Capturing the effect of treatment breaks in users of hormonal contraception

Alessandra Meddis, Amalie Lykkemark Møller

Section of Biostatistics, Copenhagen University

DSTS Two-day meeting,
November 13 2024

*work in collaboration with the Danish Cancer Research Institute and Nordsjællands Hospital

Disclaimer: The focus of this talk is not on the analysis itself, but rather on the underlying challenges

Disclaimer: The focus of this talk is not on the analysis itself, but rather on the underlying challenges

- Present the main motivation
- Emphasis on challenges in addressing the specific research question
- Discuss what is feasible in this context

Motivation

It is well known that use of hormonal contraception leads to an increase on the rates of venous thrombosis event (VTE). It seems higher at start and then reduces over the first year of use and remains stable thereafter[1].

[1] Suissa, Samy, et al. "First-time use of newer oral contraceptives and the risk of venous thromboembolism." Contraception 56.3 (1997): 141-146.

Motivation

It is well known that use of hormonal contraception leads to an increase on the rates of venous thrombosis event (VTE). It seems higher at start and then reduces over the first year of use and remains stable thereafter[1].

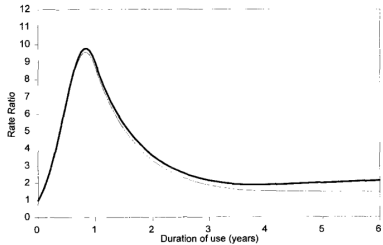


Figure: Rate Ratio of venous thrombosis as a function of duration of CHC use among first-time users and non-users.

[1] Suissa, Samy, et al. "First-time use of newer oral contraceptives and the risk of venous thromboembolism." Contraception 56.3 (1997): 141-146.

Motivation

It is well known that use of hormonal contraception leads to an increase on the rates of venous thrombosis event (VTE). It seems higher at start and then reduces over the first year of use and remains stable thereafter[1].

Guidelines discourage taking a break of hormonal contraception[2]:

"...Risk of VTE is highest in the months immediately after initiation of CHC or when ***restarting after a break of at least 1 month***[3]*. The risk then reduces over the first year of use and remains stable thereafter.* ***Frequent stopping and starting of CHC is therefore discouraged****..."*

Very little evidence

[1] Suissa, Samy, et al. "First-time use of newer oral contraceptives and the risk of venous thromboembolism." Contraception 56.3 (1997): 141-146.

[2] Faculty of Sexual and Reproductive Healthcare, UK, Clinical Guideline: Combined Hormonal Contraception (October 2023)

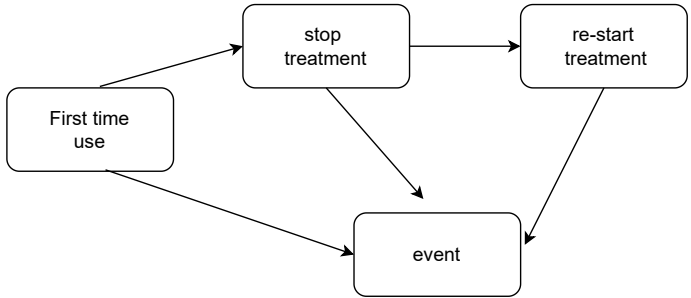
[3] Dinger J, Möhner S, Do Minh T. Early use effects on the risk of venous thromboembolism after initiation of oral contraceptive use. Eur J Contracept Reprod Health Care 2010;15 (Suppl 1):43.

Research question

Does the risk of venous thrombosis increase when women re-start hormonal contraception after their first break?

Research question

Does the risk of venous thrombosis increase when women re-start hormonal contraception after their first break?



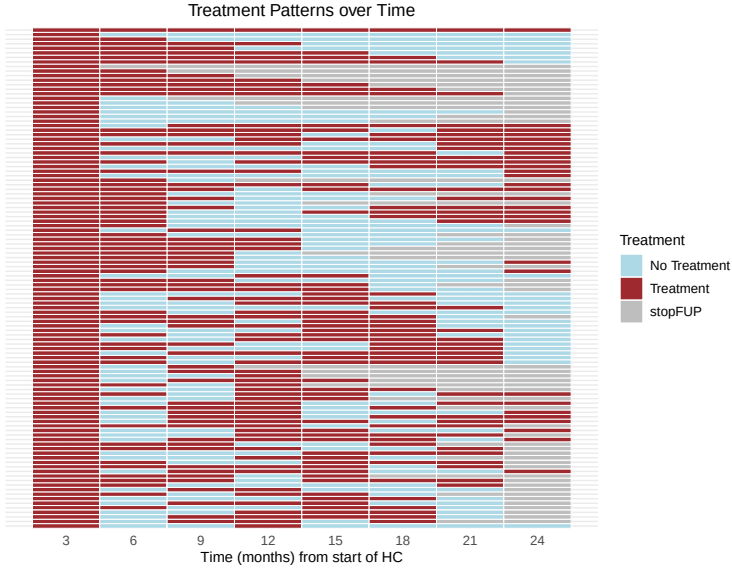
Data collection

Danish register data from women who are 10-34 years old at any time point between 2000 and 2023 and had started hormonal contraception for the first time (one prescription) and had lived in Denmark for the past 5 years.[4]

- women have heterogeneous treatment patterns

[4] meeting some inclusion/exclusion criteria on comorbidities and previous pregnancies

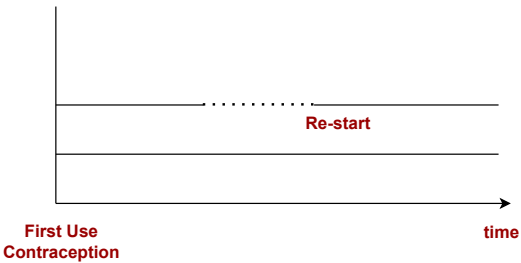
Data collection



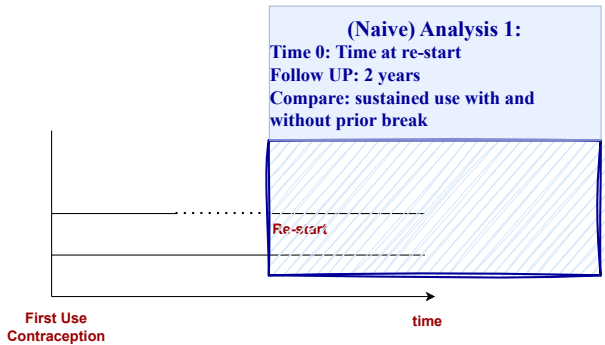
Two Distinct Approaches

We explore the research question through two separate analyses

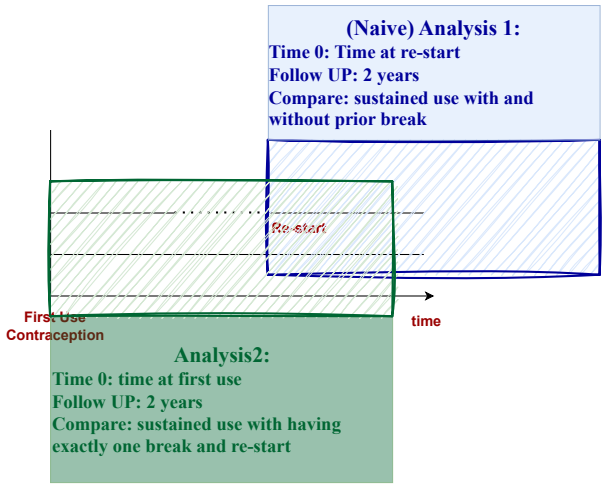
- within a causal framework, defining the scientific questions in terms of counterfactuals, to clarify causal relationships
- addressing distinct aspects of the research question



Two Distinct Approaches



Two Distinct Approaches



Observational study with longitudinal data

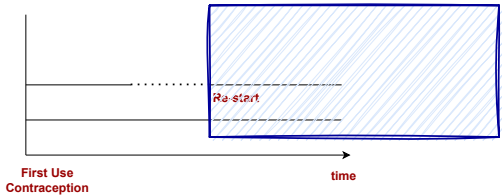
On a discrete time scale: $\{t_0, t_1, t_2, .., t_K\}$, at time t_k, we denote :

- treatment variable A_k
- binary outcome $Y_k = I(T \leq t_k)$
- censoring indicator C_k
- vector of time-dependent covariates L_k

We consider the following data structure:

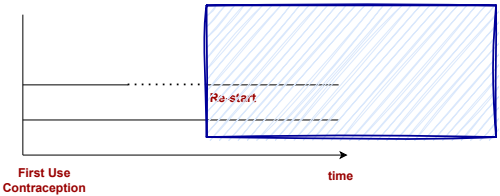
$$O = (L_0, A_0, C_1, Y_1, L_1, A_1, ..., Y_K) \sim P_0$$

and let $O_1, ..., O_n$ denote a sample of n i.i.d observations of O.
We indicate treatment and covariate history up to time t_k with $\overline{A}_k$ and $\overline{L}_k$.
We denote with $\overline{d}_k$ a treatment intervention and $Y^{\overline{d}_K}$ is what would be observed at time t_k if a subject had received the treatment $\overline{d}_k$.



(Naive) Analysis 1

Isolate the effect of re-starting. What is the risk when women re-start hormonal contraception for the first time?



(Naive) Analysis 1

Isolate the effect of re-starting. What is the risk when women re-start hormonal contraception for the first time?

*"...Risk of VTE is highest when **restarting after a break of at least 1 month**..."*

Analysis 1: Definition of Target Parameter?

Time 0: time of re-start of the hormonal contraception.
We would like to compare the 2 years absolute risk of venous thrombosis if women had sustained use of hormonal contraception without a break (d^1) and with one prior break (d^2)

Analysis 1: Definition of Target Parameter?

Time 0: time of re-start of the hormonal contraception.
We would like to compare the 2 years absolute risk of venous thrombosis if women had sustained use of hormonal contraception without a break (d^1) and with one prior break (d^2)

Analysis 1: Definition of Target Parameter?

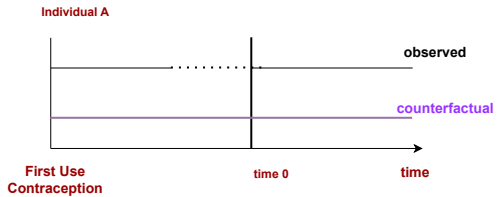
Time 0: time of re-start of the hormonal contraception.
We would like to compare the 2 years absolute risk of venous thrombosis if women had sustained use of hormonal contraception without a break (d^1) and with one prior break (d^2)

$$\Psi(P_0) = \mathbb{E}[Y_K^{\bar{d}^1} - Y_K^{\bar{d}^2}]$$

Analysis 1: Definition of Target Parameter?

Time 0: time of re-start of the hormonal contraception.
We would like to compare the 2 years absolute risk of venous thrombosis if women had sustained use of hormonal contraception without a break (d^1) and with one prior break (d^2)

$$\Psi(P_0) = \mathbb{E}[Y_K^{\bar{d}_K^1} - Y_K^{\bar{d}_K^2}] = ???$$



We can define counterfactuals among women that *survived* at re-start. Here, the two groups are not comparable if treatment has an effect.

Identifiability of the Target Parameter

Assume that

- the probability of having a break is independent to any covariates
- the risk of event only depends on the treatment status

Identifiability of the Target Parameter

Assume that

- the probability of having a break is independent to any covariates
- the risk of event only depends on the treatment status

Consider a Cox model with an individual Gamma frailty Z, where we fix the time of the break (T^b) and of re-start (T^r).
For $t > T^b$

$$\lambda(t, A, Z) = \begin{cases} Z\lambda_0(t)\exp(\beta_1 A) & \text{if } t \leq T^r \\ Z\lambda_0(t)\exp(\beta_1 A + \gamma\tilde{A}) & \text{if } t > T^r \end{cases}$$

$\tilde{A} = 1$ if the individuals had a break.

Identifiability of the Target Parameter

We compare the expectations for the frailty for survivors at time T^r in the two groups.

$$\mathbb{E}[Z|T > T^r, A] = \frac{\int_0^\infty zS(T^r|Z,A)f_z(z)dz}{\int_0^\infty S(T^r|Z,A)f_z(z)dz} = \frac{\theta}{\theta + \Lambda_0(T_r)\exp(\beta_1 A)}$$

where θ is the frailty parameter and $\Lambda_0(t) = \int_0^\infty \lambda_0(s)ds$

Identifiability of the Target Parameter

We compare the expectations for the frailty for survivors at time T^r in the two groups.

$$\mathbb{E}[Z|T > T^r, A] = \frac{\int_0^\infty zS(T^r|Z,A)f_z(z)dz}{\int_0^\infty S(T^r|Z,A)f_z(z)dz} = \frac{\theta}{\theta + \Lambda_0(T_r)\exp(\beta_1 A)}$$

where θ is the frailty parameter and $\Lambda_0(t) = \int_0^\infty \lambda_0(s)ds$

For continuous users:

$$\mathbb{E}[Z|T > T^r, A = 1] = \frac{\theta}{\theta + \Lambda_0(T_r)\exp(\beta_1)}$$

For re-starters:

$$\mathbb{E}[Z|T > T^r, A = 0] = \frac{\theta}{\theta + \Lambda_0(T_r)}$$

Identifiability of the Target Parameter

We compare the expectations for the frailty for survivors at time T^r in the two groups.

$$\mathbb{E}[Z|T > T^r, A] = \frac{\int_0^\infty zS(T^r|Z,A)f_z(z)dz}{\int_0^\infty S(T^r|Z,A)f_z(z)dz} = \frac{\theta}{\theta + \Lambda_0(T_r)\exp(\beta_1 A)}$$

where θ is the frailty parameter and $\Lambda_0(t) = \int_0^\infty \lambda_0(s)ds$

For continuous users:

$$\mathbb{E}[Z|T > T^r, A = 1] = \frac{\theta}{\theta + \Lambda_0(T_r)\exp(\beta_1)}$$

For re-starters:

$$\mathbb{E}[Z|T > T^r, A = 0] = \frac{\theta}{\theta + \Lambda_0(T_r)}$$

with $\beta_1 > 0 \quad \mathbb{E}[Z|T > T^r, A = 1] < \mathbb{E}[Z|T > T^r, A = 0]$

At time of re-start (T^r) they are not comparable and subjects in the continuous use group are less frail. Even when $\gamma = 0$ we might observe an effect.

Estimation of the Target Parameter

We could define the continuous users group by sequential matching at time of re-start.

When a women re-start the treatment, she is matched to another women that started the treatment on the same date but have not had a break yet.
The time of re-start is set as the index date for the matched pair.

However, as discussed before, what would be the causal interpretation?
The two groups are not comparable at time 0.

Estimation of the Target Parameter

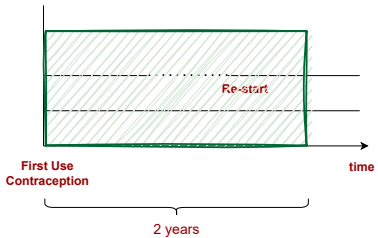
We could define the continuous users group by sequential matching at time of re-start.

When a women re-start the treatment, she is matched to another women that started the treatment on the same date but have not had a break yet.
The time of re-start is set as the index date for the matched pair.

However, as discussed before, what would be the causal interpretation?
The two groups are not comparable at time 0.

Analysis 2

Does having the first break increase the risk of venous thrombosis?



Definition of treatment interventions

Time 0: first use of hormonal contraception.

Interest lies on comparing the risk of the event in 2 years if women had been under sustained use of the treatment and if women had one treatment break within 2 years.

We consider the two interventions:

Sustained use:
static intervention

$$d_k^1 = \mathbb{P}(A_k = 1) = 1 \quad \forall k$$

Having one break:
stochastic intervention

$$d_k^2 = \mathbb{P}(A_k = 1 | \overline{A}_{k-1}, \overline{L}_k) = \begin{cases} \pi_k(\overline{L}_k) & \text{if } \textbf{not-yet} \text{ a break} \\ \tilde{\pi}_k(\overline{L}_k) & \text{if } \textbf{in} \text{ a break} \\ 1 & \text{if } \textbf{already} \text{ a break} \end{cases}$$

Definition of treatment interventions

Time 0: first use of hormonal contraception.

Interest lies on comparing the risk of the event in 2 years if women had been under sustained use of the treatment and if women had one treatment break within 2 years.

We consider the two interventions:

Sustained use:
static intervention

$$d_k^1 = \mathbb{P}(A_k = 1) = 1 \quad \forall k$$

Having one break:
stochastic intervention

$$d_k^2 = \mathbb{P}(A_k = 1 | \overline{A}_{k-1}, \overline{L}_k) = \begin{cases} \pi_k(\overline{L}_k) & \text{if } \textbf{not-yet} \text{ a break} \\ \tilde{\pi}_k(\overline{L}_k) & \text{if } \textbf{in} \text{ a break} \\ 1 & \text{if } \textbf{already} \text{ a break} \end{cases}$$

We consider the **combination of being in a break and re-starting the treatment**.

Example of stochastic Intervention

Stochastic interventions often aim to mimic more realistic or feasible treatment practices, rather than forcing a strict treatment strategy.

Example of stochastic Intervention

Stochastic interventions often aim to mimic more realistic or feasible treatment practices, rather than forcing a strict treatment strategy.

Example:
The stochastic intervention "prescribes" that women, without a break, will stay under hormonal contraception with a 40% during the first year and 70% after the first year of use.

$$\pi_k = 0.4, k < 4, \pi_k = 0.7, k \geq 4$$

In the break, women will re-start hormonal contraception with a 70% probability if they are younger(≤ 22) and 30% if older (>22)

$$\tilde{\pi}_k = 0.7I(age \leq 22) + 0.3I(age > 22), \quad \forall k$$

Definition of the Target Parameter

2 years - Absolute Risk Difference

$$\Psi(P_0) = \mathbb{E}[Y_K^{\overline{d}_K^1}] - \mathbb{E}[Y_K^{\overline{d}_K^2}]$$

Ψ is identifiable under the assumption of **consistency, sequential exchangeability and positivity**.

[4] Robins, J. M. (2000). Robust estimation in sequentially ignorable missing data and causal inference models. Proceedings of the American Statistical Association Section on Bayesian Statistical Science, 6–10.

Definition of the Target Parameter

2 years - Absolute Risk Difference

$$\Psi(P_0) = \mathbb{E}[Y_K^{\overline{d}_K^1}] - \mathbb{E}[Y_K^{\overline{d}_K^2}]$$

Ψ is identifiable under the assumption of **consistency, sequential exchangeability and positivity**.

In the longitudinal setting, the parameter can be expressed as a sequential representation [4]:

$$\mathbb{E}[Y_K = 1 \mid \overline{d}_K] = \underbrace{\int \ldots \underbrace{\int \mathbb{E}[Y_K \mid \overline{L}_{K-1}, \overline{A}_{K-1}] \prod_{s=0}^{K-1} d_s(A_s \mid \overline{L}_s, \overline{A}_{s-1}) f(\overline{L}_s \mid \overline{L}_{s-1}, \overline{A}_{s-1}) dA_{K-1} dL_{K-1}}_{\mathbb{E}_{d,f}[Y_K \mid \overline{A}_K, \overline{L}_K]} \cdots dA_0 \, dL_0}_{\mathbb{E}[.. \mid \mathbb{L}_{\nu'}]}$$

[4] Robins, J. M. (2000). Robust estimation in sequentially ignorable missing data and causal inference models. Proceedings of the American Statistical Association Section on Bayesian Statistical Science, 6–10.

Definition of the Target Parameter

2 years - Absolute Risk Difference

$$\Psi(P_0) = \mathbb{E}[Y_K^{\bar{d}_K^1}] - \mathbb{E}[Y_K^{\bar{d}_K^2}]$$

Expressed as a sequence of nested expectations:

$$\overline{Q}_{K+1}(\overline{L}_K, \overline{A}_K) = \mathbb{E}_{d,f}[Y_K|\overline{L}_K, \overline{A}_K]$$

$$\overline{Q}_K(\overline{L}_{K-1}, \overline{A}_{K-1}) = \mathbb{E}_{d,f}[\overline{Q}_{K+1}(\overline{L}_K, \overline{A}_K)|\overline{L}_{K-1}, \overline{A}_{K-1}]$$

$$\dots$$

$$\overline{Q}_k(\overline{L}_{k-1}, \overline{A}_{k-1}) = \mathbb{E}_{d,f}[\overline{Q}_{k+1}(\overline{L}_k, \overline{A}_k)|\overline{L}_{k-1}, \overline{A}_{k-1}]$$

$$\dots$$

$$\overline{Q}_1(L_0) = \mathbb{E}[\overline{Q}_2(L_1)|L_0] = \mathbb{E}[\mathbb{E}_{d,f}[\dots\mathbb{E}_{d,f}[Y_K|\overline{L}_K, \overline{A}_K]\dots]|L_0]$$

Estimation of the Target Parameter

Longitudinal Targeted Minimum Loss Estimation (LTMLE) where we estimate the target parameter Ψ_0 solving the efficient influence (EI) curve equation.

- double robustness
- targeting step using the EIF
- integrate Machine Learning models

Estimation of the Target Parameter

Longitudinal Targeted Minimum Loss Estimation (LTMLE) where we estimate the target parameter Ψ_0 solving the efficient influence (EI) curve equation.

- double robustness
- targeting step using the EIF
- integrate Machine Learning models

The idea is to construct the estimate via an iterative procedure where at each time point k:

1. Estimate an initial estimator for the conditional expectation $\overline{Q}_k(\overline{L}_{k-1}, \overline{A}_{k-1})$
2. Update it to solve the efficient influence function at k $\overline{Q}_k^*(\overline{L}_{k-1}, \overline{A}_{k-1})$

We obtain the estimator $\overline{Q}_1(L_0)$ obtained by estimating a sequence of nested expectations, designed to solve the efficient influence function.

Simulation Study

We **do not compare** the two analyses, as they target two different research questions. The goal is to illustrate how each method works in a straightforward scenario.

Simulation Study

We **do not compare** the two analyses, as they target two different research questions. The goal is to illustrate how each method works in a straightforward scenario.

Generate data for 10000 individuals where everyone starts under treatment and at each visit (3 months):

- treatment depends only on its history: $A_k \sim Bin(\pi_k(\overline{A}_{k-1}))$
- time-to-event by piecewise constant hazard model

$$\lambda(t_k) = \lambda_0(t_k)\exp(\beta_1^k A_k + \gamma \tilde{A}_k)$$

with $\exp(\beta_1^k) \geq 1$ with higher treatment effect in the first 3 months and $\exp(\beta_1^{12}) = 1$

Simulation Study

We **do not compare** the two analyses, as they target two different research questions. The goal is to illustrate how each method works in a straightforward scenario.

Generate data for 10000 individuals where everyone starts under treatment and at each visit (3 months):

- treatment depends only on its history: $A_k \sim Bin(\pi_k(\overline{A}_{k-1}))$
- time-to-event by piecewise constant hazard model

$$\lambda(t_k) = \lambda_0(t_k)\exp(\beta_1^k A_k + \gamma \tilde{A}_k)$$

with $\exp(\beta_1^k) \geq 1$ with higher treatment effect in the first 3 months and $\exp(\beta_1^{12}) = 1$

Provide the estimation of the target parameter for the two analyses in the context of

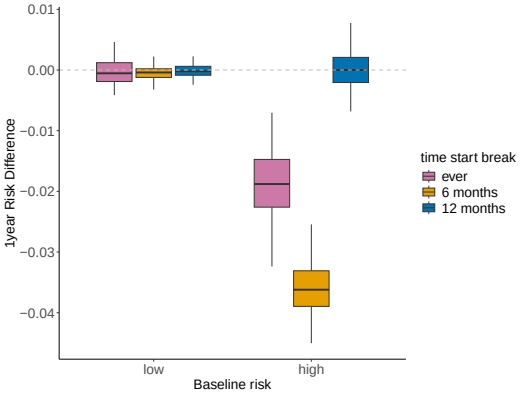
- no time-dependent confounding
- only administrative censoring
- no competing event

Simulation Study: Estimation

- Analysis 1: Exact matching + Kaplan-Meier
  - Match on the date of first use
  - Artificial Censor women that start the break after being matched
  - Estimate the 1-year Absolute Risk Difference comparing the use of hormonal contraception with and without prior break by Kaplan Meier
- Analysis 2: LTMLE with **static interventions**
  - 2-years Absolute Risk Difference comparing the risk under the two interventions of **sustained use** and having a break **of 6 months, after 6 months from first use**
  - True Value calculated via simulation of counterfactuals for 100000 individuals.

Simulation Results: Analysis 1

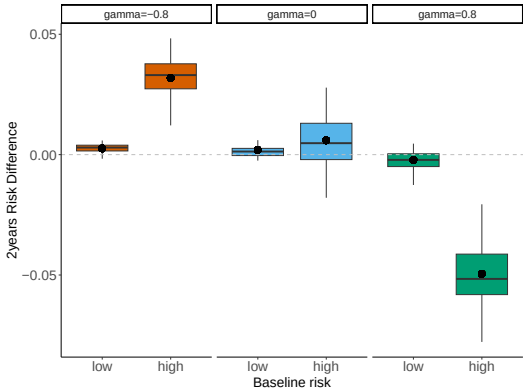
Results for $\gamma = 0$ at varying of start of break



- low/high risk: such that the Risk after 2 years of continuous use is 0.07 % (mimicking the motivation data) and 1% respectively
- estimate a negative break effect because of the selection of less frail individuals at time of re-start
- generated data such that $\exp(\beta^{12}) = 1$, no Bias when having the break at 12 months

Simulation Results: Analysis 2

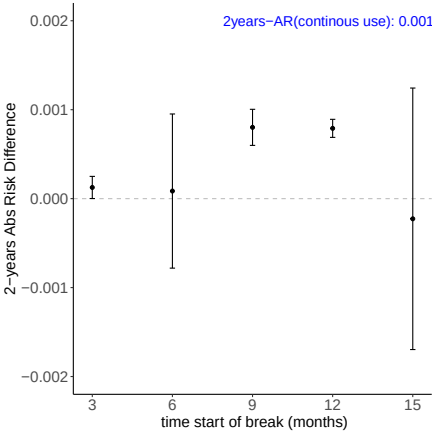
Results at varying of γ for the static interventions of sustained use and one 6 months long break at 6 months:



- low/high risk: such that the Risk after 2 years of continuous use is 0.07 % (mimicking the motivation data) and 1% respectively
- consistently estimate the Target Parameter with higher risk under continuous use for $\gamma < 0$ and lower risk for $\gamma > 0$

Preliminary Results

LTMLE analysis with static intervention for 6 months long break at varying of time of break start.



Discussion

- Conclusions like *The risk of VTE is highest when restarting treatment* are complex to achieve and to interpret. Highest compared to ?
  - If we compare to the moment just before the break ended, we overlook that if women had continued treatment without interruption, they would be exposed to higher risk
- It's difficult to separate the effect of re-starting from the break itself, especially since the intervention begins with the break.

Discussion

- Conclusions like *The risk of VTE is highest when restarting treatment* are complex to achieve and to interpret. Highest compared to ?
  - If we compare to the moment just before the break ended, we overlook that if women had continued treatment without interruption, they would be exposed to higher risk
- It's difficult to separate the effect of re-starting from the break itself, especially since the intervention begins with the break.
- Use of LTMLE for the estimation of the ATE with stochastic interventions, it is not ready yet but we are working on it.
- I haven not addressed competing events here, which include not only death (a minor concern in our case) but also pregnancy, as many women discontinue treatment to pursue pregnancy.

References

Suissa, Samy, et al. "First-time use of newer oral contraceptives and the risk of venous thromboembolism." Contraception 56.3 (1997): 141-146.

Dinger J, Möhner S, Do Minh T. Early use effects on the risk of venous thromboembolism after initiation of oral contraceptive use. Eur J Contracept Reprod Health Care 2010;15 (Suppl 1):43.

Robins, J. M. (2000). Robust estimation in sequentially ignorable missing data and causal inference models. Proceedings of the American Statistical Association Section on Bayesian Statistical Science, 6–10.

Hernán MA. The hazards of hazard ratios. Epidemiology. 2010 Jan;21(1):13-5.

Van der Laan, Mark J., and Sherri Rose. Targeted learning: causal inference for observational and experimental data. Vol. 4. New York: Springer, 2011.

van der Laan, Mark J., et al. "Stochastic treatment regimes." Targeted learning in data science: Causal inference for complex longitudinal studies (2018): 219-232.

Thank You!

Assumptions for identifiability

- **Positivity**: for each possible history, there is a non-zero chance that an individual could follow the specified treatment intervention

$$d_k(A_k|\overline{A}_{k-1} = \overline{a}_{k-1}, \overline{L}_k = \overline{l}_k) > 0 \quad \forall k$$

- **Sequential Exchangeability** (unmeasured confounders)

$$Y^{\overline{d}_k} \perp\!\!\!\perp \overline{A}_k|_{\overline{L}_k, \overline{A}_{k-1}}$$

- **Consistency**: clear definition of the treatment strategy such that

$$\text{if } \overline{A}_K = \overline{a}_K, \text{according to } d, \text{ then } Y^{\overline{a}_K} = Y$$

Efficient Influence Curve

In the Longitudinal setting, the efficient influence curve for the absolute risk is:

$$EIF_\Psi(O) = \overline{Q}_1(L_0) - \Psi + \sum_{k=1}^{K+1} \prod_{j=0}^{k-1} \frac{d_j(a_j|\overline{a}_{j-1}, \overline{l}_j)}{\hat{d}_j(a_j|\overline{a}_{j-1}, \overline{l}_j)} (\overline{Q}_k(\overline{L}_{k-1}, \overline{A}_{k-1}) - \overline{Q}_{k-1}(\overline{L}_{k-2}, \overline{A}_{k-2}))$$

We have that the estimated parameter is asimptotically efficient: (unbiased and smallest variance)

$$\hat{\Psi} = \Psi + \frac{1}{n} \sum_{i=1}^{n} EIF_\Psi(O_i) + o_P(n^{-1/2})$$

Efficient Influence Curve

In the Longitudinal setting, the efficient influence curve for the absolute risk is:

$$EIF_\Psi(O) = \overline{Q}_1(L_0) - \Psi + \sum_{k=1}^{K+1} \prod_{j=0}^{k-1} \frac{d_j(a_j|\overline{a}_{j-1}, \overline{l}_j)}{\hat{d}_j(a_j|\overline{a}_{j-1}, \overline{l}_j)} (\overline{Q}_k(\overline{L}_{k-1}, \overline{A}_{k-1}) - \overline{Q}_{k-1}(\overline{L}_{k-2}, \overline{A}_{k-2}))$$

We have that the estimated parameter is asimptotically efficient: (unbiased and smallest variance)

$$\hat{\Psi} = \Psi + \frac{1}{n} \sum_{i=1}^{n} EIF_\Psi(O_i) + o_P(n^{-1/2})$$

- log-likelihood loss function and logistic regression for the outcome.
- one step update to solve the efficient influence function at each time