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

Kelly Van Lancker
joint work with Stijn Vansteelandt and An Vandebosch
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
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
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 ($T = 1$)

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

Low dose ($D = l$)

Low + High ($D = lh$)

Deterministic rule for switching complicates inverse probability weighting: positivity violation

Available information too scarce: no arm assigned to fixed low dose

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
FLEXIBLE DOSING TRIAL ($T = 1$)

FLEXIBLE DOSE ($D = f$)

(FLEX) PLACEBO ($D = p$)

$S = 1$ → LOW+HIGH ($D = lh$)

$S = 0$ → LOW DOSE ($D = l$)

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

FLEXIBLE DOSING TRIAL ($T = 1$)

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

$S = 1$  $S = 0$

LOW + HIGH ($D = lh$)
LOW DOSE ($D = l$)

Not comparable!

- 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

FLEXIBLE DOSING TRIAL \( (T = 1) \)

- **FLEXIBLE DOSE** \( (D = f) \)
- **(FLEX) PLACEBO** \( (D = p) \)
- **LOW+HIGH** \( (D = lh) \)
- **LOW DOSE** \( (D = l) \)

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

Not comparable!
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**

### 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

**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
In flexible dosing study, for those who required switching:
How different would the average response $Y$ have been for them, had they not switched:

$$E[Y^{lh} - Y^l|T = 1, D = f, S = 1]$$
In flexible dosing study, for those who required switching:
How different would the average response $Y$ have been for them, had they not switched:

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

Some calculations

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

### Goal

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

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

Note: expectation is an ATT effect!
Proposed Estimator

Goal

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

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

Note: expectation is an ATT effect!
Proposed Estimator

Goal

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

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

- $P(S = 1|T = 1, D = f)$: 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:*

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

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

1. Fitting a regression model for $Y$ given baseline covariates $X$ 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.

$$E[Y_f - Y_l | T = 1, D = f] / P(S = 1 | T = 1, D = f)$$
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
An efficient estimator can be obtained by

1. Fitting a regression model for $Y$ given baseline covariates $X$ among the patients on the flexible dose.
An efficient estimator can be obtained by

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

2. Predicting the outcomes for all patients in flexible dosing trial.
An efficient estimator can be obtained by

1. Fitting a regression model for $Y$ given baseline covariates $X$ 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

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

- 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

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

- 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

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

- 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 \( X^1 \) (transportability):
    \[
    E(Y^l | T = 1, X) = E(Y^l | T = 0, X) = E(Y^l | X).
    \]

\(^1\)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
  - Mean exchangeability w.r.t. \( T \), conditional on baseline covariates \( X^1 \) (transportability):
    \[ E(Y_l | T = 1, X) = E(Y_l | T = 0, X) = E(Y_l | X). \]
  - Positivity of trial assignment: \( 0 < P(T = 1 | X) < 1. \)

\(^1\)where differences in mean potential outcomes can be explained by imbalances across studies in the vector of baseline covariates.
An estimator for $E(Y^l | T = 1, D = f)$ is obtained by
An estimator for $E(Y^l | T = 1, D = f)$ is obtained by

1. Estimating the parametric model for the selection model $P(T = 1 | X)$

\[
\pi(X, \gamma) = \expit(\gamma'X)
\]

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

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

2. Fitting a weighted regression model for $Y$ given $X$ among the patients on the low dose with weights $\hat{\pi}(X, \hat{\gamma})/(1 - \hat{\pi}(X, \hat{\gamma}))$
   
   e.g., $m(X, \beta) = \beta'X$ for continuous $Y$
An estimator for \( E(Y^l|T = 1, D = f) \) is obtained by

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

2. Fitting a weighted regression model for \( Y \) given \( X \) among the patients on the low dose with weights \( \hat{\pi}(X, \hat{\gamma})/(1 - \hat{\pi}(X, \hat{\gamma})) \)
   
   e.g., \( m(X, \beta) = \beta'X \) 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
    \[\Rightarrow \text{Asymptotically unbiased}\] when either model is correctly specified

- Achieves the \textbf{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 ($j = 2, 3$)

- $S|X_1 \sim Ber(\text{expit}(0.7X_1))$

- 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

<table>
<thead>
<tr>
<th>Misspecification</th>
<th>Method</th>
<th>Bias</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>Proposed Estimator</td>
<td>−0.0002</td>
<td>0.0764</td>
</tr>
<tr>
<td></td>
<td>G-computation</td>
<td>−0.0002</td>
<td>0.0751</td>
</tr>
<tr>
<td>Outcome misspec.</td>
<td>Proposed Estimator</td>
<td>0.0225</td>
<td>0.1002</td>
</tr>
<tr>
<td></td>
<td>G-computation</td>
<td>0.6430</td>
<td>0.1146</td>
</tr>
<tr>
<td>SM³ misspec.</td>
<td>Proposed Estimator</td>
<td>−0.0001</td>
<td>0.0757</td>
</tr>
<tr>
<td></td>
<td>G-computation</td>
<td>−0.0002</td>
<td>0.0751</td>
</tr>
</tbody>
</table>

²Misspecification: $X_1$ is replaced by $\log |X_1|$ in the working models
³SM: 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?
**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.


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.

That efficiency can be gained when the selection model is known, may seem surprising as it is not the case for ATE.
Proposal: Remark

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

- However, for estimation of (components of) ATT, the selection model is no longer ancillary.
That efficiency can be gained when the selection model is known, may seem surprising 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. Taking the sum of $\hat{\pi}(X_i, \hat{\gamma}) \hat{m}(X_i; \hat{\beta})$ over all patients,
3. and dividing by the number of patients in the flexible dosing trial

Note: similar reasoning for $E(Y^f | T = 1, D = f)$
That efficiency can be gained when the selection model is known, may seem surprising 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
Proposal: Remark

- That efficiency can be gained when the selection model is known, may seem surprising 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 $\hat{\pi}(X_i, \hat{\gamma})\hat{m}(X_i; \hat{\beta})$ over all patients,
That efficiency can be gained when the selection model is known, may seem surprising 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 $\hat{\pi}(X_i, \hat{\gamma}) \hat{m}(X_i; \hat{\beta})$ over all patients,
3. and dividing by the number of patients in the flexible dosing trial
That efficiency can be gained when the selection model is known, may seem surprising 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 $\hat{\pi}(X_i, \hat{\gamma}) \hat{m}(X_i; \hat{\beta})$ over all patients,
3. and dividing by the number of patients in the flexible dosing trial

Note: similar reasoning for $E(Y^f|T = 1, D = f)$
Interestingly, this estimator is, generally, no longer doubly robust

- Generally not unbiased when selection model is misspecified

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

\[^4\text{with a set of covariates that includes the covariates used in the model } m(X; \beta)\]
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 \( m(X; \beta) \)
Proposition: 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 $\pi(X, \gamma)^4$

\(^4\) with a set of covariates that includes the covariates used in the model $m(X; \beta)$
## Simulation Results

### Operational characteristics for treatment effect in switchers

<table>
<thead>
<tr>
<th>Misspecification (^5)</th>
<th>Method</th>
<th>Bias</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>NP(^6)</td>
<td>−0.0002</td>
<td>0.0764</td>
</tr>
<tr>
<td></td>
<td>SP(^7)</td>
<td>−0.0002</td>
<td>0.0761</td>
</tr>
<tr>
<td></td>
<td>G-computation</td>
<td>−0.0002</td>
<td>0.0751</td>
</tr>
<tr>
<td>Outcome misspec.</td>
<td>NP</td>
<td>0.0225</td>
<td>0.1002</td>
</tr>
<tr>
<td></td>
<td>SP</td>
<td>0.0200</td>
<td>0.1031</td>
</tr>
<tr>
<td></td>
<td>G-computation</td>
<td>0.6430</td>
<td>0.1146</td>
</tr>
<tr>
<td>SM(^8) misspec.</td>
<td>NP</td>
<td>−0.0001</td>
<td>0.0757</td>
</tr>
<tr>
<td></td>
<td>SP</td>
<td>−0.1614</td>
<td>0.0714</td>
</tr>
<tr>
<td></td>
<td>G-computation</td>
<td>−0.0002</td>
<td>0.0751</td>
</tr>
</tbody>
</table>

---

\(^5\)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