# The Use of the Prevalent New User Design in the Setting of Complex Treatment Patterns: A Simulation Study Protocol

## Project description

This project will investigate the effect of different matching strategies on the estimation of treatment effects when conducting a cohort study using a prevalent-new user study design (PNUD) in the setting of complex treatment patterns.

Complexity in treatment patterns can refer to the number of prior treatments, the type of prior treatment, e.g., treatments that have cumulative effects, or evolving disease severity where treatment may change over time. This project will focus on the situation where the complexity is due to switching between two or more treatments over time.

A simulated dataset will be generated that represents individual dispensing claims for a cohort of individuals dispensed up to three different medicines.  Outcome events associated with the target medicine of interest will be simulated according to different patterns of prior exposure to alternative medicines. We will compare estimation of treatment effects using different matching strategies for selecting comparators (i.e., users that don’t switch treatment) to prevalent new users (treatment switchers) when conducting a PNUD cohort study.

The impact of different matching strategies for selecting comparators for each prevalent new user on treatment effect estimates will be assessed using measures of bias, precision (empirical and model SE), and coverage.

## Motivating example

The scenario that the simulated data represents is a cohort of patients with an autoimmune disease such as rheumatoid arthritis who are treated with either a biologic medicine (e.g., TNFi or non-TNFi) or are untreated.  We wish to measure the risk of herpes zoster associated with the initiation of a new medicine that is introduced to the market, JAKi.  Each patient is either newly treated with a treatment of interest (JAKi) (incident users) or switch to JAKi after having received prior treatment with another medicine for that condition for some time prior to initiating JAKi (prevalent new users).

In a traditional study design, such as an active comparator new user design (ACNU), we would compare incident users of JAKi to incident users of alternative medicines and those who were prevalent new users would be excluded from the study. However, this limits the population size, and results in evidence that relates only to the newly diagnosed and newly treated. We then have no evidence of risk or safety for those with prior treatment that switched to the new medicine, which is a group of medicine users that are also usually excluded from clinical trials.

In the prevalent new user design patients that switch treatment are included in the study cohort, allowing for comparing those that switch with those that continued treatment. In this example, the prevalent new users have been treated with one or both biologics (i.e., may have switched between the two) and will be compared with a user that continued their treatment with the biologic, accounting for any prior treatment.

Two forms of confounding: 1) due to prior medication history (prevalent user bias) and 2) due to patient characteristics that increase the risk of herpes zoster, e.g., older age, more severe disease activity and concomitant glucocorticoid dispensing (confounding by indication), need to be accounted for in the study design.

Matching on treatment history is used to avoid prevalent user bias and propensity scores are used to account for confounding by indication. Propensity scores are an established method for confounding control and are not the focus of this study. Our focus is on trialling different matching strategies to account for prior treatment history and their impact on reduction of bias.

## Methods

### Aims

To evaluate the impacts of different matching strategies on the estimation of treatment effects when outcomes are generated using different risk models in cohort studies using PNUD.

### Data-generating mechanism

We will consider nine data-generating mechanisms. For all nine, data are simulated on patients, representing medicines claims dispensing. Each data-generating mechanism represents a different outcome scenario, which is modelled dependent on the medicine dispensing and covariates.

This document describes one of the data-generating mechanisms, however all will follow the same basic process, the difference occurs in the regression coefficients and covariates related to the outcome generation.

#### Simulated dataset

Three medicines will be assigned, medicines A and B are of the same medicine type (e.g., biologics) and represent existing treatments for a condition (e.g., rheumatoid arthritis); a new medicine, X (e.g., JAKi), becomes available for treatment of the same condition as A and B. The dataset represents new use of medicines A and B between January 2019 and December 2024. Medicine X is only available from January 2022. It is assumed that all medicines are dispensed in quantities of 30, taken one per day, however randomisation is used to inject variability in the exposure durations.

Patients that have switched treatment are prevalent new users (PNU) (treatment patterns AX, BX, ABX, BAX). Each patient that has not been treated with medicine X are potential comparator users. For the purposes of the simulation, treatment patterns will be assigned such that each group has a designated matching group. For example, treatment pattern AX with be matched with A.

When simulating the data the treatment patterns for PNU and their comparators are determined dependent on their covariates at two transition points. At the first transition a patient will either continue on their current treatment (A or B) or switch to the other available treatment. At the second transition they will either continue on their current treatment or switch to the new medicine X. For example, someone may initiate treatment with medicine A, at the first transition they continue on A, at the second transition they switch to medicine X, resulting in treatment pattern AX. See Table 1 for the Simulation Protocol for assignment of these parameters.

The steps to generate the initial dataset () are as follows for each individual :

1. Generate the initial treatment (.
2. Generate the three covariates (.
3. Generate result of first transition,
4. Generate result of second transition,
5. Determine treatment pattern, , based on
6. Generate the treatment start date between January 1, 2019, and December 31, 2021.
7. (For treatment patterns: AB, ABX, BA, BAX) generate the date of switch between medicine A and B between January 1, 2022, to December 31, 2024. This will be known as the swap date.
8. Generate the switch date between January 1, 2022, to December 31, 2024. For treatment patterns AX, ABX, BX and BAX this is the date they switched to medicine X; for treatment patterns A, AB, B and BA this is the date that they could have switched to medicine X but instead continued with the treatment received at that time. For treatment patterns: AB, ABX, BA, BAX, select the earliest date between the swap and switch dates as the swap date, the other date is the switch date.
9. Determine the time on treatment (switch date minus start date).
10. (For treatment patterns: AB, ABX, BA, BAX) determine the time on first medicine (swap date minus start date) and the time on second medicine (switch date minus swap date).
11. Determine the proportion of time spent on medicine A and medicine B, using the values from step 10, or for treatment patterns A and AX, time on treatment A is 1.0, time on treatment B is 0, and vice versa for treatment patterns B and BX. Depending on treatment pattern determine the which is the first proportion () and the second proportion ().

Table 1. Simulation protocol for variables used to generate the dataset

|  |  |  |
| --- | --- | --- |
| **Variable** | **Notation** | **Simulation Protocol** |
| Initial treatment: (0 = medicine A, 1 = medicine B) |  |  |
| Age at treatment initiation, years |  |  |
| Disease Activity Score (DAS), (low, moderate, high) |  |  |
| Concomitant steroids: 1 = yes, 0 = no |  |  |
| Switching from initial treatment: 1 = yes, 0 = no |  |  |
| Switching at second transition: 1 = yes, 0 = no |  |  |
| Time to event, years  (see below for details) |  |  |
| Outcome: 1 = yes, 0 = no |  |  |

Treatment patterns are allocated as shown in Table 2 based on the values for :

Table 2. Treatment patterns that can be allocated

|  |  |  |
| --- | --- | --- |
| **Treatment pattern,** |  | **Prevalent new user, P** |
| A | 0, 0, 0 | No |
| AX | 0, 0, 1 | Yes |
| AB | 0, 1, 0 | No |
| ABX | 0, 1,1 | Yes |
| B | 1, 0, 0 | No |
| BX | 1, 0, 1 | Yes |
| BA | 1, 1, 0 | No |
| BAX | 1, 1, 1 | Yes |

To account for the treatment history of prevalent new users, the PNUD utilises a multi-step matching strategy. This is implemented by first defining an ‘exposure set’ consisting of all potential comparators to the prevalent new user who are at the same point in the course of disease as measured by their time on treatment, number of prescriptions or some treatment related factor. Time-conditional propensity scores are then computed for all individuals in the exposure set to identify the closest match for the prevalent new user. These propensity scores are time conditional since they are calculated using patient characteristics measured at the time of entering the exposure set.

For each of the prevalent new users (PNU), create an exposure set with all comparators that meet the following conditions:

1. Time on treatment for the comparator will be within 90 days of the PNU time on treatment
2. The comparator will have the same type of treatment prior to switching to X (i.e., A will match with AX, BA will match with BAX)
3. The PNU and comparator will have the same proportion of prior treatment for medicine A and medicine B.

Calculate the time conditional propensity score for all individuals using a logistic regression model, conditional on the exposure set, adjusted for the three covariates (

The observation with the propensity score (PS) closest to that of the PNU in each exposure set is the match. Prior to matching, ensure the positivity assumption is met (the PNU PS is between the lowest and highest PS within the exposure set) and that only observations that have not already been matched are assigned as a match.

Note that although the initial dataset consisted of individuals, not all PNU will be matched so the number of individuals for the rest of this section will be approximately half, e.g., .

The PNUs and their matches can now be assigned their event and time to event. The event times ()are simulated for each individual based on a Weibull distribution using the cumulative hazard inversion method1:

Where , is the scale parameter, is the shape parameter, is the vector of desired regression coefficients, and is the vector of covariates. The values of and vary depending on the outcome scenario. For the preliminary example the values are:

With corresponding values:

The true treatment effect (as induced by the calculation of ) is .

To simulate the scenario where only a proportion of individuals experience the event we introduced the presence of censoring. To incorporate censoring, we modified the data-generating process so that for each individual we simulated an event time and a censoring time. Censoring times were simulated from an exponential distribution. For each individual, the observed time to event was the minimum of the simulated event time and the simulated censoring time. Individuals that are not censored will experience the event (

The values of the event time and event status are appended to the originally generated dataset, however due to the matching this means that ~10,000 individuals will not have an assigned event and event time. The unassigned will have biased values injected using the same calculation for as above, however, the values for and will be:

With corresponding values:

The biased treatment effect for these individuals (as induced by the calculation of ) is .

This is now a complete data set ready to compare three matching strategies. Note that the final project will be assessing nine matching strategies, only three are described here.

### Matching strategies

Similarly to the process described above exposure sets will be generated and time-conditional propensity scores will be used to identify comparators to match to the PNU. However, this time three matched study cohorts will be generated based on three different strategies.

In the first, exposure sets use just the first criteria listed above, time on treatment.

In the second, exposure sets use time on treatment and same treatment pattern.

In the third, exposure sets use time on treatment, treatment pattern and same proportion of each prior medicine.

Each simulated matched study cohort will be analysed using Cox proportional hazards models. The estimated treatment effect, will be extracted from the fitted model.

### Performance Measures

We will assess bias, coverage, empirical and model-based standard errors for .

Our primary performance measure is the bias of the treatment effect estimate. It is estimated by

 Where is the true treatment effect () and is the effect estimate from simulation .

Our secondary performance measures are empirical standard error and coverage. Empirical standard error is estimated by

where is the mean of across repetitions.

Coverage is estimated by

We will report Monte Carlo uncertainty in tables (MCSEs next to the estimand performance measures) and in plots (error bars with MCSE around estimated performance measures).

We will perform 1,600 repetitions per condition. We determined this number by aiming for a MCSE of 0.005 for the bias. We will assume that , meaning that (A conservative estimate based on the initial small simulation run.) Given that

This implies that we need 1600 repetitions. If coverage of all methods is 95%, the implication of using is

With 50% coverage, the MCSE is maximised at 1.25. We consider this satisfactory and so proceed with (to be revised if, for example, )