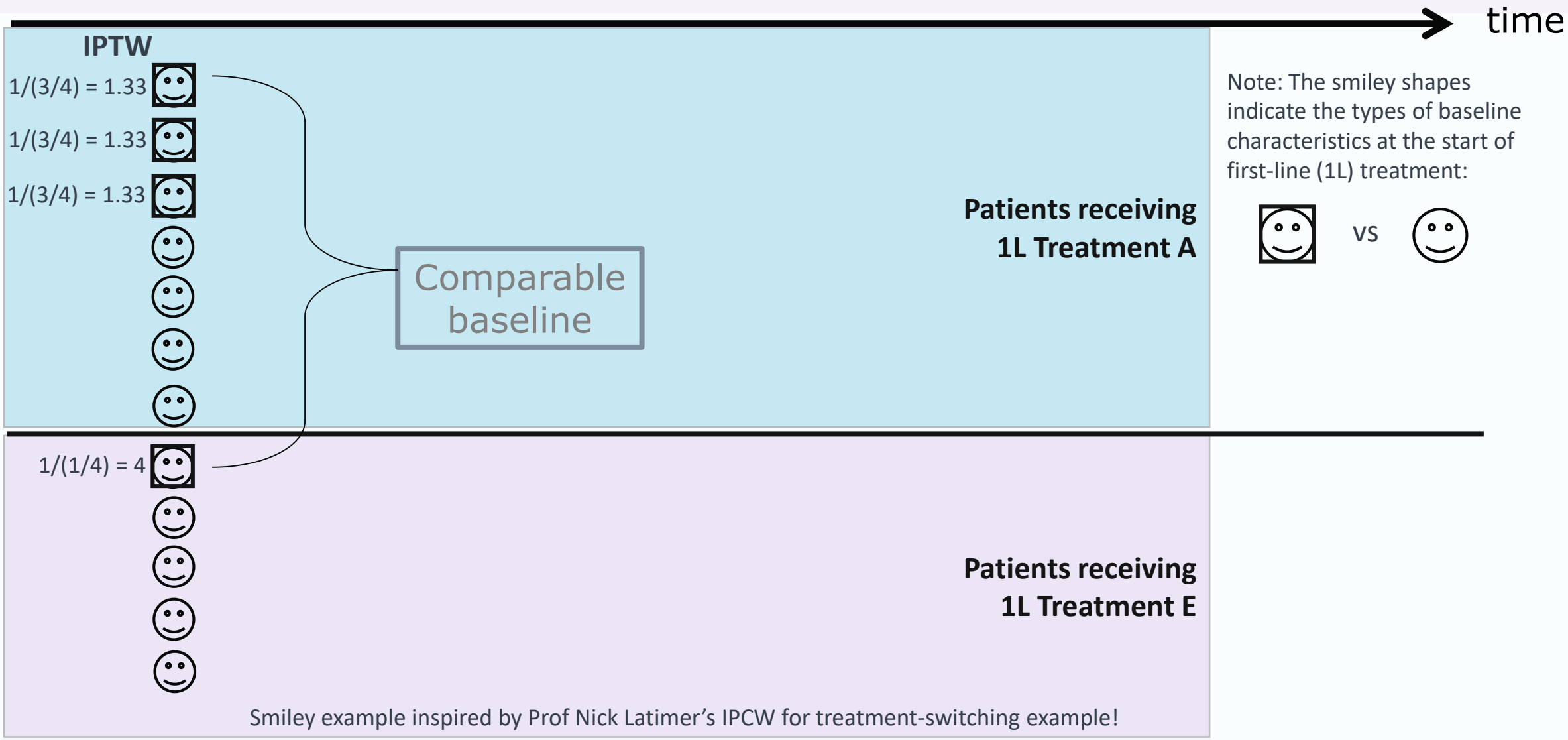
What is the causal effect on OS if all patients received A → E versus E → A?

Baseline & time-varying confounding

IPTW*IPCW$_{txdev}$ can be used to address this. $W_t^{T,D} = W^T \times W_t^D$ These weights are *multiplied* as the probability of continuing a treatment sequence within each arm is *conditioned* on receiving a specific 1L treatment. IPCW$_{txdev}$ is modelled separately using each arm's data, unlike IPTW, which uses the total population.

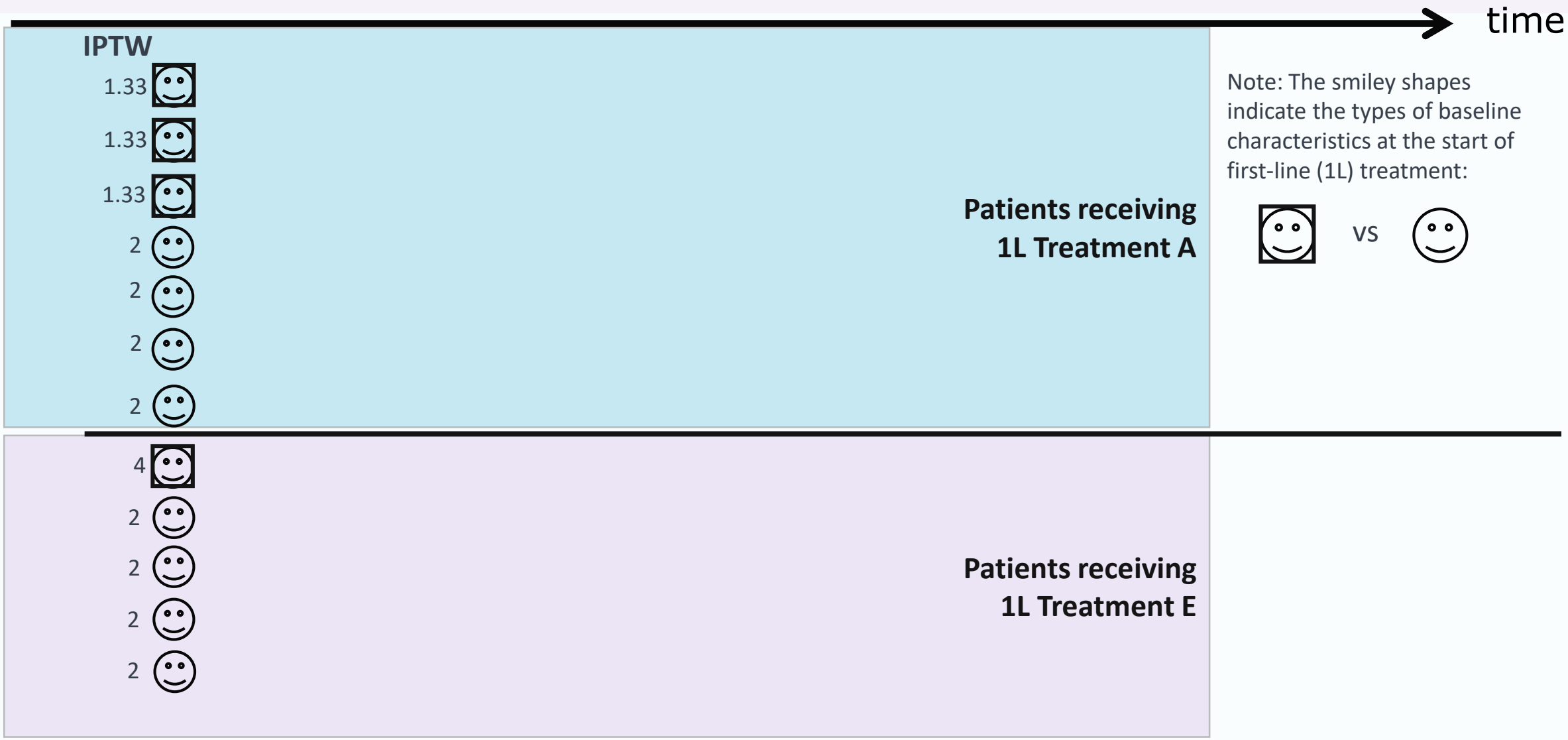
The IPTW*IPCW$_{txdev}$ method adapts IPW techniques from:
1. Robins JM & Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. Biometrics, 56(3), 779-788.
2. Huang et al. Estimation of the causal effects on survival of two-stage nonrandomized treatment sequences for recurrent diseases. Biometrics, 2006; 62(3), 901-909.
3. Hernán MA. How to estimate the effect of treatment duration on survival outcomes using observational data. BMK; 2018;360:k182

IPTW*IPCW$_{txdev}$ – the concept (over time)

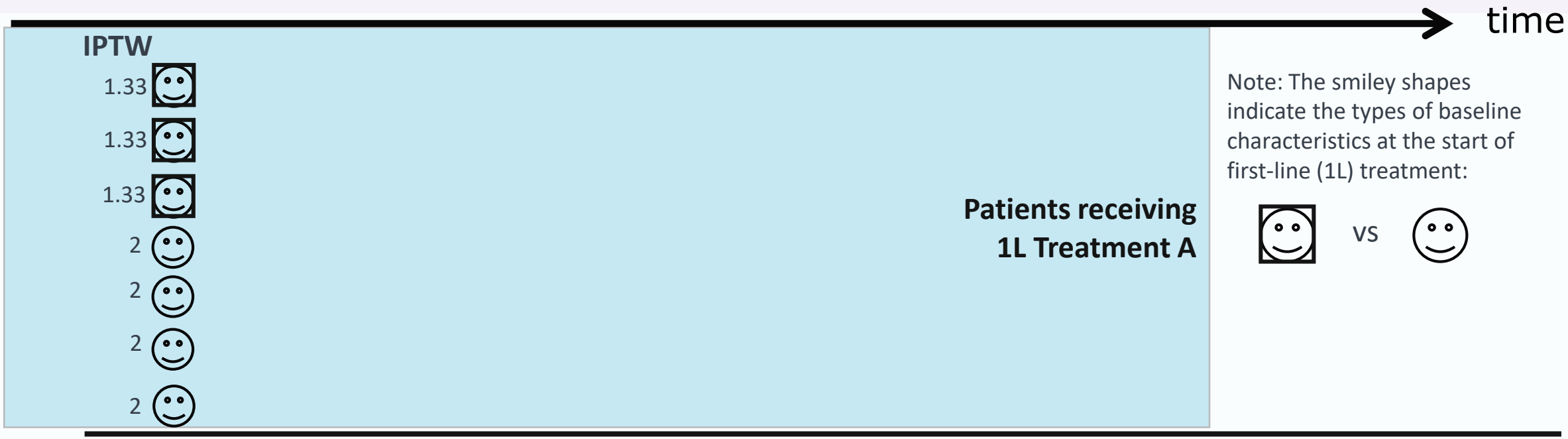


Smiley example inspired by Prof Nick Latimer's IPCW for treatment-switching example!

IPTW*IPCW$_{txdev}$ – the concept (over time)

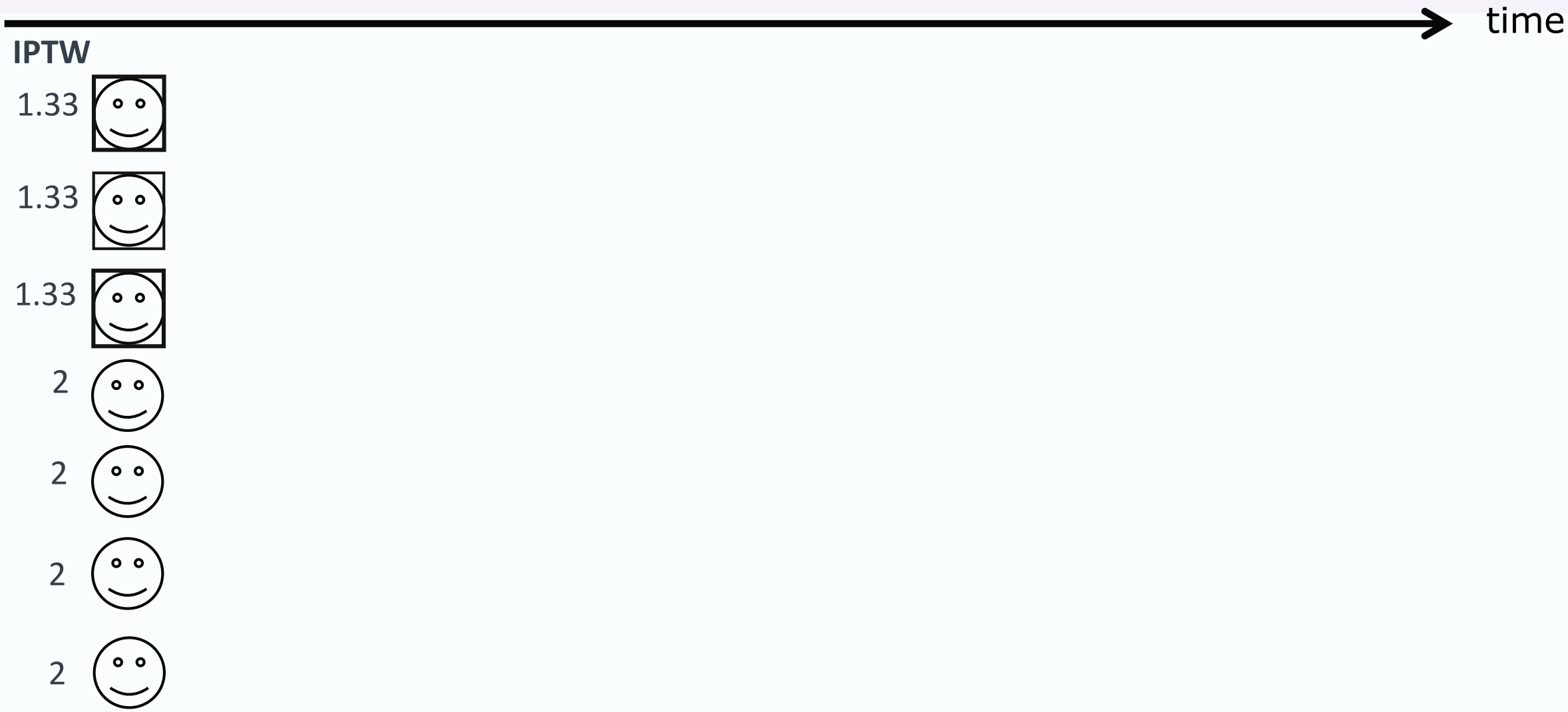


IPTW*IPCW$_{txdev}$ – the concept (over time)

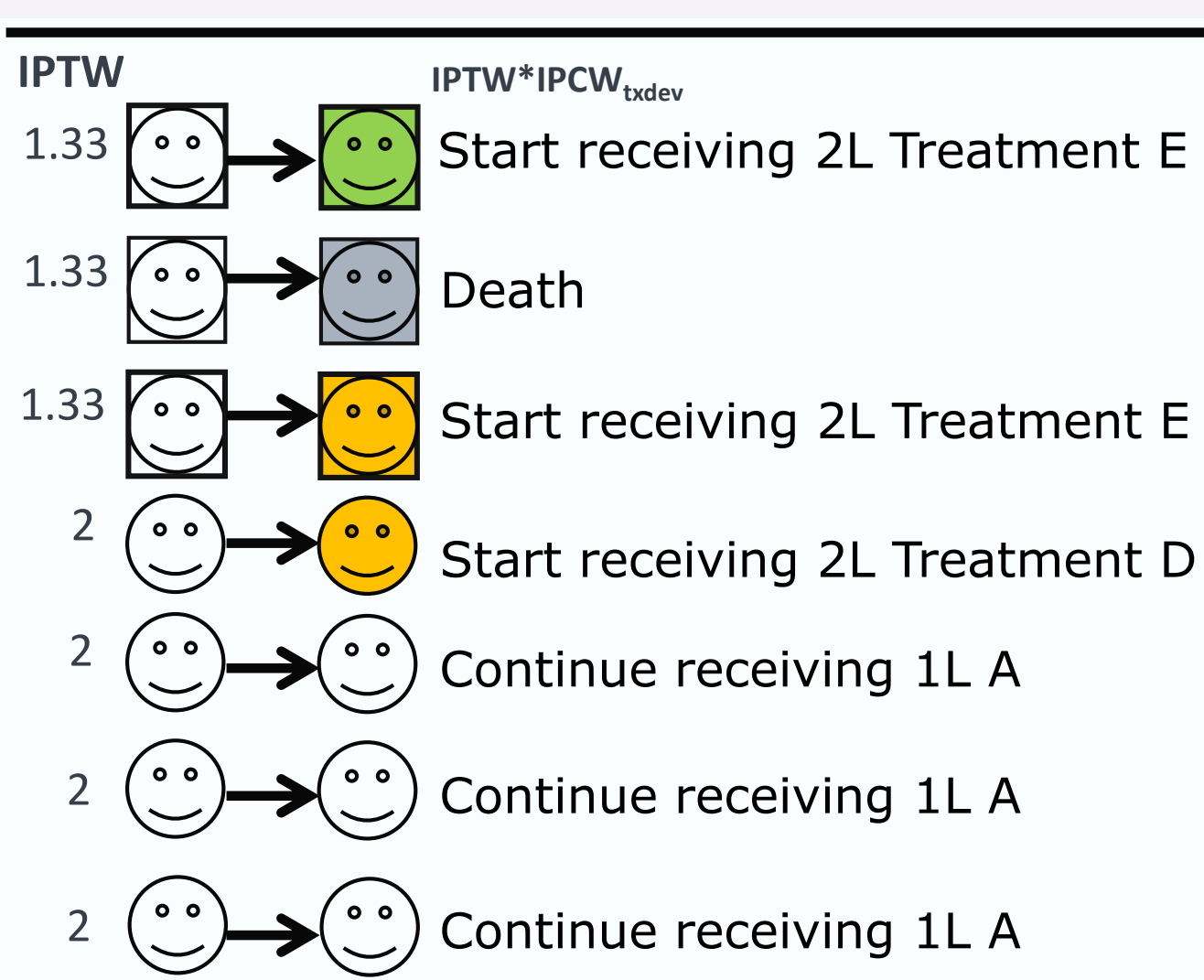


For simplicity, let's focus on implementing IPCW$_{txdev}$ for a single treatment arm.

IPTW*IPCW$_{txdev}$ – the concept (focusing on a single-arm over time)

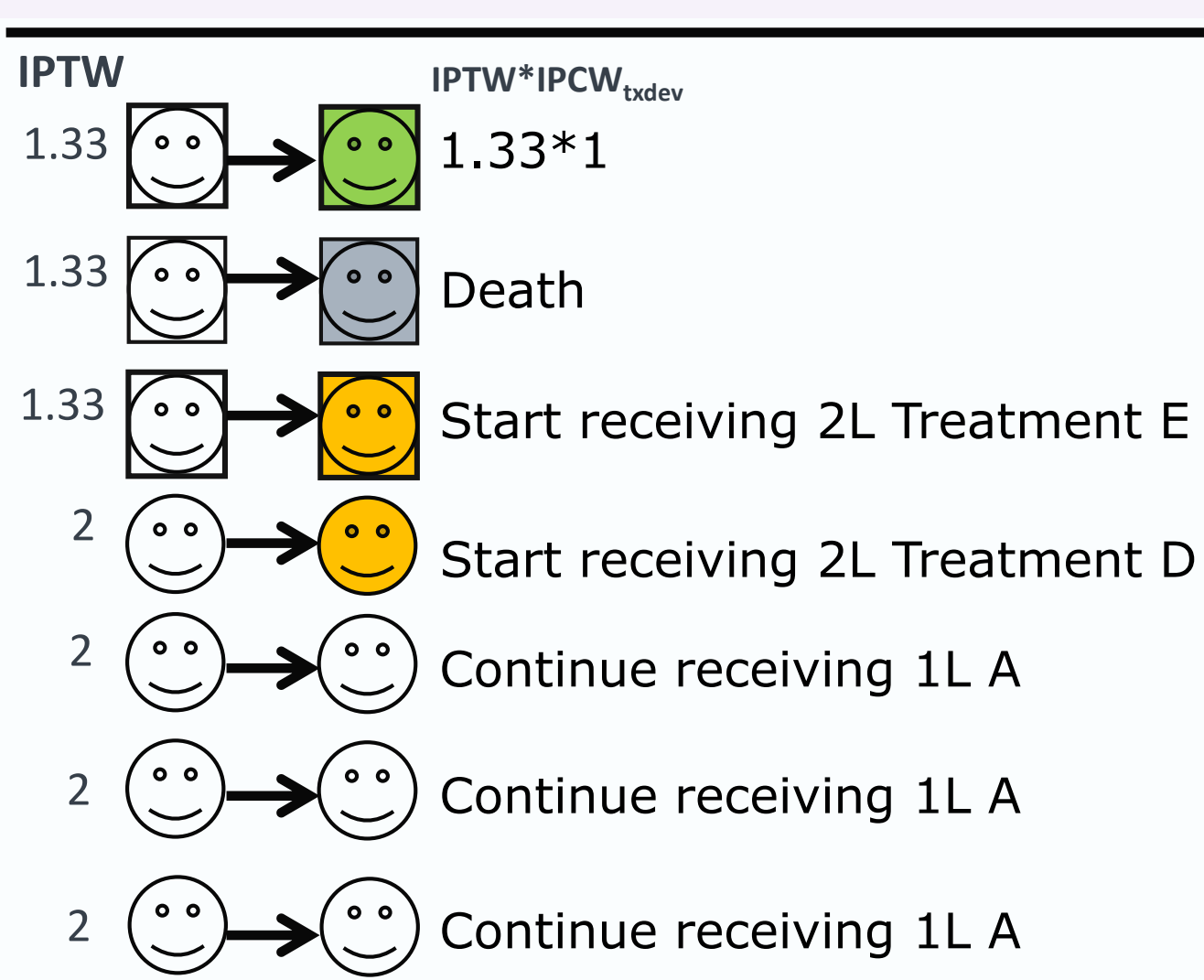


IPTW*IPCW$_{txdev}$ – the concept (focusing on a single-arm over time)



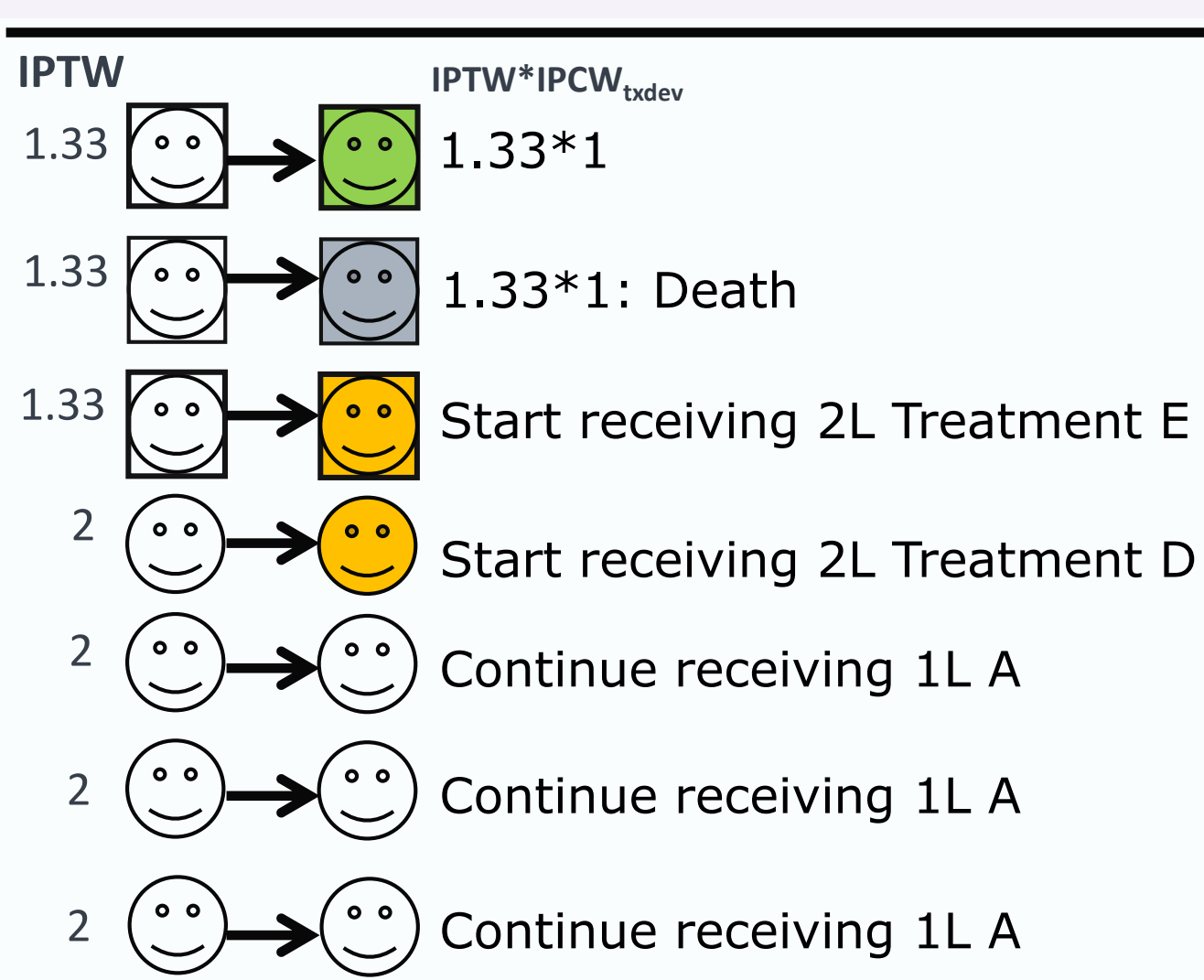
Note: The colour of the smileys indicate the types of baseline characteristics at the start of second-line (2L) treatment:
Green vs orange vs purple

IPTW*IPCW$_{txdev}$ – the concept (focusing on a single-arm over time)



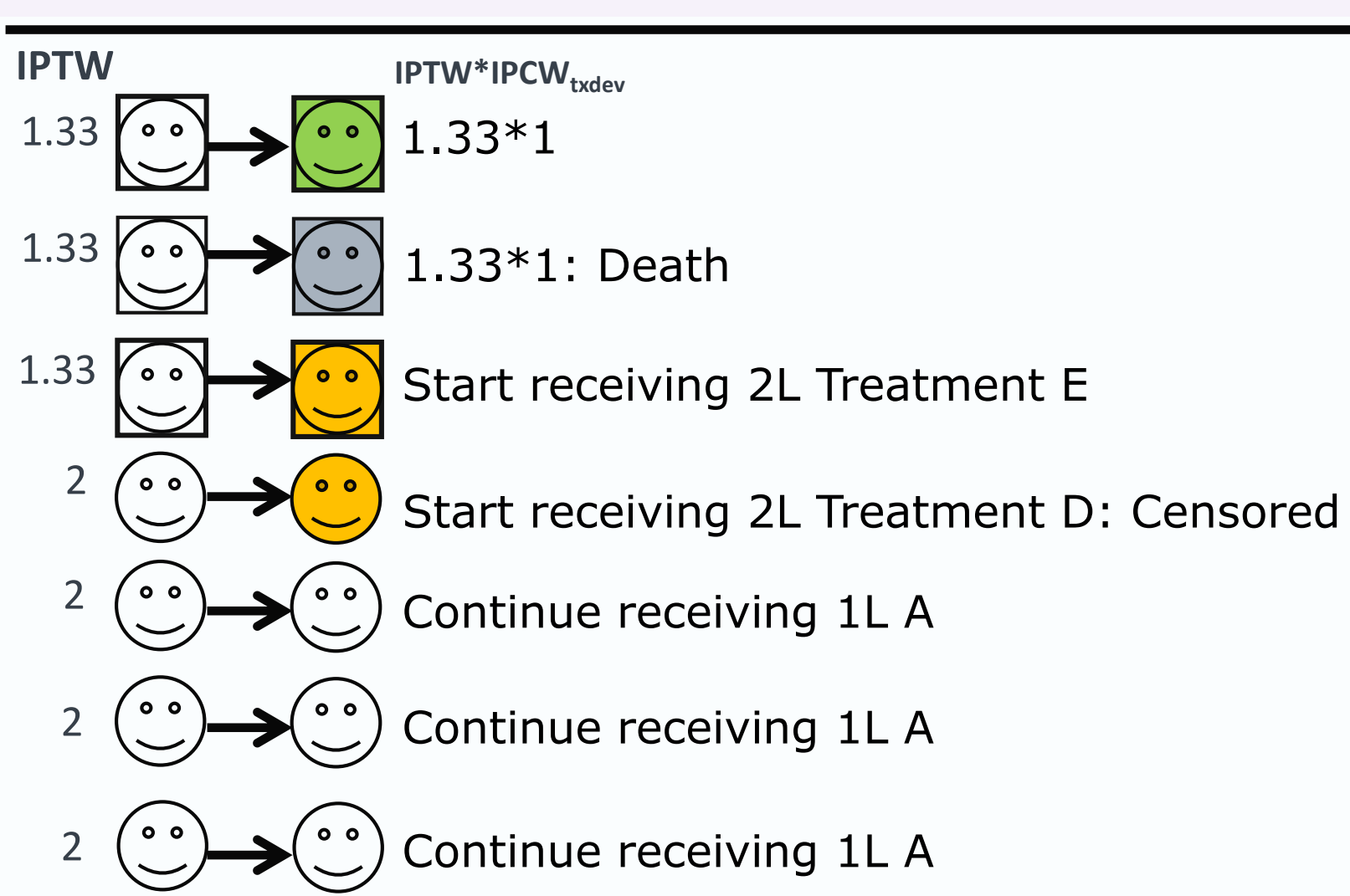
Note: The colour of the smileys indicate the types of baseline characteristics at the start of second-line (2L) treatment:
Green vs orange vs purple

IPTW*IPCW$_{txdev}$ – the concept (focusing on a single-arm over time)



Note: The colour of the smileys indicate the types of baseline characteristics at the start of second-line (2L) treatment:
Green vs orange vs purple

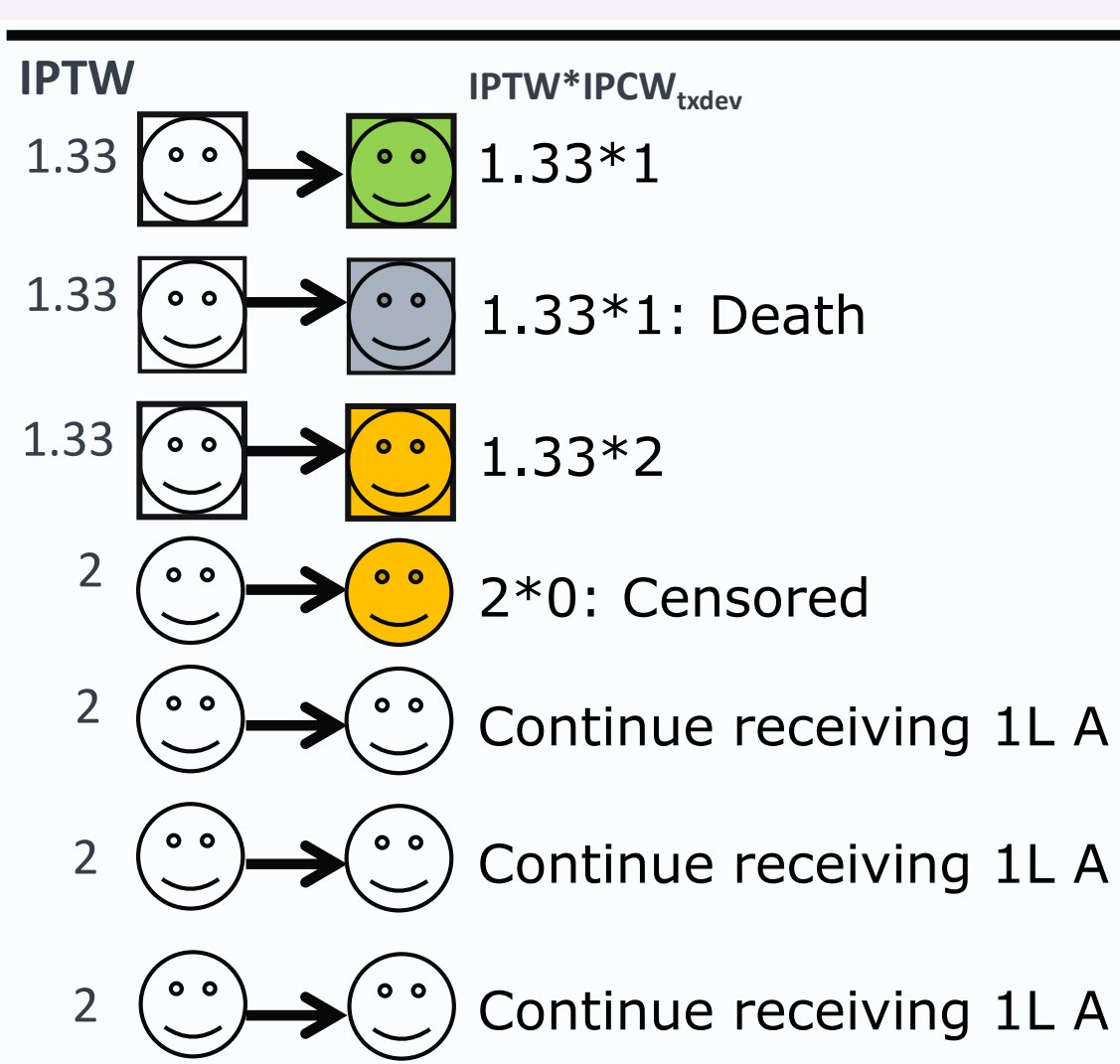
IPTW*IPCW$_{txdev}$ – the concept (focusing on a single-arm over time)



Note: The colour of the smileys indicate the types of baseline characteristics at the start of second-line (2L) treatment:
Green vs orange vs purple

Comparable 2L baseline

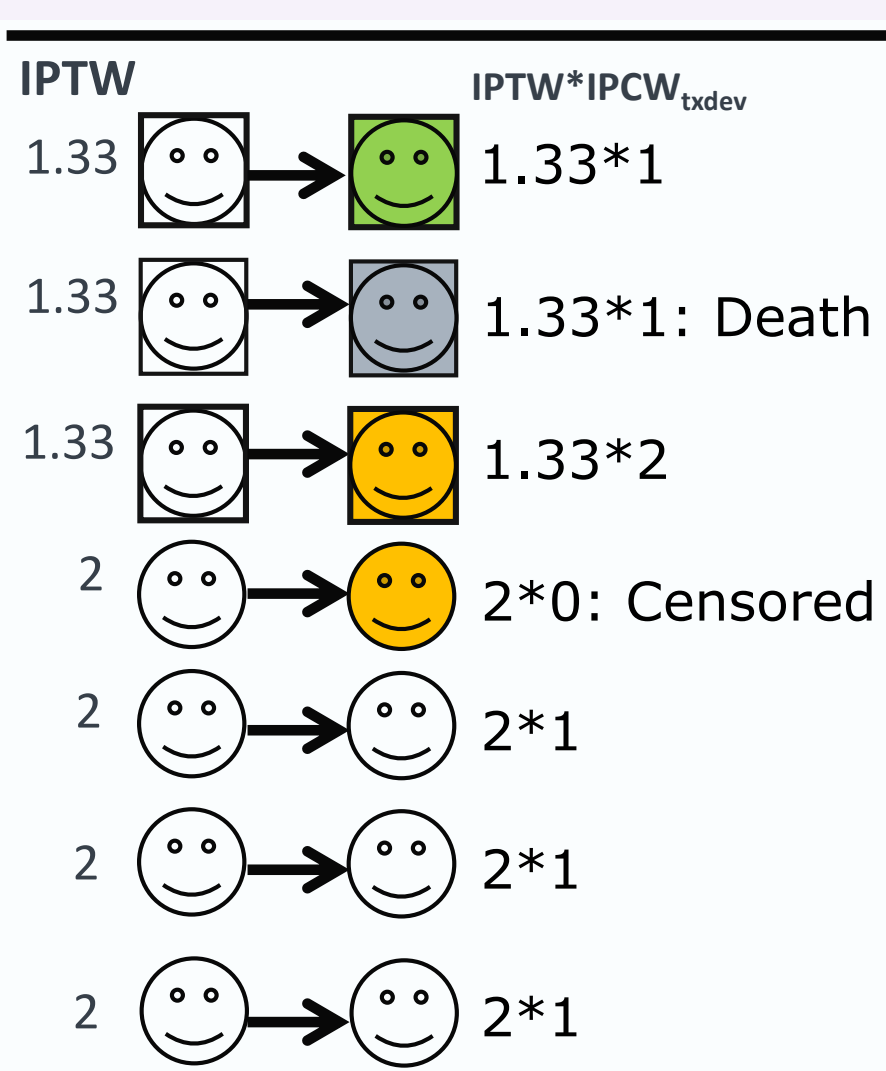
IPTW*IPCW$_{txdev}$ – the concept (focusing on a single-arm over time)



Note: The colour of the smileys indicate the types of baseline characteristics at the start of second-line (2L) treatment:
Green vs orange vs purple

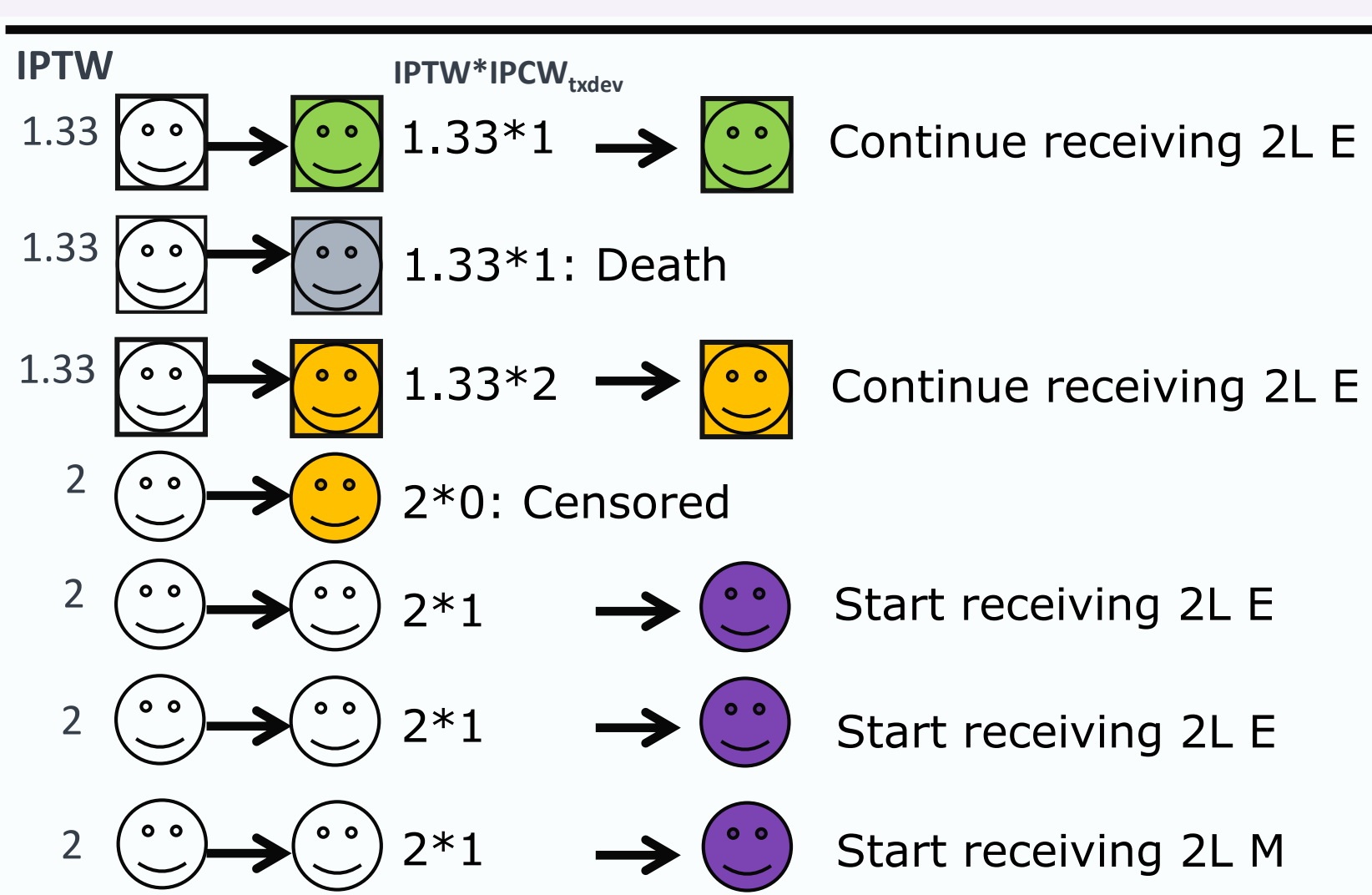
Comparable 2L baseline

IPTW*IPCW$_{txdev}$ – the concept (focusing on a single-arm over time)



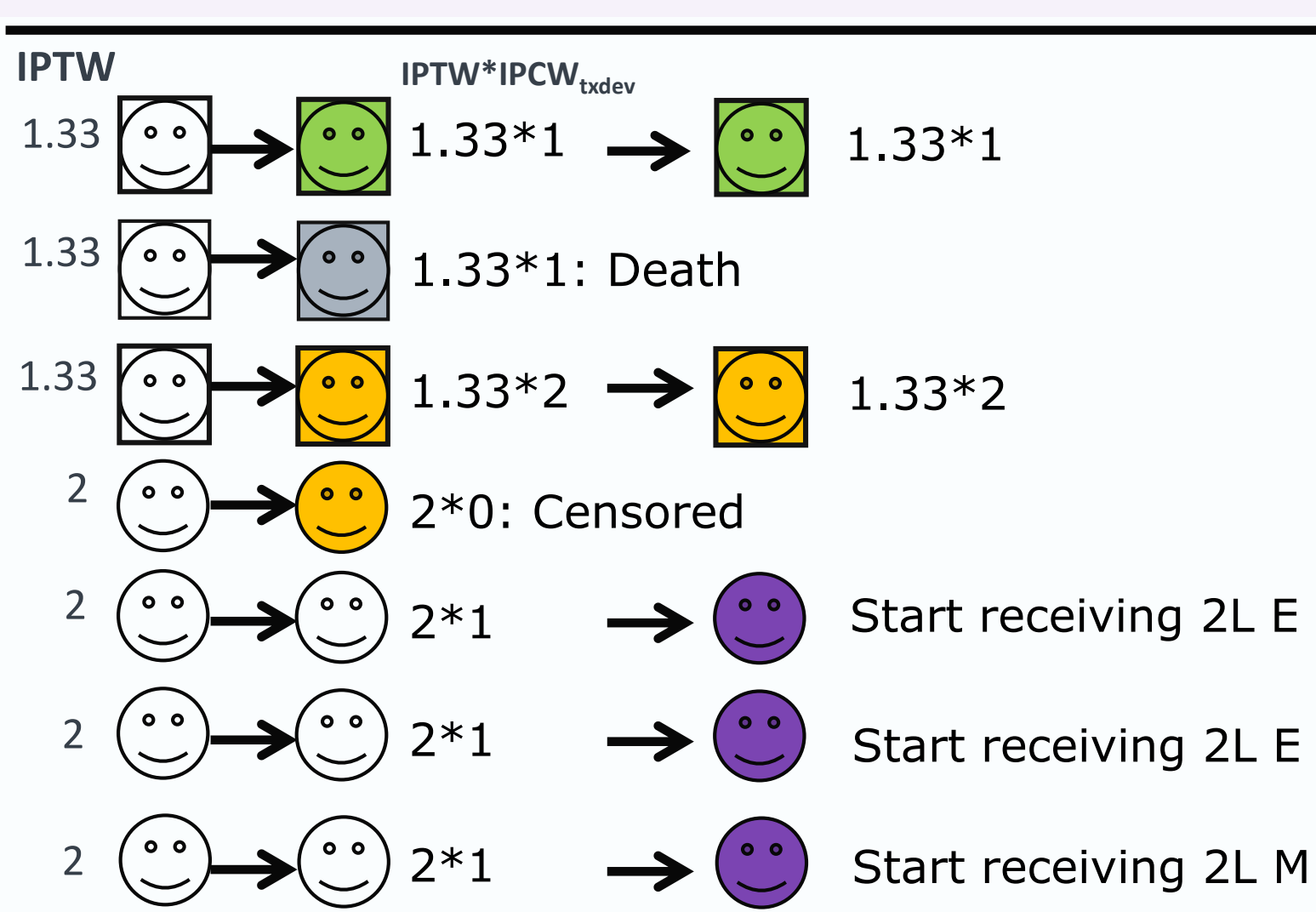
Note: The colour of the smileys indicate the types of baseline characteristics at the start of second-line (2L) treatment:
Green vs orange vs purple

IPTW*IPCW$_{txdev}$ – the concept (focusing on a single-arm over time)



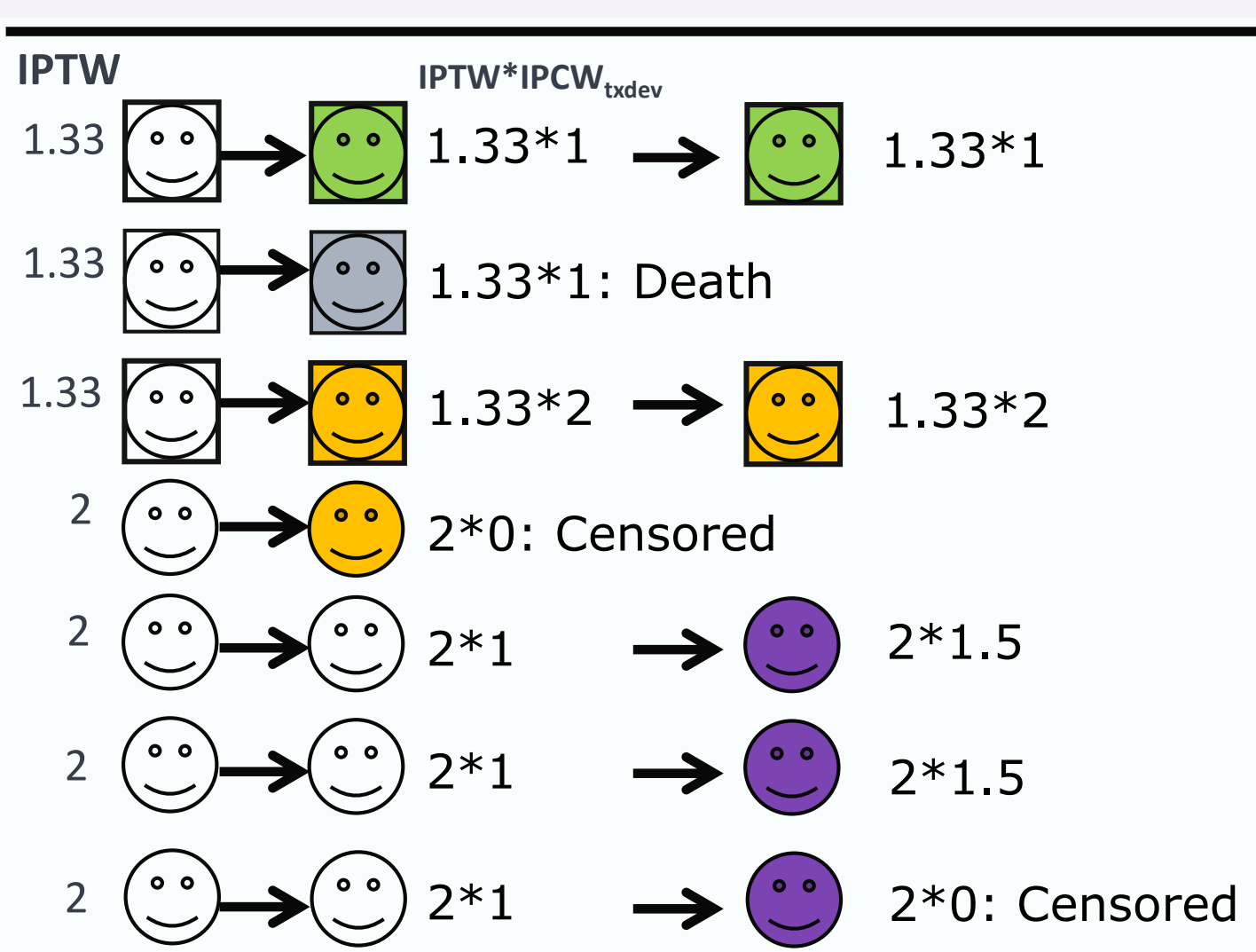
Note: The colour of the smileys indicate the types of baseline characteristics at the start of second-line (2L) treatment:
Green vs orange vs purple

IPTW*IPCW$_{txdev}$ – the concept (focusing on a single-arm over time)



Note: The colour of the smileys indicate the types of baseline characteristics at the start of second-line (2L) treatment:
Green vs orange vs purple

IPTW*IPCW$_{txdev}$ – the concept (focusing on a single-arm over time)

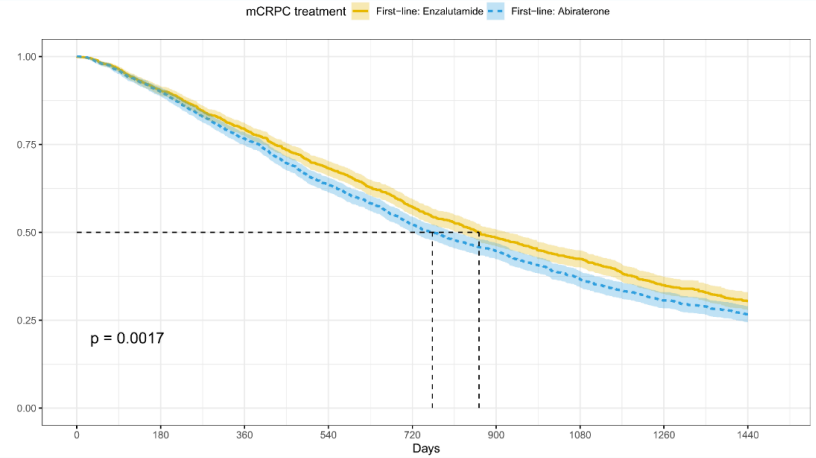


Note: The colour of the smileys indicate the types of baseline characteristics at the start of second-line (2L) treatment:
Green vs orange vs purple

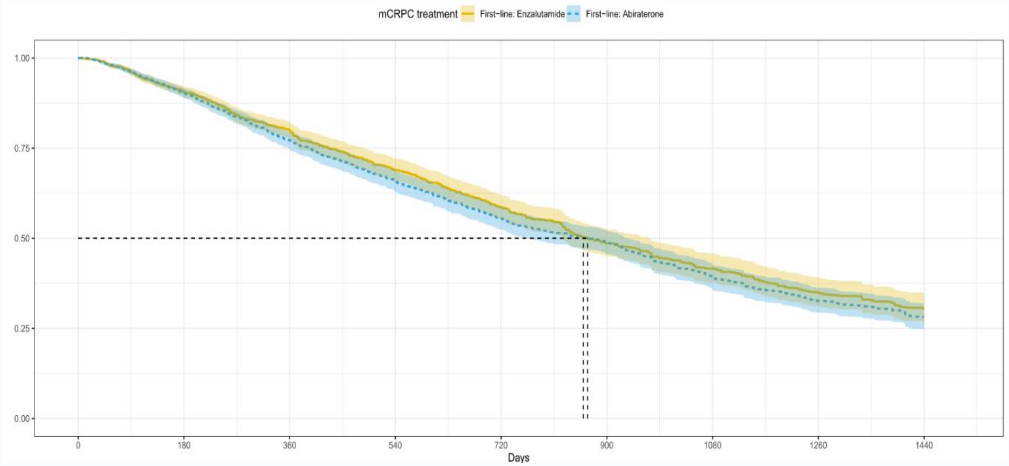
Comparable 2L baseline

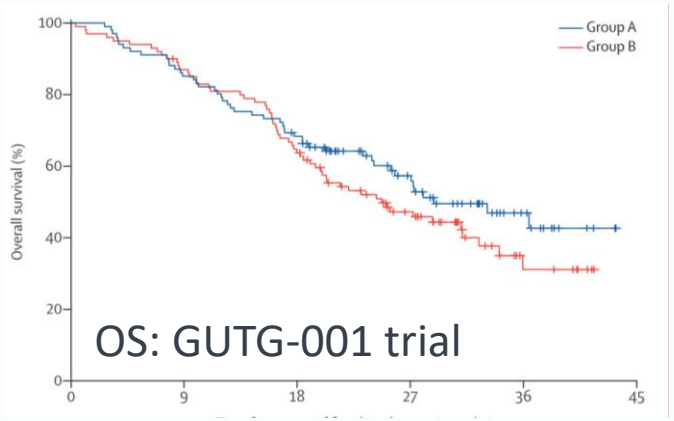
Emulated GUTG-001 Analogue Target Trial using IPTW*IPCW$_{txdev}$

Unadjusted comparison: OS



IPTW*IPCW$_{txdev}$ adjusted comparison: OS



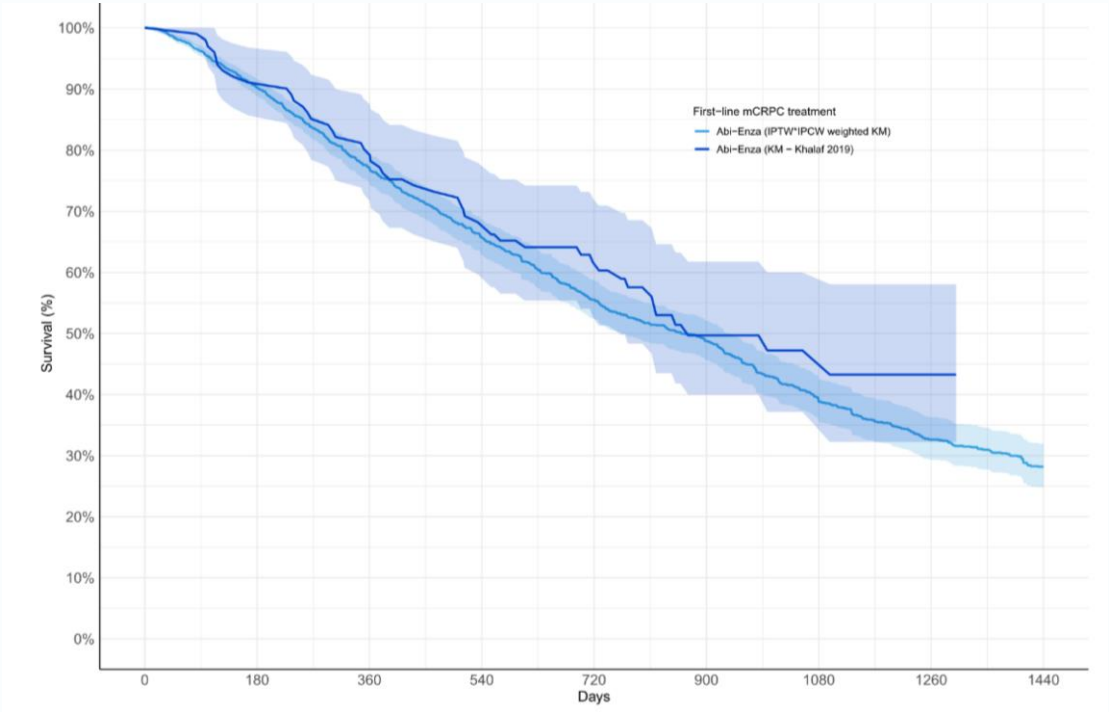


-Journal publication in progress-

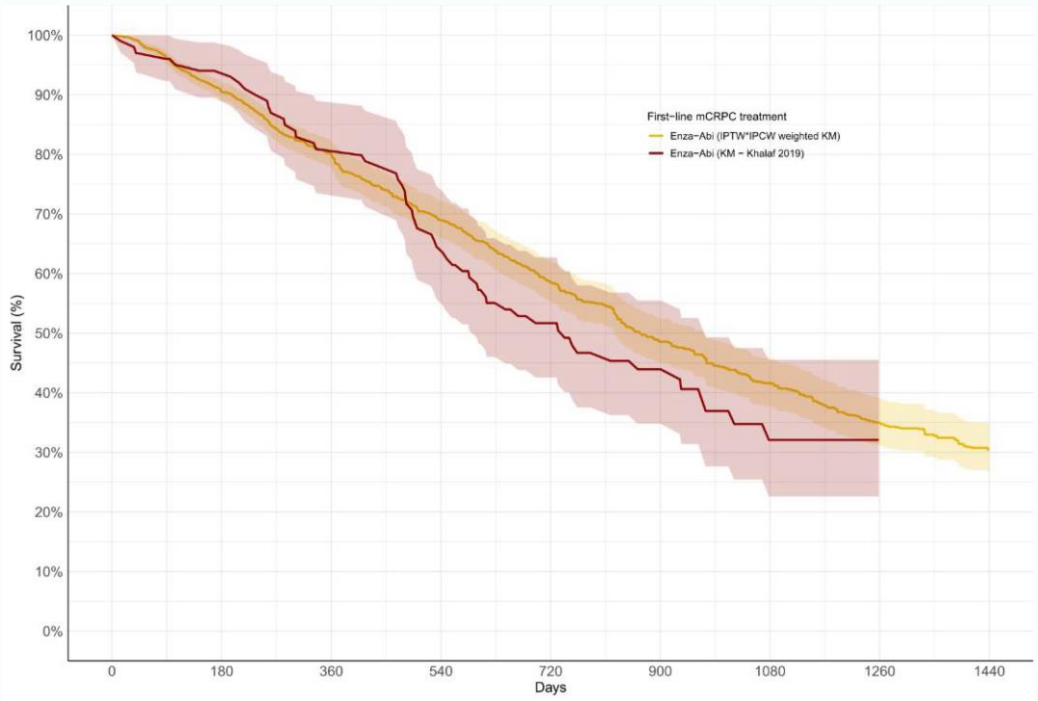
Ref: Chang JYA. Investigating the Application of Causal Inference Methods for Modelling the Impact of Treatment Sequences in Health Economic Evaluations 2024. (PhD Thesis, University of Sheffield).
Khalaf DJ, et al. Optimal sequencing of enzalutamide and abiraterone acetate plus prednisone in metastatic castration-resistant prostate cancer: a multicentre, randomised, open-label, phase 2, crossover trial. *The Lancet Oncology*, 2019; *20*(12), 1730-1739.

RCT-RWE agreement assessment:
GUTG-001 trial vs. emulated GUTG-001 Analogue Target Trial using IPTW*IPCW$_{txdev}$

Abiraterone → Enzalutamide



Enzalutamide → Abiraterone

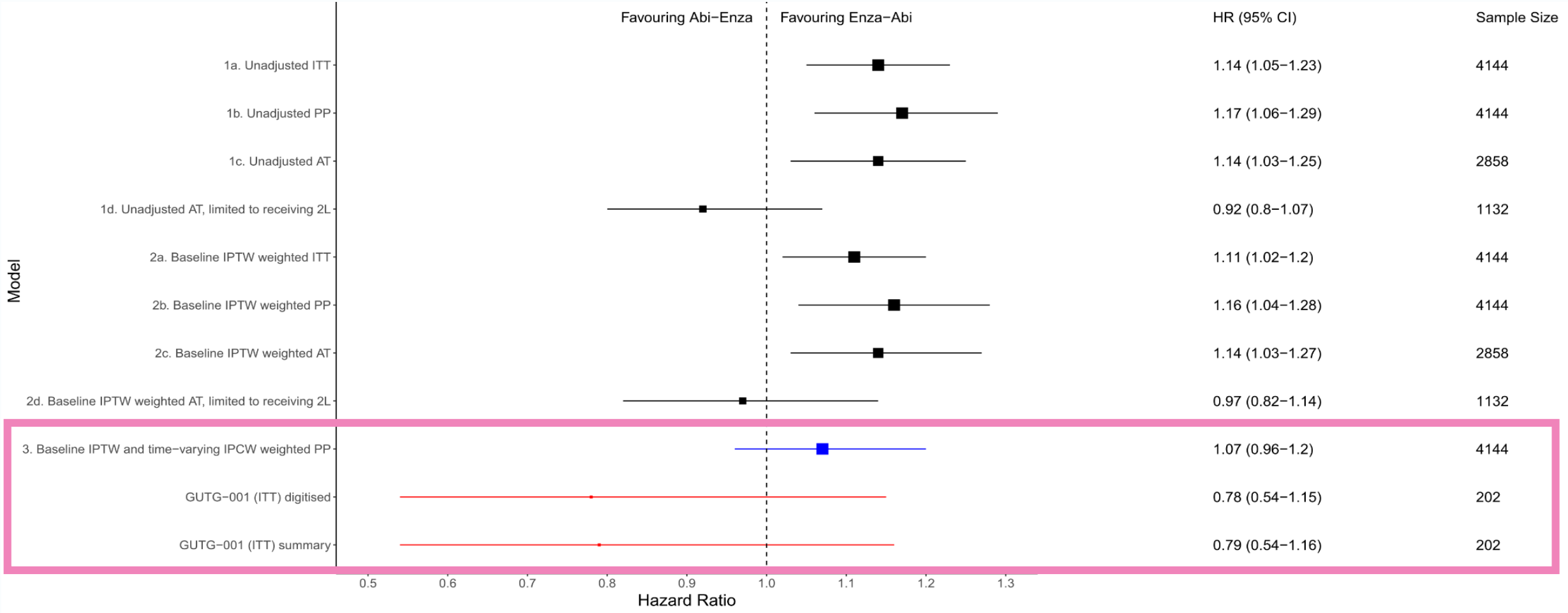


-Journal publication in progress-

Ref: Chang JYA. Investigating the Application of Causal Inference Methods for Modelling the Impact of Treatment Sequences in Health Economic Evaluations 2024. (PhD Thesis, University of Sheffield).

RCT-RWE agreement assessment: Hazard Ratios

GUTG-001 trial vs. emulated GUTG-001 Analogue Target Trial using IPTW*IPCW$_{txdev}$



-Journal publication in progress-

Ref: Chang JYA. Investigating the Application of Causal Inference Methods for Modelling the Impact of Treatment Sequences in Health Economic Evaluations 2024. (PhD Thesis, University of Sheffield).

RCT-RWE agreement assessment with extended matrix

Criteria	Findings
1. Regulator agreement	**Agreed** The GUTG-001 trial shows no significant HR difference between treatment sequences (0.79, 0.54-1.16), consistent with the our emulation (1.07, 0.96-1.20).
2. Estimate agreement	**Agreed** - The HR point estimate from the emulation (1.07) falls within the GUTG-001 trial's 95% CI (0.54-1.16). - The median OS estimates for both treatment sequences fall within the GUTG-001 trial's 95% CIs: abiraterone → enzalutamide is 28.7 (trial CI: 28.8–not reached), and enzalutamide → abiraterone is 28.9 (trial CI: 18.8–34.0).
3. Exploratory – standardised difference	**Agreed** Z = -1.48 (for HR from final emulation versus GUTG-001), indicating no significant difference (i.e., < 1.96) between estimates from RWE and RCT.
4. Exploratory – survival curve comparison	**Largely aligned** The survival point estimate from the emulation mostly fell within the GUTG-001 trial's 95% CIs, except during the first 3 months in the abiraterone → enzalutamide group.

Please see our poster on Nov 18, 2024 afternoon: "MSR64: Enhanced Randomised Controlled Trials-Real-World Evidence (RCT-RWE) Agreement Assessment Metrics for Health Technology Assessment (HTA)"

-Journal publication in progress-

Ref: Chang JYA. Investigating the Application of Causal Inference Methods for Modelling the Impact of Treatment Sequences in Health Economic Evaluations 2024. (PhD Thesis, University of Sheffield).

Summary & future research

- The IPTW*IPCW$_{txdev}$ method with Target Trial Emulation effectively estimates comparative treatment sequence effectiveness in Flatiron data, benchmarked against an existing RCT.
- Limitations:
  - Currently no adjustment for additional treatment lines
  - Potential unmeasured confounders and positivity assumption violations
- Future research
  - Extended studies in Flatiron and English Cancer Registry
  - Explore alternative methods (e.g. g-formula)
  - Streamline survival extrapolation with parametric survival models using adjusted OS.
  - Simulation studies for further validation
  - Explore adaption for external control arms

Post-workshop discussion

Post-workshop polling

Should we be estimating the effectiveness/cost-effectiveness of different sequences?

- Yes
- No
- Depending on the scenario

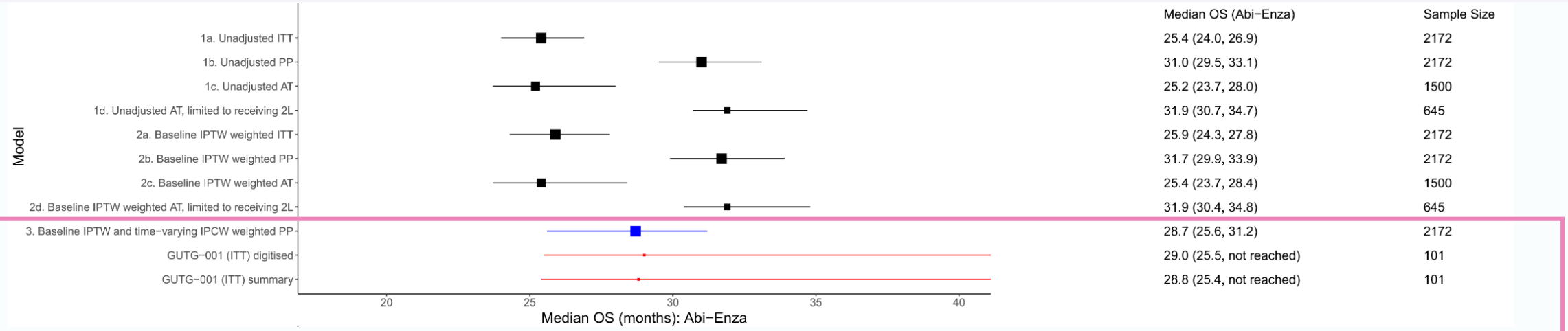
Discussion

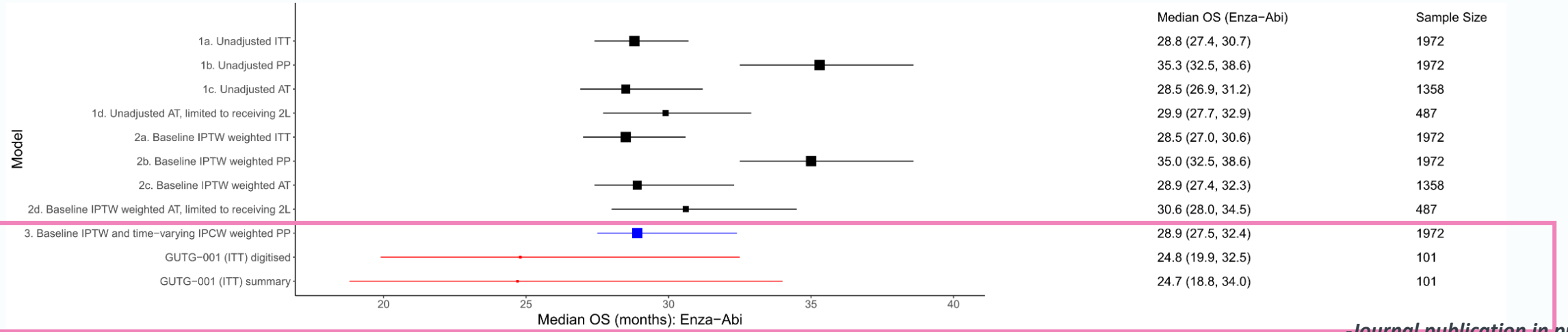
- Should we be evaluating treatment sequences in HTA?
- What do you think is the best approach for doing this?
- Or should we stick to just evaluating lines of therapy in isolation?

Appendix

RCT-RWE agreement assessment: Median Overall Survival

GUTG-001 trial vs. emulated GUTG-001 Analogue Target Trial using IPTW*IPCW$_{txdev}$





-Journal publication in progress-

Ref: Chang JYA. Investigating the Application of Causal Inference Methods for Modelling the Impact of Treatment Sequences in Health Economic Evaluations 2024. (PhD Thesis, University of Sheffield).