# Efficient, doubly robust estimation of the effect of dose switching for switchers in a randomised clinical trial

UNIVERSITEIT GENT

Kelly Van Lancker

joint work with Stijn Vansteelandt and An Vandebosch

janssen

# Background

- Phase 3 program of a new experimental compound for patients with a chronic condition
- Consisted of multiple studies, including

**FIXED DOSING TRIAL ($T = 0$)**

HIGH DOSE ($D = h$)

LOW DOSE ($D = l$)

(FIXED) PLACEBO

**FLEXIBLE DOSING TRIAL ($T = 1$)**

FLEXIBLE DOSE ($D = f$)

(FLEX) PLACEBO

# Background

- Phase 3 program of a new experimental compound for patients with a chronic condition
- Consisted of multiple studies, including

**FIXED DOSING TRIAL ($T = 0$)**

HIGH DOSE ($D = h$)

LOW DOSE ($D = l$)

(FIXED) PLACEBO

**FLEXIBLE DOSING TRIAL ($T = 1$)**

FLEXIBLE DOSE ($D = f$)

(FLEX) PLACEBO

- Examples: **neuroscience** (e.g. Invega, Spravato, . . . )

# Background

- Phase 3 program of a new experimental compound for patients with a chronic condition
- Consisted of multiple studies, including

**FIXED DOSING TRIAL ($T = 0$)**

HIGH DOSE ($D = h$)

LOW DOSE ($D = l$)

(FIXED) PLACEBO

**FLEXIBLE DOSING TRIAL ($T = 1$)**

FLEXIBLE DOSE ($D = f$)

(FLEX) PLACEBO

- Examples: **neuroscience** (e.g. Invega, Spravato, . . . )
- Why?
  - Need 2 positive efficacy trials for approval
  - Fixed: dose-response for efficacy evaluation
  - Flexible: presumed dosing strategy in clinical practice

# Study Results

- Protocols are similar in
  - Target patient population (I/E criteria)
  - Primary endpoint
  - Treatment effect measure for primary endpoint

# Study Results

- Protocols are similar in
  - Target patient population (I/E criteria)
  - Primary endpoint
  - Treatment effect measure for primary endpoint
- Flexible dosing trial:
  - Flexible dose significantly better than placebo

# Study Results

- Protocols are similar in
  - Target patient population (I/E criteria)
  - Primary endpoint
  - Treatment effect measure for primary endpoint
- Flexible dosing trial:
  - Flexible dose significantly better than placebo
- Fixed dosing trial:
  - Treatment effect observed at both doses
  - No dose-response observed

# Study Results

- Protocols are similar in
  - Target patient population (I/E criteria)
  - Primary endpoint
  - Treatment effect measure for primary endpoint
- Flexible dosing trial:
  - Flexible dose significantly better than placebo
- Fixed dosing trial:
  - Treatment effect observed at both doses
  - No dose-response observed

## Research Question

*Is* **flexible dosing potentially beneficial** *(in terms of treatment effect* **compared to the low dose***) for* **switchers** *in the treatment arm of the flexible dosing study?*

# Problem Setting

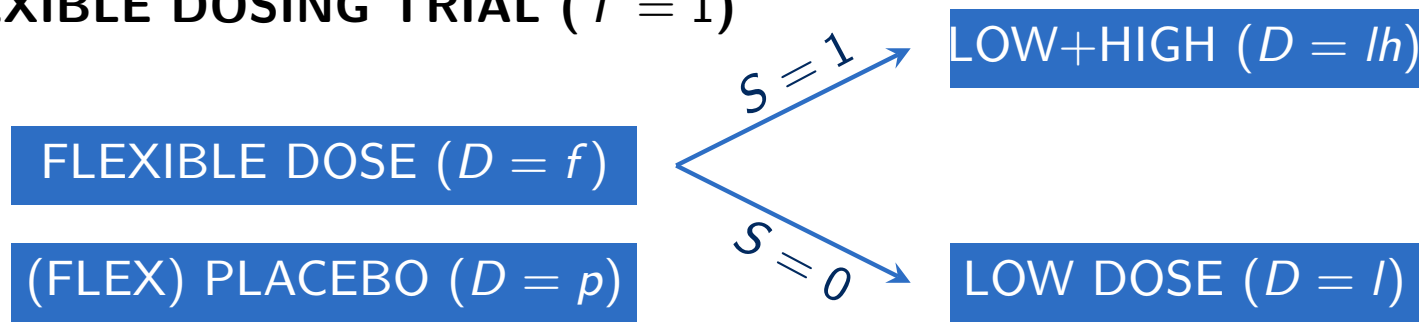
**FLEXIBLE DOSING TRIAL ($T = 1$)**

FLEXIBLE DOSE ($D = f$)

(FLEX) PLACEBO ($D = p$)

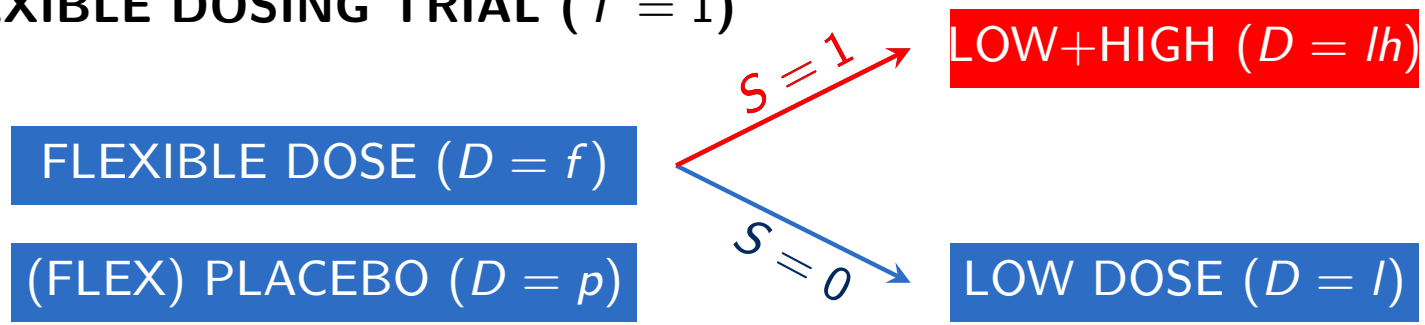
# Problem Setting

**FLEXIBLE DOSING TRIAL ($T = 1$)**

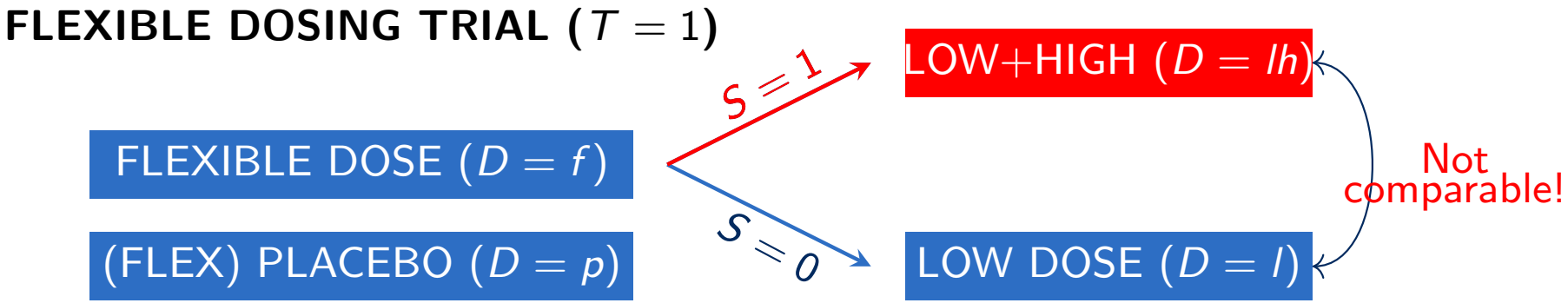


# Problem Setting

**FLEXIBLE DOSING TRIAL ($T = 1$)**

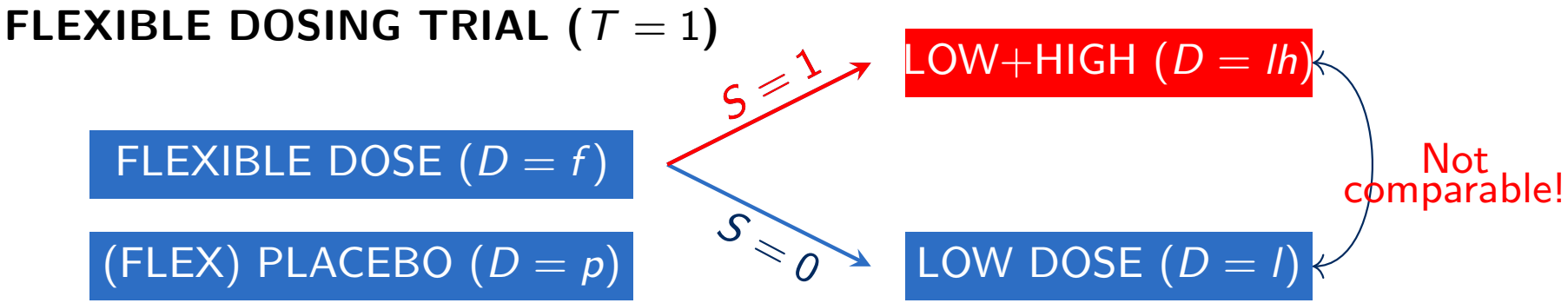


# Problem Setting



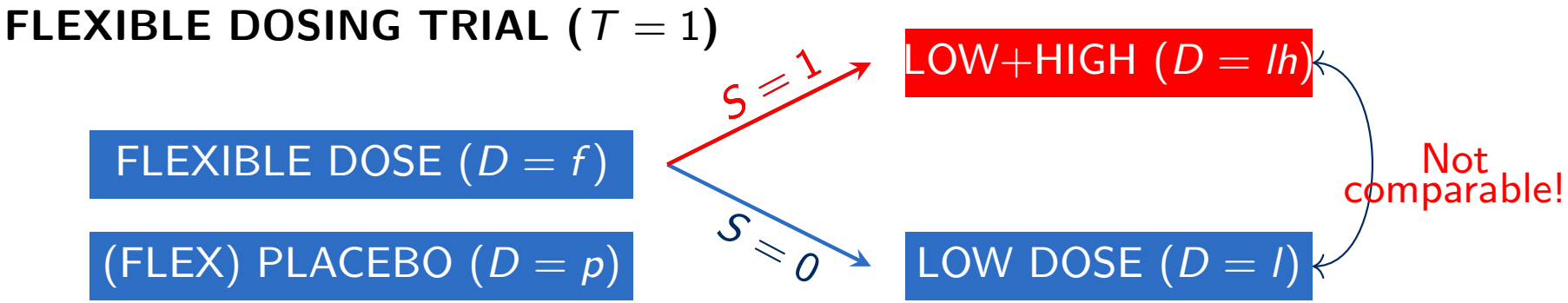

- Comparing directly with those who stayed on low dose does not entail a satisfactory evaluation
  - e.g., latter patients usually in a better health condition

# Problem Setting



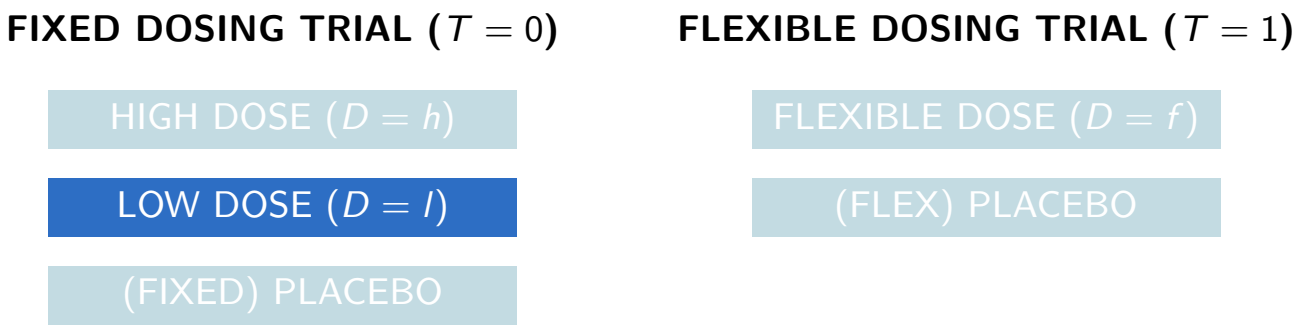

- Comparing directly with those who stayed on low dose does not entail a satisfactory evaluation
  - e.g., latter patients usually in a better health condition
- Deterministic rule for switching complicates inverse probability weighting: positivity violation

# Problem Setting



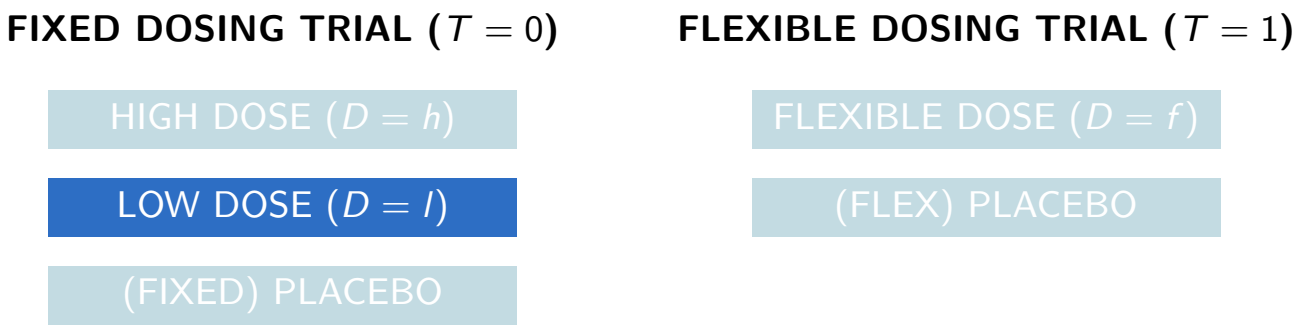

- Comparing directly with those who stayed on low dose does not entail a satisfactory evaluation
  - e.g., latter patients usually in a better health condition
- Deterministic rule for switching complicates inverse probability weighting: positivity violation
- Available information too scarce: no arm assigned to fixed low dose

# Problem Setting: Possible Solution




- Possible Solution
  - Can we employ data from the fixed dosing trial (i.e., low dose)?
  - Possibly correcting for imbalances between trials?

# Problem Setting: Possible Solution




- Possible Solution
  - Can we employ data from the fixed dosing trial (i.e., low dose)?
  - Possibly correcting for imbalances between trials?
- Transport data from low dose arm of fixed dosing trial
  - using similar techniques as for transporting inferences from trial participants to new target population

# Question Formalized - Estimand

## Goal

*In flexible dosing study, for those who required switching:*

*How different would the average response $Y$ have been for them, had they not switched:*

$$E[Y^{lh} - Y^{l} | T = 1, D = f, S = 1]$$

# Question Formalized - Estimand

## Goal

*In flexible dosing study, for those who required switching:*

*How different would the average response $Y$ have been for them, had they not switched:*

$$E[Y^{lh} - Y^{l} | T = 1, D = f, S = 1]$$

$\Downarrow$ Some calculations

$$E[Y^{\mathrm{f}} - Y^{l} | T = 1, D = f] / P(S = 1 | T = 1, D = f)$$

# Proposed Estimator

## Goal

*How different would the average response $Y$ have been for the switchers, had they not switched:*

$$E[Y^f - Y^l | T = 1, D = f] / P(S = 1 | T = 1, D = f)$$

# Proposed Estimator

## Goal

*How different would the average response $Y$ have been for the switchers, had they not switched:*

$$E[Y^f - Y^l | T = 1, D = f] / \boldsymbol{P(S = 1 | T = 1, D = f)}$$

# Proposed Estimator

## Goal

*How different would the average response $Y$ have been for the switchers, had they not switched:*

$$E[Y^f - Y^l | T = 1, D = f] / \boldsymbol{P(S = 1 | T = 1, D = f)}$$

- $P(S = 1 | T = 1, D = f)$: estimated as **proportion of switchers in flexible dosing arm**

# Proposed Estimator

## Goal

*How different would the average response $Y$ have been for the switchers, had they not switched:*

$$\boldsymbol{E[Y^f - Y^l | T = 1, D = f]} / P(S = 1 | T = 1, D = f)$$

- $P(S = 1 | T = 1, D = f)$: estimated as **proportion of switchers in flexible dosing arm**
- Note: expectation is an ATT effect!

# Proposed Estimator

$$\boldsymbol{E[Y^f - Y^l | T = 1, D = f]} / P(S = 1 | T = 1, D = f)$$

# Proposed Estimator

$$\boldsymbol{E[Y^f - Y^l | T = 1, D = f]} / P(S = 1 | T = 1, D = f)$$

**FIXED DOSING TRIAL ($T = 0$)**

- HIGH DOSE ($D = h$)
- LOW DOSE ($D = l$)
- (FIXED) PLACEBO

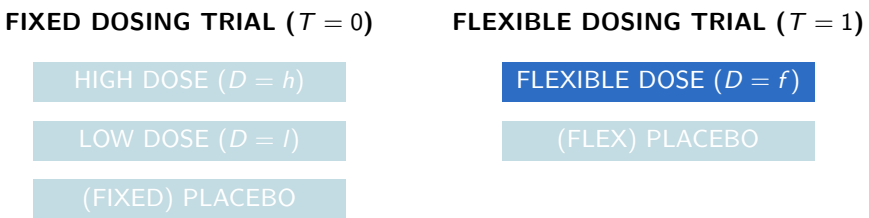
**FLEXIBLE DOSING TRIAL ($T = 1$)**

- FLEXIBLE DOSE ($D = f$)
- (FLEX) PLACEBO

An efficient estimator can be obtained by

# Proposed Estimator

$$E[Y^f - Y^l | T = 1, D = f] / P(S = 1 | T = 1, D = f)$$

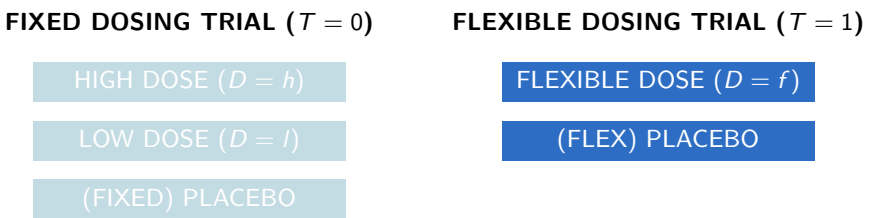



An efficient estimator can be obtained by

1. Fitting a regression model for $Y$ given baseline covariates $\boldsymbol{X}$ among the patients on the flexible dose.

# Proposed Estimator

$$E[Y^f - Y^l | T = 1, D = f] / P(S = 1 | T = 1, D = f)$$

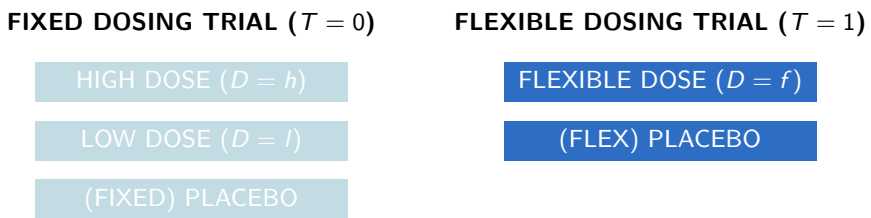



An efficient estimator can be obtained by

1. Fitting a regression model for $Y$ given baseline covariates $\boldsymbol{X}$ among the patients on the flexible dose.
2. Predicting the outcomes for all patients in flexible dosing trial.

# Proposed Estimator

$$E[Y^f - Y^l | T = 1, D = f] / P(S = 1 | T = 1, D = f)$$




An efficient estimator can be obtained by

1. Fitting a regression model for $Y$ given baseline covariates $\boldsymbol{X}$ among the patients on the flexible dose.
2. Predicting the outcomes for all patients in flexible dosing trial.
3. Taking the average of the predicted values over all patients in flexible dosing trial.

# Proposed Estimator

$$\boldsymbol{E[Y^f - Y^l | T = 1, D = f]} / P(S = 1 | T = 1, D = f)$$

- Cannot be directly estimated from flexible dosing trial

[1] where differences in mean potential outcomes can be explained by imbalances across studies in the vector of baseline covariates

# Proposed Estimator

$$E[Y^f - Y^l | T = 1, D = f] / P(S = 1 | T = 1, D = f)$$

- Cannot be directly estimated from flexible dosing trial
- Transport data from the fixed dosing arm - correcting for imbalances between the studies

[1] where differences in mean potential outcomes can be explained by imbalances across studies in the vector of baseline covariates

# Proposed Estimator

$$E[Y^f - Y^l | T = 1, D = f] / P(S = 1 | T = 1, D = f)$$

- Cannot be directly estimated from flexible dosing trial
- Transport data from the fixed dosing arm - correcting for imbalances between the studies
- Assumptions to transport inferences

[1] where differences in mean potential outcomes can be explained by imbalances across studies in the vector of baseline covariates

# Proposed Estimator

$$\boldsymbol{E[Y^f - Y^l | T = 1, D = f]} / P(S = 1 | T = 1, D = f)$$

- Cannot be directly estimated from flexible dosing trial
- Transport data from the fixed dosing arm - correcting for imbalances between the studies
- Assumptions to transport inferences
  - **Mean exchangeability w.r.t. $T$, conditional on baseline covariates $\boldsymbol{X}$[1] (transportability):**
    $E(Y^l | T = 1, \boldsymbol{X}) = E(Y^l | T = 0, \boldsymbol{X}) = E(Y^l | \boldsymbol{X})$.

[1] where differences in mean potential outcomes can be explained by imbalances across studies in the vector of baseline covariates

# Proposed Estimator

$$\boldsymbol{E[Y^f - Y^l | T = 1, D = f]} / P(S = 1 | T = 1, D = f)$$

- Cannot be directly estimated from flexible dosing trial
- Transport data from the fixed dosing arm - correcting for imbalances between the studies
- Assumptions to transport inferences
  - **Mean exchangeability w.r.t. $T$, conditional on baseline covariates $\boldsymbol{X}$[1] (transportability):**
    $E(Y^l | T = 1, \boldsymbol{X}) = E(Y^l | T = 0, \boldsymbol{X}) = E(Y^l | \boldsymbol{X})$.
  - **Positivity of trial assignment:** $0 < P(T = 1 | \boldsymbol{X}) < 1$.

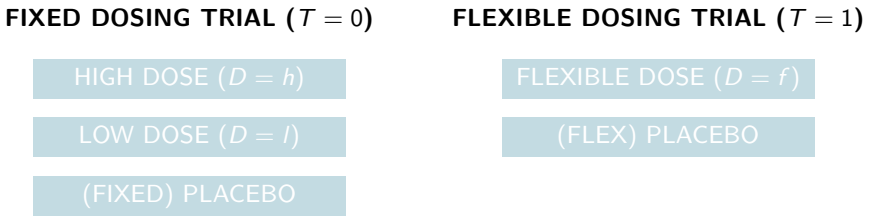
[1] where differences in mean potential outcomes can be explained by imbalances across studies in the vector of baseline covariates

# Proposed Estimator

| **FIXED DOSING TRIAL** ($T = 0$) | **FLEXIBLE DOSING TRIAL** ($T = 1$) |
|---|---|
| HIGH DOSE ($D = h$) | FLEXIBLE DOSE ($D = f$) |
| LOW DOSE ($D = l$) | (FLEX) PLACEBO |
| (FIXED) PLACEBO | |

An estimator for $E(Y^l | T = 1, D = f)$ is obtained by

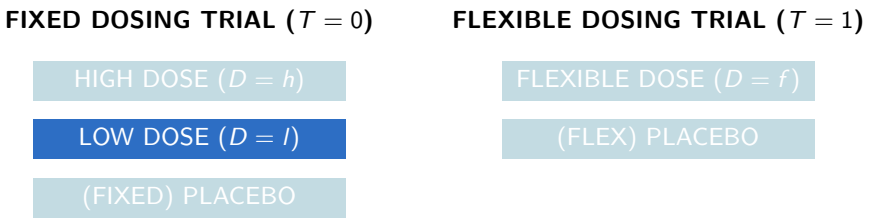
# Proposed Estimator




An estimator for $E(Y^l | T = 1, D = f)$ is obtained by

1. Estimating the parametric model for the selection model $P(T = 1|\boldsymbol{X})$
   e.g., $\pi(\boldsymbol{X}, \gamma) = expit(\gamma' \boldsymbol{X})$ for binary $T$

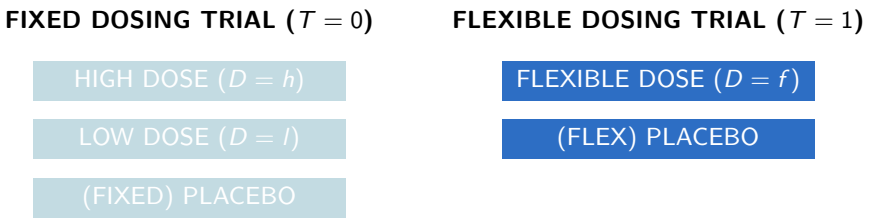
# Proposed Estimator




An estimator for $E(Y^l | T = 1, D = f)$ is obtained by

1. Estimating the parametric model for the selection model $P(T = 1|\boldsymbol{X})$
   e.g., $\pi(\boldsymbol{X}, \gamma) = expit(\gamma' \boldsymbol{X})$ for binary $T$
2. Fitting a weighted regression model for $Y$ given $\boldsymbol{X}$ among the patients on the low dose with weights $\hat{\pi}(\boldsymbol{X}, \hat{\gamma})/(1 - \hat{\pi}(\boldsymbol{X}, \hat{\gamma}))$
   e.g., $m(\boldsymbol{X}, \beta) = \beta' \boldsymbol{X}$ for continuous $Y$

# Proposed Estimator




An estimator for $E(Y^l|T = 1, D = f)$ is obtained by

1. Estimating the parametric model for the selection model $P(T = 1|\boldsymbol{X})$
   e.g., $\pi(\boldsymbol{X}, \gamma) = expit(\gamma'\boldsymbol{X})$ for binary $T$
2. Fitting a weighted regression model for $Y$ given $\boldsymbol{X}$ among the patients on the low dose with weights $\hat{\pi}(\boldsymbol{X}, \hat{\gamma})/(1 - \hat{\pi}(\boldsymbol{X}, \hat{\gamma}))$
   e.g., $m(\boldsymbol{X}, \beta) = \beta'\boldsymbol{X}$ for continuous $Y$
3. Taking the average of the predicted values over all patients in flexible dosing trial.

(Shu and Tan, 2018)

# Proposed Estimator

- This semi-parametric estimator, relies on
  - Selection Model for the association between trial and patients characteristics
  - Outcome Model

    $\Rightarrow$ **Asymptotically unbiased** when either model is correctly specified
- Achieves the **non-parametric efficiency bound** when both models are correctly specified

  (Shu and Tan, 2018; Dahabreh et al., 2018)

# Simulation Settings (Similar as Dahabreh et al., 2018)

- 10.000 simulations, $n = 500$ (100 in each arm)
- Randomization: $1:1$ in flexible and $1:1:1$ in fixed dosing trial
- 3 covariates: one imbalanced, two balanced between trials
  - $X_1 \sim N(0,1)$ in fixed dosing trial;
    $X_1 \sim N(0.5,1)$ in flexible dosing trial
  - $X_j \sim N(0.5,1)$ in both trials $(j = 2,3)$
- $S|X_1 \sim Ber(expit(0.7X_1))$
- Outcome $Y$ normally distributed with variance 1 and means
  - $2.25X_1 + X_2 + X_3$ when assigned to flexible dose and switched
  - $1.75X_1 + X_2 + X_3$ when assigned to flexible dose and not switched
  - $1.75X_1 + X_2 + X_3$ when assigned to fixed low dose
  - $2.5X_1 + X_2 + X_3$ when assigned to fixed high dose

# Simulation Results

**Impact of Misspecification - Operational Characteristics for treatment effect in switchers**

| Misspecification[2] | Method | Bias | SE |
|---|---|---|---|
| Correct | Proposed Estimator | −0.0002 | 0.0764 |
| | G-computation | −0.0002 | 0.0751 |
| Outcome misspec. | Proposed Estimator | 0.0225 | 0.1002 |
| | G-computation | 0.6430 | 0.1146 |
| SM[3] misspec. | Proposed Estimator | −0.0001 | 0.0757 |
| | G-computation | −0.0002 | 0.0751 |

[2]Misspecification: $X_1$ is replaced by $\log|X_1|$ in the working models

[3]SM: selection model

# Discussion

- **Business case**: enabled evaluation of a potential beneficial effect of higher dose for **subgroup** of patients switching to higher dose
  - Subgroup actually observed in data rather than one defined in terms of counterfactuals (*Principal Stratification*)

# Discussion

- **Business case**: enabled evaluation of a potential beneficial effect of higher dose for **subgroup** of patients switching to higher dose
  - Subgroup actually observed in data rather than one defined in terms of counterfactuals (*Principal Stratification*)
- **Estimands**: improved implementation design stage to improve model assumptions?
  - One trial with 5 arms: fixed/flexible blinded
  - (Stratified) randomization between trials: selection model known
  - In case of two trials: which baseline factors should be measured?

# Discussion

- **Business case**: enabled evaluation of a potential beneficial effect of higher dose for **subgroup** of patients switching to higher dose
  - Subgroup actually observed in data rather than one defined in terms of counterfactuals (*Principal Stratification*)
- **Estimands**: improved implementation design stage to improve model assumptions?
  - One trial with 5 arms: fixed/flexible blinded
  - (Stratified) randomization between trials: selection model known
  - In case of two trials: which baseline factors should be measured?
- **Future work**: improve performance under model misspecification via specialised nuisance parameter estimators

(e.g. Robins, Sued, Lei-Gomez, and Rotnitzky, 2007; Cao, Tsiatis and Davidian, 2009; Vermeulen and Vansteelandt, 2015)

# Thank you for your attention!


This project has received funding from VLAIO under the Baekeland grant agreement HBC.2017.0219.

Van Lancker, Vandebosch and Vansteelandt(2020): arXiv:2009.02136

# References



Issa J. Dahabreh, Sarah E. Robertson, Elizabeth A. Stuart, and Miguel A. Hernan, *Transporting inferences from a randomized trial to a new target population*, arXiv e-prints (2018), arXiv:1805.00550.


James Robins, Mariela Sued, Quanhong Lei-Gomez, and Andrea Rotnitzky, *Comment: Performance of double-robust estimators when inverse probability weights are highly variable*, Statist. Sci. **22** (2007), no. 4, 544–559.


Heng Shu and Zhiqiang Tan, *Improved Estimation of Average Treatment Effects on the Treated: Local Efficiency, Double Robustness, and Beyond*, arXiv e-prints (2018), arXiv:1808.01408.


Anastasios A. Tsiatis, Marie Davidian, and Weihua Cao, *Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data*, Biometrika **96** (2009), no. 3, 723–734.


Karel Vermeulen and Stijn Vansteelandt, *Bias-reduced doubly robust estimation*, Journal of the American Statistical Association **110** (2015), no. 511, 1024–1036 (eng).

# Proposal: Remark

- That efficiency can be gained when the selection model is known, may seem suprising as it is not the case for ATE.

# Proposal: Remark

- That efficiency can be gained when the selection model is known, may seem suprising as it is not the case for ATE.
- However, for estimation of (components of) ATT, the selection model is no longer ancillary.

# Proposal: Remark

- That efficiency can be gained when the selection model is known, may seem suprising as it is not the case for ATE.
- However, for estimation of (components of) ATT, the selection model is no longer ancillary.
- If we know the selection model, then a more efficient estimator for $E(Y^I | T = 1, D = f)$ can be attained by

# Proposal: Remark

- That efficiency can be gained when the selection model is known, may seem suprising as it is not the case for ATE.
- However, for estimation of (components of) ATT, the selection model is no longer ancillary.
- If we know the selection model, then a more efficient estimator for $E(Y^l | T = 1, D = f)$ can be attained by
  1. Fitting the same weighted regression as before

# Proposal: Remark

- That efficiency can be gained when the selection model is known, may seem suprising as it is not the case for ATE.
- However, for estimation of (components of) ATT, the selection model is no longer ancillary.
- If we know the selection model, then a more efficient estimator for $E(Y^I | T = 1, D = f)$ can be attained by
  1. Fitting the same weighted regression as before
  2. but, taking the sum of $\hat{\pi}(\boldsymbol{X}_i, \hat{\gamma})\hat{m}(\boldsymbol{X}_i; \hat{\beta})$ over **all** patients,

# Proposal: Remark

- That efficiency can be gained when the selection model is known, may seem suprising as it is not the case for ATE.
- However, for estimation of (components of) ATT, the selection model is no longer ancillary.
- If we know the selection model, then a more efficient estimator for $E(Y^l | T = 1, D = f)$ can be attained by
  1. Fitting the same weighted regression as before
  2. but, taking the sum of $\hat{\pi}(\boldsymbol{X}_i, \hat{\gamma})\hat{m}(\boldsymbol{X}_i; \hat{\beta})$ over **all** patients,
  3. and dividing by the number of patients in the flexible dosing trial

# Proposal: Remark

- That efficiency can be gained when the selection model is known, may seem suprising as it is not the case for ATE.
- However, for estimation of (components of) ATT, the selection model is no longer ancillary.
- If we know the selection model, then a more efficient estimator for $E(Y^l | T = 1, D = f)$ can be attained by
  1. Fitting the same weighted regression as before
  2. but, taking the sum of $\hat{\pi}(\boldsymbol{X}_i, \hat{\gamma})\hat{m}(\boldsymbol{X}_i; \hat{\beta})$ over **all** patients,
  3. and dividing by the number of patients in the flexible dosing trial
- Note: similar reasoning for $E(Y^f | T = 1, D = f)$

# Proposal: Remark

- Interestingly, this estimator is, generally, no longer doubly robust
  - Generally not unbiased when selection model is misspecified

[4] with a set of covariates that includes the covariates used in the model $m(\boldsymbol{X}; \beta)$

# Proposal: Remark

- Interestingly, this estimator is, generally, no longer doubly robust
  - Generally not unbiased when selection model is misspecified
- In the motivating study, the selection model is unknown and we therefore recommend the doubly robust estimator.

[4] with a set of covariates that includes the covariates used in the model $m(\boldsymbol{X}; \beta)$

# Proposal: Remark

- Interestingly, this estimator is, generally, no longer doubly robust
  - Generally not unbiased when selection model is misspecified
- In the motivating study, the selection model is unknown and we therefore recommend the doubly robust estimator.
- Randomisation of patients over studies, would justify use of the more efficient estimator.

[4] with a set of covariates that includes the covariates used in the model $m(\boldsymbol{X}; \beta)$

# Proposal: Remark

- Interestingly, this estimator is, generally, no longer doubly robust
  - Generally not unbiased when selection model is misspecified
- In the motivating study, the selection model is unknown and we therefore recommend the doubly robust estimator.
- Randomisation of patients over studies, would justify use of the more efficient estimator.
- Note: both estimators are equivalent when using a logistic regression for $\pi(\boldsymbol{X}, \gamma)$ [4]

[4] with a set of covariates that includes the covariates used in the model $m(\boldsymbol{X}; \beta)$

# Simulation Results

## Operational characteristics for treatment effect in switchers

| Misspecification[5] | Method | Bias | SE |
|---|---|---|---|
| Correct | NP[6] | −0.0002 | 0.0764 |
| | SP[7] | −0.0002 | 0.0761 |
| | G-computation | −0.0002 | 0.0751 |
| Outcome misspec. | NP | 0.0225 | 0.1002 |
| | SP | 0.0200 | 0.1031 |
| | G-computation | 0.6430 | 0.1146 |
| SM[8] misspec. | NP | −0.0001 | 0.0757 |
| | SP | −0.1614 | 0.0714 |
| | G-computation | −0.0002 | 0.0751 |

[5] Misspecification: $X_1$ is replaced by $\log|X_1|$ in the working models

[6] NP: eff. estimator under non-parametric model

[7] SP: eff. estimator under under semi-parametric model

[8] SM: selection model