T2D

2024-10-18

Inverstigation of the effect of drug on death

The overall aim of the project is to investigate how to analyse if a drug which has an established beneficial effect on an intermediate event (onset of diabetes is delayed for people taking the drug) has an additional benefit on a terminal event (death).

Intuitively it seems that we have answered this question by answering the specific questions

- What is the effect of the drug on the probability of death?
- What is the effect of the drug on the probability of death among T2D subjects?

However, assume that it is also well-known that the intermediate event is a risk factor for the terminal event. Then the effect of the drug on the terminal event could be 100% mediated through the intermediate event. The (naively asked) question is if using and having used the drug helps people when they get diabetes. The major challenge is the selection of the people in the study over time: people are randomized at the start of the study but at the onset of diabetes they are not randomized anymore:

- people who are already dead are not there anymore but if the drug has an effect on death then the people who take the drug live longer and hence are more likely to get diabetes.
- people are older in the treatment arm at the onset of diabetes due to the beneficial effect of the drug on preventing diabetes.

In the following we simulate healthcare data from a setting with 3 different types of events: Death, Censoring and T2D. Death and Censoring are terminal events and each individual can be diagnosed with T2D once. There are two baseline covariate L_0 and L_1, and a baseline treatment A_0 that affects the intensities of the different events. The covariate L_0 is uniform on $(40, 60)$ and represents age. The binary covariate L_1 represents sex. The intensities of the different events follow the form

$$\lambda_x(t) = R^x(t)\lambda_0(x,t)\phi(x,t)$$

The function $R^x(t)$ is an at risk indicator. The baseline hazard is specified as

$$\lambda_0(x,t) = \eta_x \nu_x t^{\nu_x - 1}$$

And

$$\phi(x,t) = \exp\left(\frac{\beta_{L_0,x}}{50} L_0 + \beta_{L_1,x} L_1 + \beta_{A_0,x} A_0 + \beta_{L,x} L\right)$$

Probability of getting T2D

Let T be the random variable representing the time of being diagnosed with T2D. We in the following estimate

$$P(T \leq 5, A_0 = 1) \quad , \quad P(T \leq 5, A_0 = 1)$$

By

$$\hat{P}(T \leq 5, A_0 = 1) = \frac{\sum_{i=1}^{N} 1(T_i \leq 5, A_0 = 1)}{N} \quad , \quad \hat{P}(T \leq 5, A_0 = 0) = \frac{\sum_{i=1}^{N} 1(T_i \leq 5, A_0 = 0)}{N}$$

Where T_i is the time of the T2D diagnose of individual i.

We simulate from the T2D setting described above. We let the effect of T2D on Death, $\beta_{L \to D}$, vary from 0 to 1 by 0.1. For each value of $\beta_{L \to D}$ we generate three data sets, where the effect of the initial treatment on the development of T2D is 0, -0.5 and -1 respectively. For each data set we estimate the probabilities $P(T \leq 5, A_0 = 1)$ and $P(T \leq 5 A_0 = 1)$

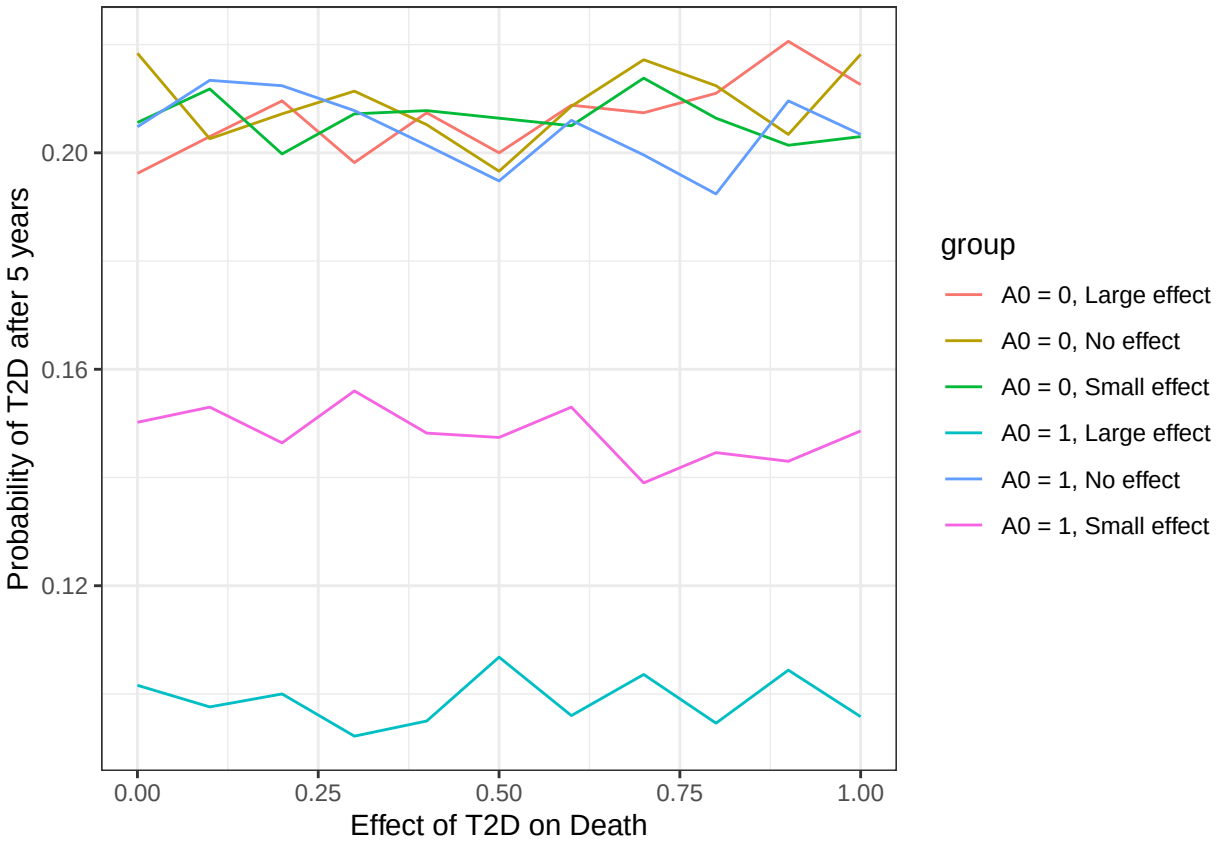
```
estimator1 <- function(data, N) {
  # Finding all the T2D people and setting T_0 to debut time of diabetes
    T2D_events <- data[Delta == 3]

    # Save estimate of absolute risk of developing T2D for people on Treatment and people not on Treatm
    return(c(nrow(T2D_events[A0 == 0 & Time < 5])/N, nrow(T2D_events[A0 == 1 & Time < 5])/N))
}
```

```
res1 <- compare_effects(estimator = estimator1, N = 5000, beta_L_D = seq(0,1,by = 0.1))
```

```
## [1] 1
## [1] 2
## [1] 3
## [1] 4
## [1] 5
## [1] 6
## [1] 7
## [1] 8
## [1] 9
## [1] 10
## [1] 11
```

```
plot_compare(res1) +
      ylab("Probability of T2D after 5 years")+
      xlab("Effect of T2D on Death")
```



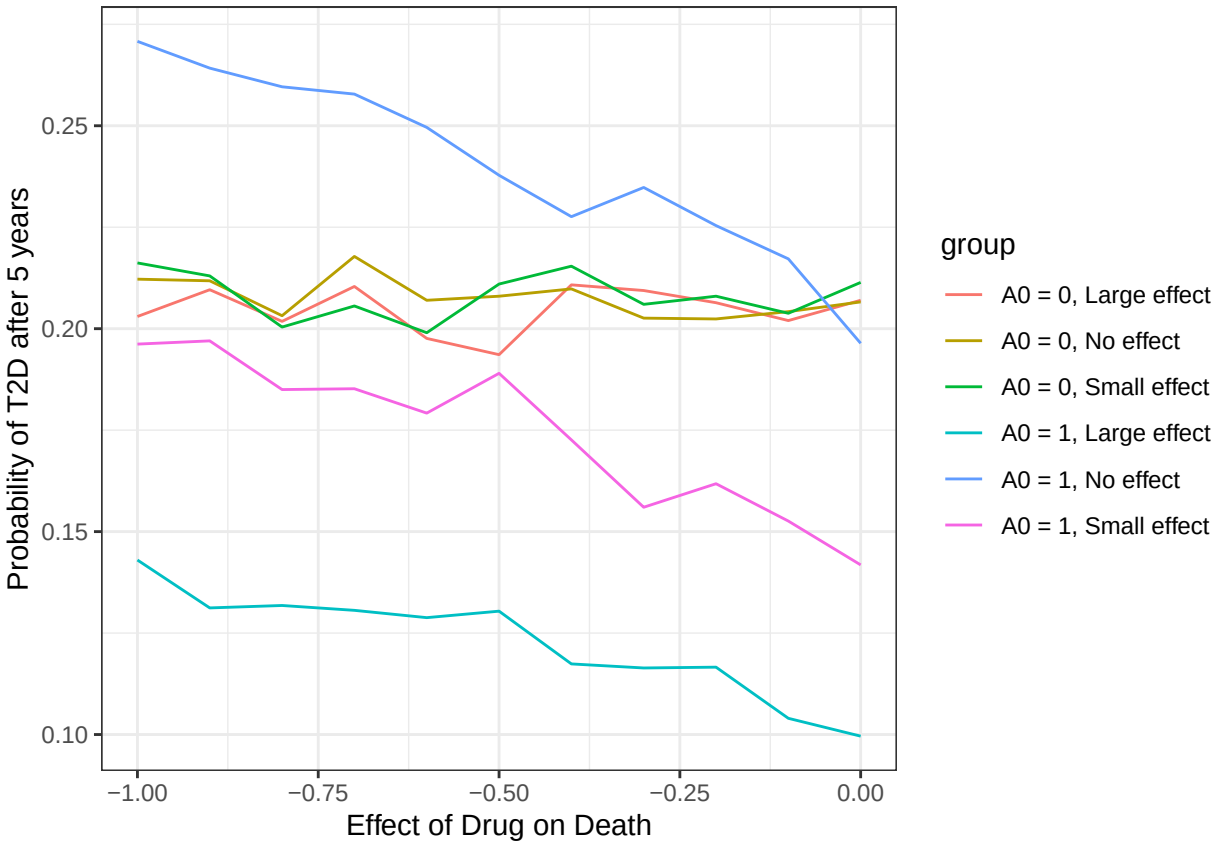
We see that the probability of getting T2D is reduced by the drug (as we have simulated), and the reduction does not depend on the effect T2D has on Death.

We now vary the effect of the Drug on Death $\beta_{A_0 \to D}$, and see how this affects the T2D probability.

```
res2 <- compare_effects(estimator = estimator1, N = 5000, beta_A0_D = seq(-1,0,by = 0.1))
```

```
## [1] 1
## [1] 2
## [1] 3
## [1] 4
## [1] 5
## [1] 6
## [1] 7
## [1] 8
## [1] 9
## [1] 10
## [1] 11
```

```
plot_compare(res2, plot_no = 1, diff_betas <- seq(-1,0,by = 0.1)) +
      ylab("Probability of T2D after 5 years")+
      xlab("Effect of Drug on Death")
```



For the people without initial treatment, the probability of T2D is constant throughout the graph. For the people on treatment the probability of getting T2D after 5 years is largest when the drug has an large effect on Death. Because less people die, there is a larger population ready to develop T2D. As the effect of the drug on T2D increases the probability of getting T2D gets smaller for the people in the Treatment arm.

Survival probability 3 years after debut of T2D

Let U be the random variable representing time of death after being diagnosed with T2D. We in the following estimate

$$P(U > 3, A_0 = 1, L_0 = 0.5, L_1 = 0) \quad , \quad P(U > 3, A_0 = 0, L_0 = 0.5, L_1 = 0)$$

We estimate the survival probabilities by simulating data as previously. For each data set we estimate the survival function with the Kaplan Meyer estimator. For each fit we predict the survival probability 3 years after debut of T2D for an observation with $L_0 = 0.5$, $L1 = 0$ and $A_0 = 0$ and $A_0 = 1$. That is a person on initial treatment, and a person not on initial treatment.

```
estimator2 <- function(data, N) {
  # Finding all the T2D people and setting T_0 to debut time of diabetes
    T2D_events <- data[Delta == 3]
    T2D_peeps <- data[ID %in% T2D_events$ID]
    T2D_peeps[, Time_T2D := Time - min(Time), by = ID]
    T2D_peeps <- T2D_peeps[Delta != 3]
    T2D_peeps[, Status := Delta == 1]

    # Kaplan meyer fit
    fit <- prodlim(Hist(Time_T2D, Status) ~ L0 + L1 + A0, data = T2D_peeps)
```

```
    # Save estimate of survival probability
    preds <- predict(fit, times = 3, newdata = data.frame(L0 = c(0.5,0.5), A0 = c(0,1), L1 = c(0,0)))
    return(c(preds$`L0=0.5, L1=0, A0=0`, preds$`L0=0.5, L1=0, A0=1`))
}

res3 <- compare_effects(estimator = estimator2, N = 10000, beta_L_D = seq(0,1,by = 0.1))
```

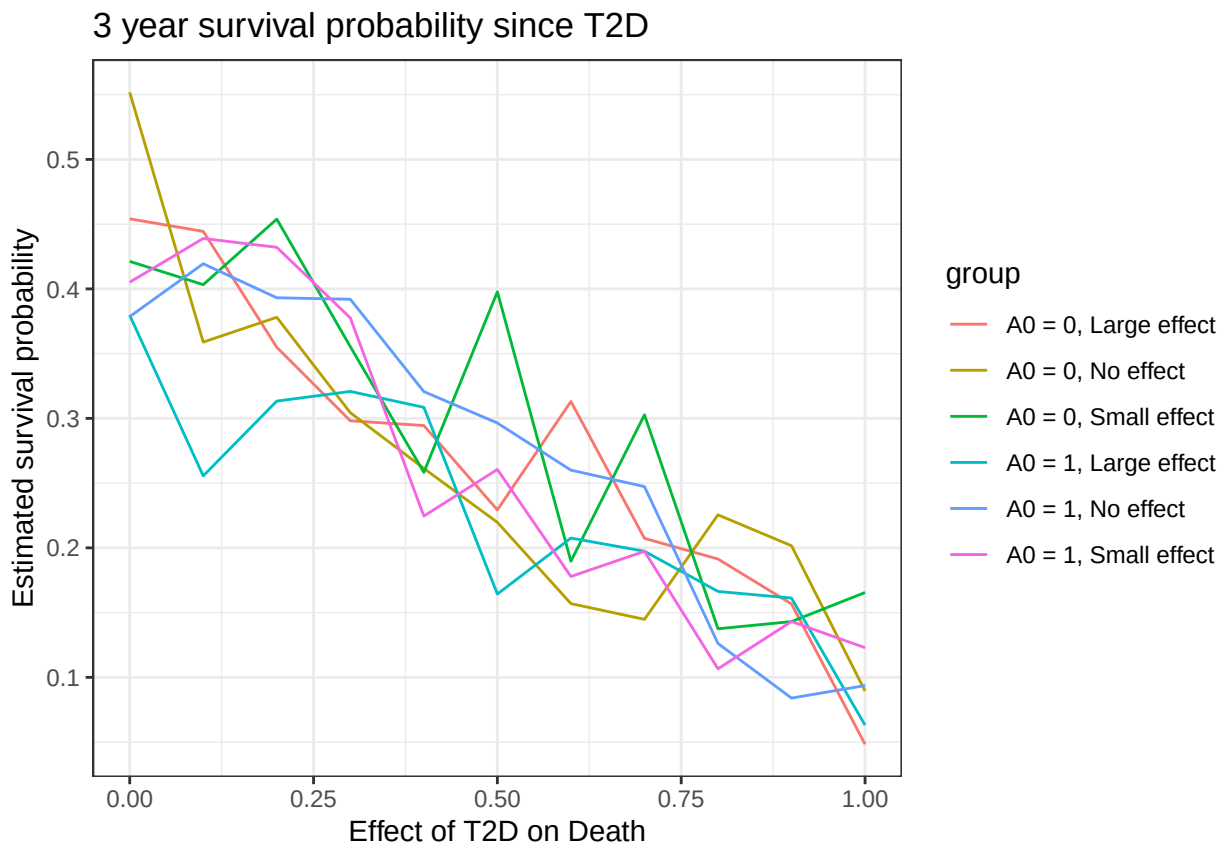
```
## [1] 1
## [1] 2
## [1] 3
## [1] 4
## [1] 5
## [1] 6
## [1] 7
## [1] 8
## [1] 9
## [1] 10
## [1] 11
```

```
res4 <- compare_effects(estimator = estimator2, N = 10000, beta_A0_D = seq(-1,0,by = 0.1))
```

```
## [1] 1
## [1] 2
## [1] 3
## [1] 4
## [1] 5
## [1] 6
## [1] 7
## [1] 8
## [1] 9
## [1] 10
## [1] 11
```

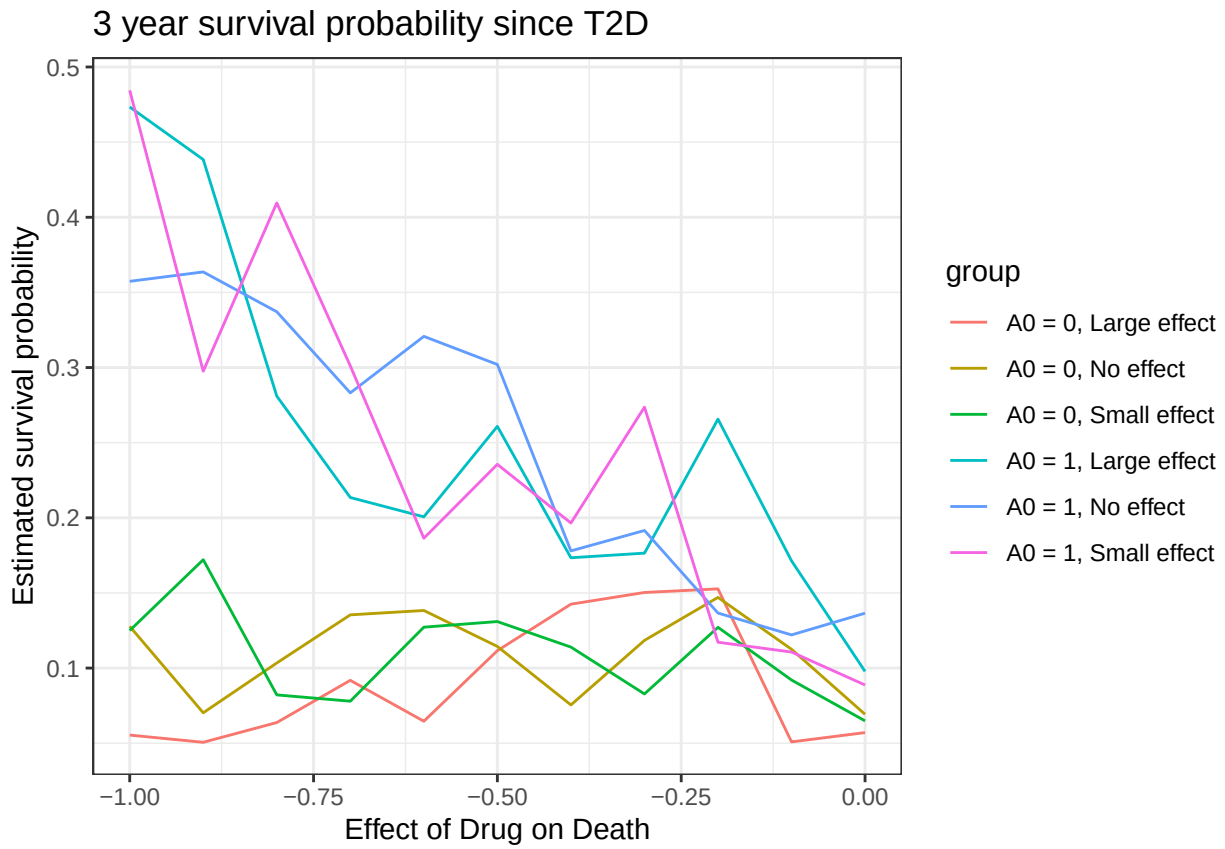
Visual illustration

```
plot_compare(res3)+
      ylab("Estimated survival probability")+
      xlab("Effect of T2D on Death")+
      labs(title = "3 year survival probability since T2D")
```



We see that as the effect of T2D on Death increases, the probability of survival decreases. It does so equally for both treatment and placebo group and for all effects of the drug on T2D. This makes sense as the drug only affects death through a later T2D diagnose. After the diagnose, both groups stand equally well. Hmm... what about the difference in age between the two groups?

```
plot_compare(res4, diff_betas = seq(-1,0,by = 0.1))+
      ylab("Estimated survival probability")+
      xlab("Effect of Drug on Death")+
      labs(title = "3 year survival probability since T2D")
```



When the drug has an effect on Death, the survival probability after T2D diagnose is larger in the Treatment Group. Quite interestingly this survival probability does not depend on the size of the effect of the drug on T2D. This is weird since the larger the effect of the drug, the older the T2D treatment group. As the effect of the drug on Death decreases, the survival probability approaches the one in the placebo group.

Covariate differences

```
data0 <- sim_data_setting2(N = 10^4, beta_L_D = 1, beta_A0_D = -0.1,
                           beta_L0_L = 1, beta_A0_L = 0)
data0.5 <- sim_data_setting2(N = 10^4, beta_L_D = 1, beta_A0_D = -0.1,
                           beta_L0_L = 1, beta_A0_L = -0.5)
data1 <- sim_data_setting2(N = 10^4, beta_L_D = 1, beta_A0_D = -0.1,
                           beta_L0_L = 1, beta_A0_L = -1)

# Find age difference between the two groups after T2D diagnose

T2D_events0 <- data0[Delta == 3]
T2D_events0[, .(L0_mean = mean(L0)), by = A0]
```

```
##       A0  L0_mean
##    <int>    <num>
## 1:     0 50.16391
## 2:     1 49.94636
```

```
T2D_events0[, .(Time_mean = mean(Time)), by = A0]
```

```
##       A0 Time_mean
##    <int>     <num>
## 1:     0  1.432213
## 2:     1  1.494803
```

```
data0.5[, .(L0_mean = mean(L0)), by = L]
```

```
##        L  L0_mean
##    <num>    <num>
## 1:     0 49.97637
## 2:     1 50.20750
```

```
T2D_events0.5 <- data0.5[Delta == 3]
T2D_events0.5[, .(L0_mean = mean(L0)), by = A0]
```

```
##       A0  L0_mean
##    <int>    <num>
## 1:     1 50.20224
## 2:     0 50.21174
```

```
T2D_events0.5[, .(Time_mean = mean(Time)), by = A0]
```

```
##       A0 Time_mean
##    <int>     <num>
## 1:     1  1.798118
## 2:     0  1.472623
```

```
T2D_events1 <- data1[Delta == 3]
T2D_events1[, .(L0_mean = mean(L0)), by = A0]
```

```
##       A0  L0_mean
##    <int>    <num>
## 1:     0 50.00203
## 2:     1 50.14952
```

```
T2D_events1[, .(Time_mean = mean(Time)), by = A0]
```

```
##       A0 Time_mean
##    <int>     <num>
## 1:     0  1.407090
## 2:     1  2.070898
```

Apparently both groups have same average baseline covariate, but as the effect of the drug increases, they get older in the Treatment group. The same average baseline covariate might be due to the effects of the drug on death and on development of T2D being equal.

```
# Find proportion of men and women after T2D diagnose
```

Helene forslag

Generate large data

```
N <- 50000
data_A <- sim_data_setting2(N = N, beta_L_D = 1, cens = 0)
data_A[, at_risk_cov := as.numeric(L == 0)]
```

Grouping data based on treatment

```
no_treat_group <- data_A[A0 == 0]
treat_group <- data_A[A0 == 1]
```

Fit Weibull

I feel like there is something going wrong here

```
survfit <- survreg(Surv(Time, Delta == 3) ~ L0,
                   data = no_treat_group[at_risk_cov == 1],
                   cluster = ID,
                   dist='weibull')

survfit1 <- survreg(Surv(Time, Delta == 3) ~ L0,
                   data = treat_group[at_risk_cov == 1],
                   cluster = ID,
                   dist='weibull')
```

Estimates

```
nu_est <- 1/survfit$scale; nu_est
```

```
## [1] 1.101306
```

```
eta_est <- 1/exp(survfit$coefficients[1]); eta_est
```

```
## (Intercept)
##   0.1293397
```

```
beta_l0_l_est <- - survfit$coefficients[2] / survfit$scale; beta_l0_l_est
```

```
##         L0
## 0.01919048
```

```
1/survfit1$scale
```

```
## [1] 1.105025
```

```r
1/exp(survfit1$coefficients[1])
```

```
## (Intercept)
##    0.124777
```

```r
- survfit1$coefficients[2] / survfit1$scale
```

```
##         L0
## 0.01979945
```

Generate large data set under the intervened intensity

```r
N <- 10000
data_new <- sim_data_setting2(N = N,
                              cens = 0,
                              eta = c(rep(0.1,3), eta_est),
                              nu = c(rep(1.1,3), nu_est),
                              beta_L0_L = beta_l0_l_est)
```

Generate large data set without intervened intensity

```r
data_new_no_int <- sim_data_setting2(N = N, cens = 0)
```

Proportion of subjects dying before some time τ

```r
tau <- 5
mean(data_new[Delta == 1, Time] < tau) # with intervention
```

```
## [1] 0.8766
```

```r
mean(data_new_no_int[Delta == 1, Time] < tau) # without intervention
```

```
## [1] 0.9149
```