T2D

2024-10-18

Inverstigation of T2D trial

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 time T2D diagnose. We in the following estimate

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

By

$$\hat{P}(T \leq 5|A_0 = 1) = \frac{\hat{P}(T \leq 5, A_0 = 1)}{\hat{P}(A_0 = 1)} \quad , \quad \hat{P}(T \leq 5|A_0 = 0) = \frac{\hat{P}(T \leq 5, A_0 = 0)}{\hat{P}(A_0 = 0)}$$

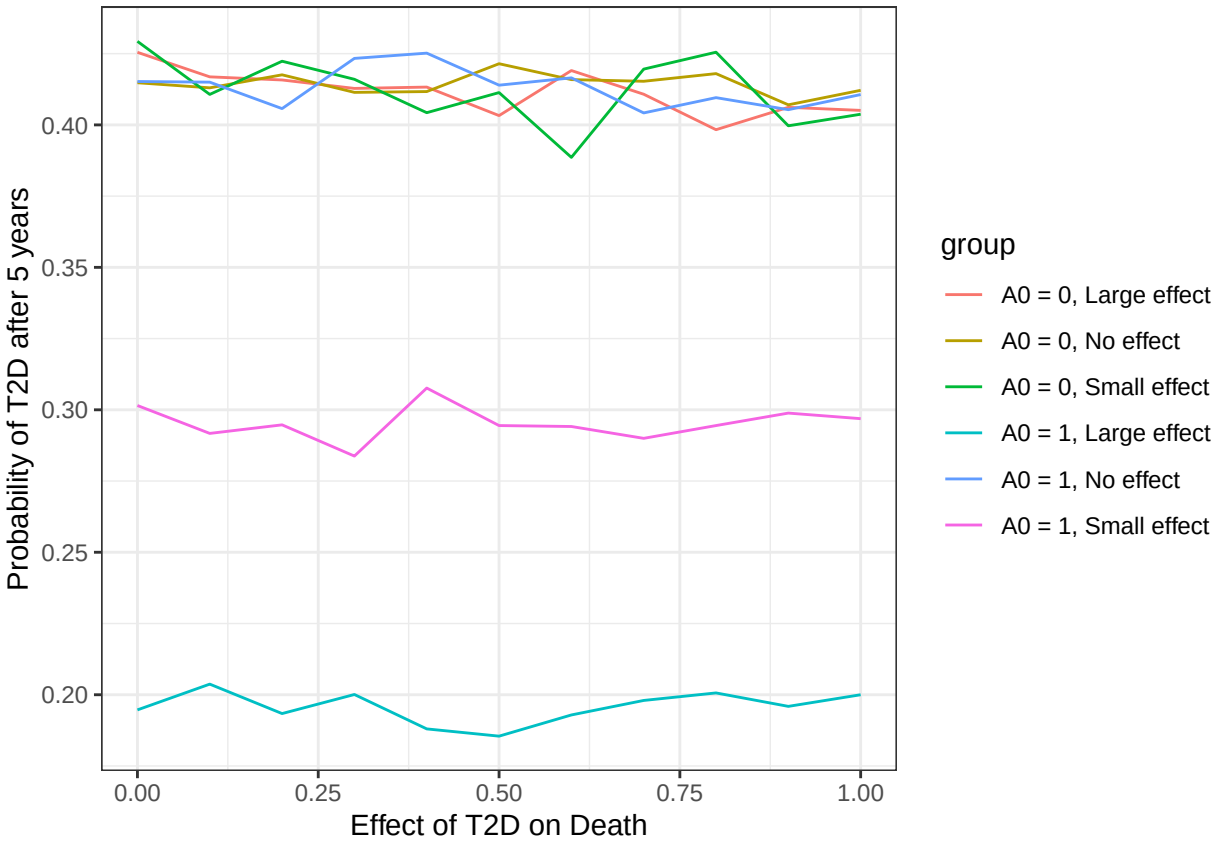
Where T_i is the time of the T2D diagnose of individual i.

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 $P(T \leq 5|A_0 = 1)$

```
N <- 5000

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

    # Estimating probability of A0 = 1 and A0 = 0
    p_a0_1 <- length(unique(data[A0 == 1, ID])) / N
    p_a0_0 <- length(unique(data[A0 == 0, ID])) / N

    p_t_0 <- (nrow(T2D_events[A0 == 0 & Time < 5]) / N) / p_a0_0
    p_t_1 <- (nrow(T2D_events[A0 == 1 & Time < 5]) / N) / p_a0_1

    return(c(p_t_0, p_t_1))
}
```

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

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

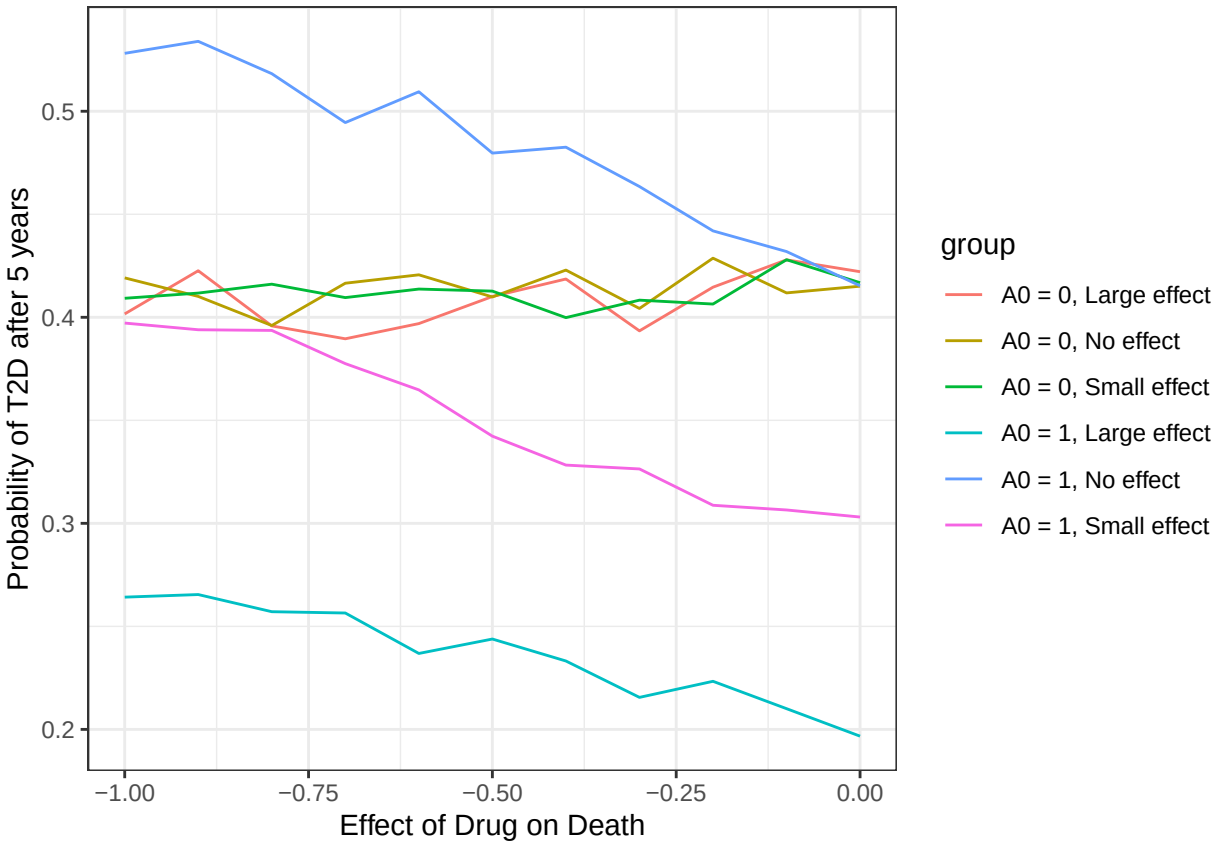


We see that the probability of getting T2D is reduced by the drug (as we have simulated), and the reduction does not depend on the effect T2D has on Death.

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 = N, beta_A0_D = seq(-1,0,by = 0.1))
```

```
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

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

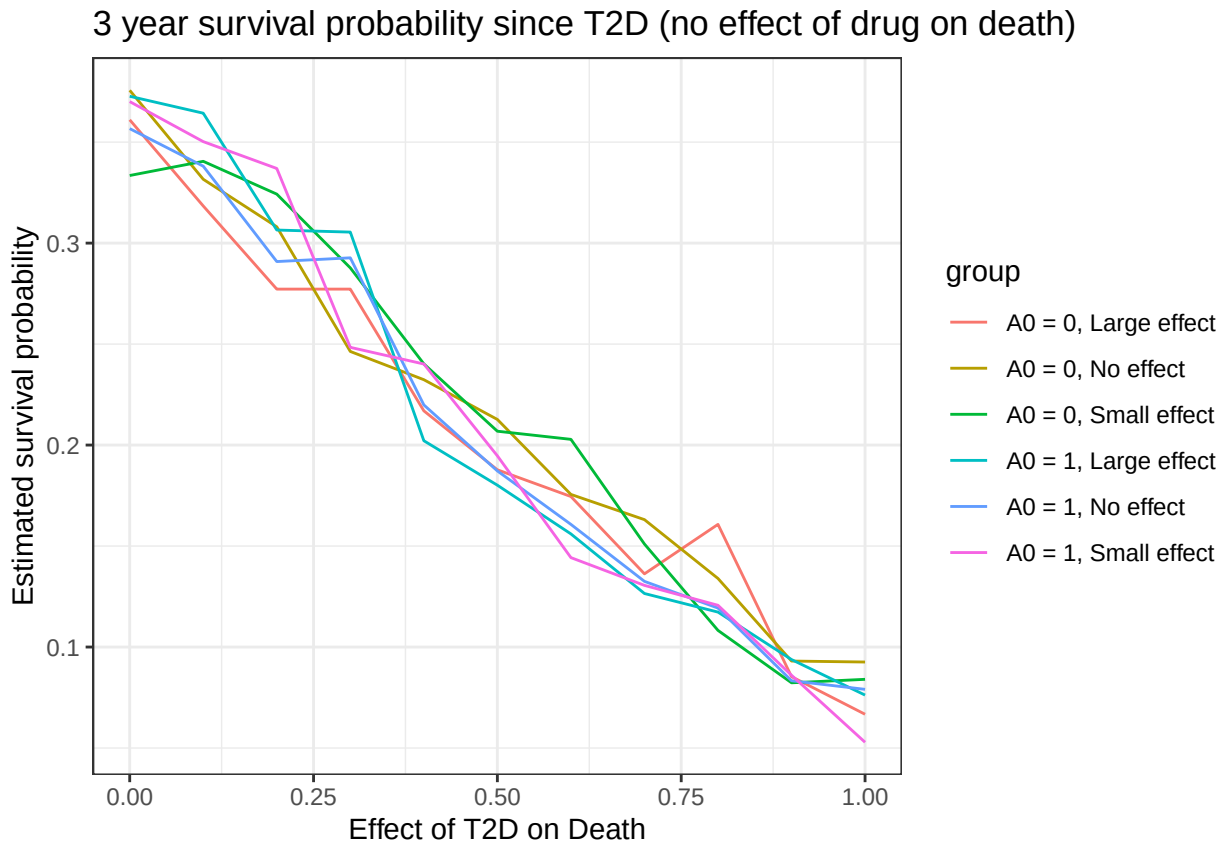
We estimate the survival probabilities by simulating data as previously. For each data set we estimate the survival function with the Kaplan Meyer estimator. For each fit we predict the survival probability 3 years after debut of T2D for an observation with $L_0 = 0.5$, $L_1 = 0$ and respectively $A_0 = 0$ and $A_0 = 1$. That is for a person on initial treatment, and a person not on initial treatment.

```
estimator2 <- function(data, N) {
    # Finding all the T2D people
    T2D_events <- data[Delta == 3]
    T2D_peeps <- data[ID %in% T2D_events$ID]

    # Setting T_0 to debut time of diabetes
    T2D_peeps[, Time_T2D := Time - min(Time), by = ID]

    # Removing the new Time 0
    T2D_peeps <- T2D_peeps[Delta != 3]
```

```
    # Creating a status variable
    T2D_peeps[, Status := Delta == 1]

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

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

res3 <- compare_effects(estimator = estimator2, N = N, beta_L_D = seq(0,1,by = 0.1))
res4 <- compare_effects(estimator = estimator2, N = N, beta_A0_D = seq(-1,0,by = 0.1))
```

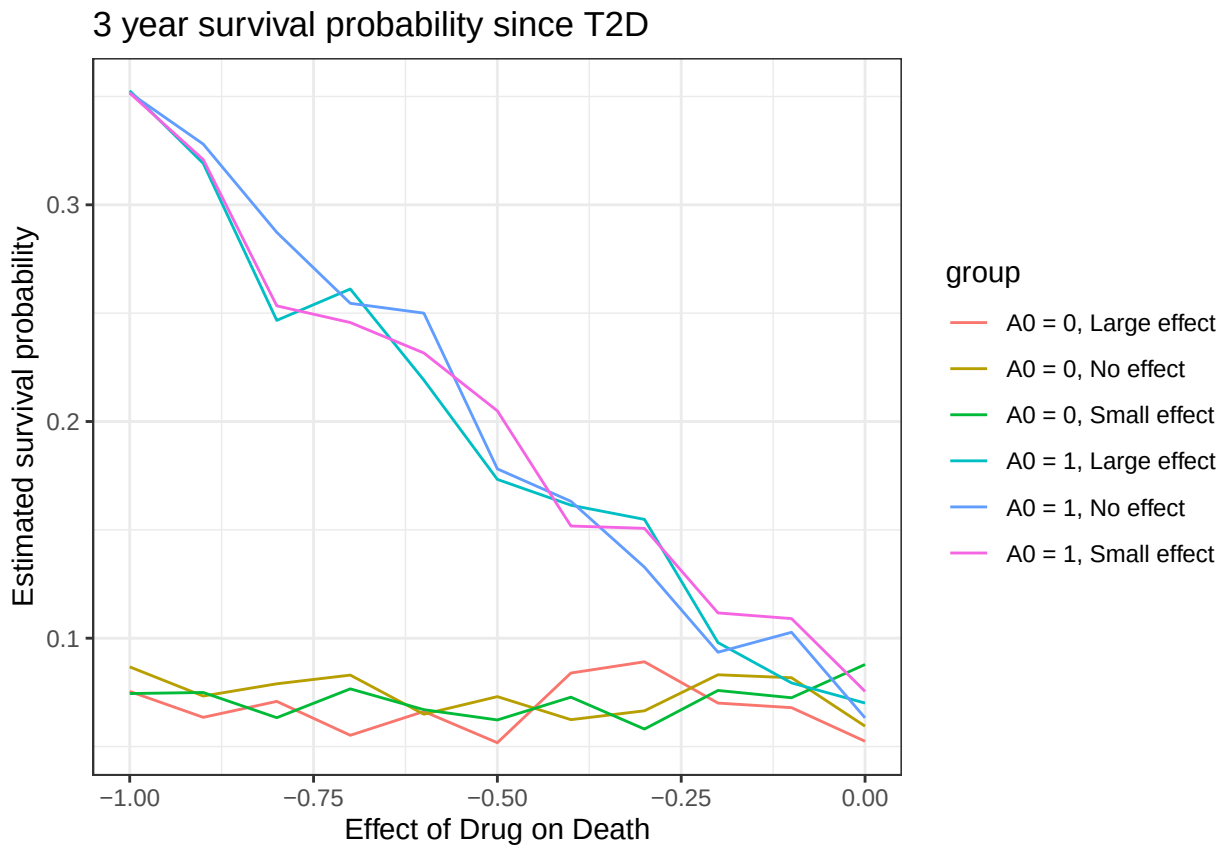
Visual illustration

```
plot_compare(res3, diff_betas= seq(0,1,by = 0.1))+
      ylab("Estimated survival probability")+
      xlab("Effect of T2D on Death")+
      labs(title = "3 year survival probability since T2D (no effect of drug on death)")
```



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 is surprising since one would think that the age difference between the two groups would play a role. We will investigate this further in the following.

```
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 one would imagine that the larger the effect of the drug, the older the T2D treatment group, the larger the intensity of Death and the smaller survival probability. This is the same effect we were pondering about above, and we will look into this below. As the effect of the drug on Death decreases, the survival probability approaches the one in the placebo group.

Interventions

Generate large data

```
data_A <- sim_data_setting2(N = 5*10^4, beta_L_D = 1, cens = 0, beta_A0_L = 1)
data_A[, at_risk_cov := as.numeric(L == 0)]
data_A_t <- trans_int_data(data_A)

survfit <- coxph(Surv(tstart, tstop, Delta == 3) ~ I(L0 / 50) + L1 + A0,
                         data = data_A_t[at_risk_cov == 1],
                         timefix = FALSE,
                         cluster = ID)
```

```
survfit$coefficients
```

```
##   I(L0/50)         L1         A0
## 0.98798772 0.01952001 1.00265226
```

Grouping data based on treatment

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

Fit Weibull

```
survfit <- survreg(Surv(Time, Delta == 3) ~ I(L0 / 50) + L1,
                   data = no_treat_group[at_risk_cov == 1],
                   cluster = ID,
                   dist='weibull')
```

Estimates in no treatment group

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

```
## [1] 1.092259
```

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

```
## (Intercept)
##   0.1028308
```

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

```
##  I(L0/50)
## 0.9748535
```

```
beta_l1_l_est <- - survfit$coefficients[3] / survfit$scale; beta_l1_l_est
```

```
##          L1
## 0.002165346
```

Generate large data set under the intervened intensity

```
data_new <- sim_data_setting2(N = 10^4,
                              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,
                              beta_A0_L = 0)
```

Generate large data set without intervened intensity

```r
data_new_no_int <- sim_data_setting2(N = 10^4, cens = 0)
```

Proportion of subjects dying before some time τ in treatment group

```r
tau <- 2
mean(data_new[Delta == 1 & A0 == 1, Time] < tau) # with intervention
```

```
## [1] 0.546398
```

```r
mean(data_new_no_int[Delta == 1 & A0 == 1, Time] < tau) # without intervention
```

```
## [1] 0.5395741
```

Basically the same...