T2D Scenarios and Estimators

2024-10-18

The Aim

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 question

- 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.

The set up

Simulations are carried out in a setting where patients can experience 2 different events: Death (1) and Type-2-Diabetes (2). Death is a terminal events and each individual can be diagnosed with T2D once. There is a baseline covariate L_0, and a baseline treatment A_0. The covariate L_0 is uniform on $(40, 60)$ and represents age. Effects of covariates L_0, A_0 and T2D on the intensities of the events can be specified by the user by the beta arguments of the simulation function. 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

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

The Death process has $\eta_{death} = 0.3$ and $\nu_{death} = 1.3$, while T2D has $\eta_{T2D} = 0.1$ and $\nu_{T2D} = 1.1$. The $\phi(x,t)$ function is defined as

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

Below we will for two different settings look at three different ways of evaluating the effect of treatment on Death. We will look at

- Proportion of dead treatment and placebo patients a year after T2D diagnose
- Kaplan-Meier estimated survival probability a year after T2D diagnose for respectively treatment and placebo patients.
- An intervention based estimate.

When we calculate the proportion of dead we use data without censoring, and when we estimate the survival probability we use data with censoring. We simulate $2 \cdot 10^4$ individuals per estimate.

Setting A

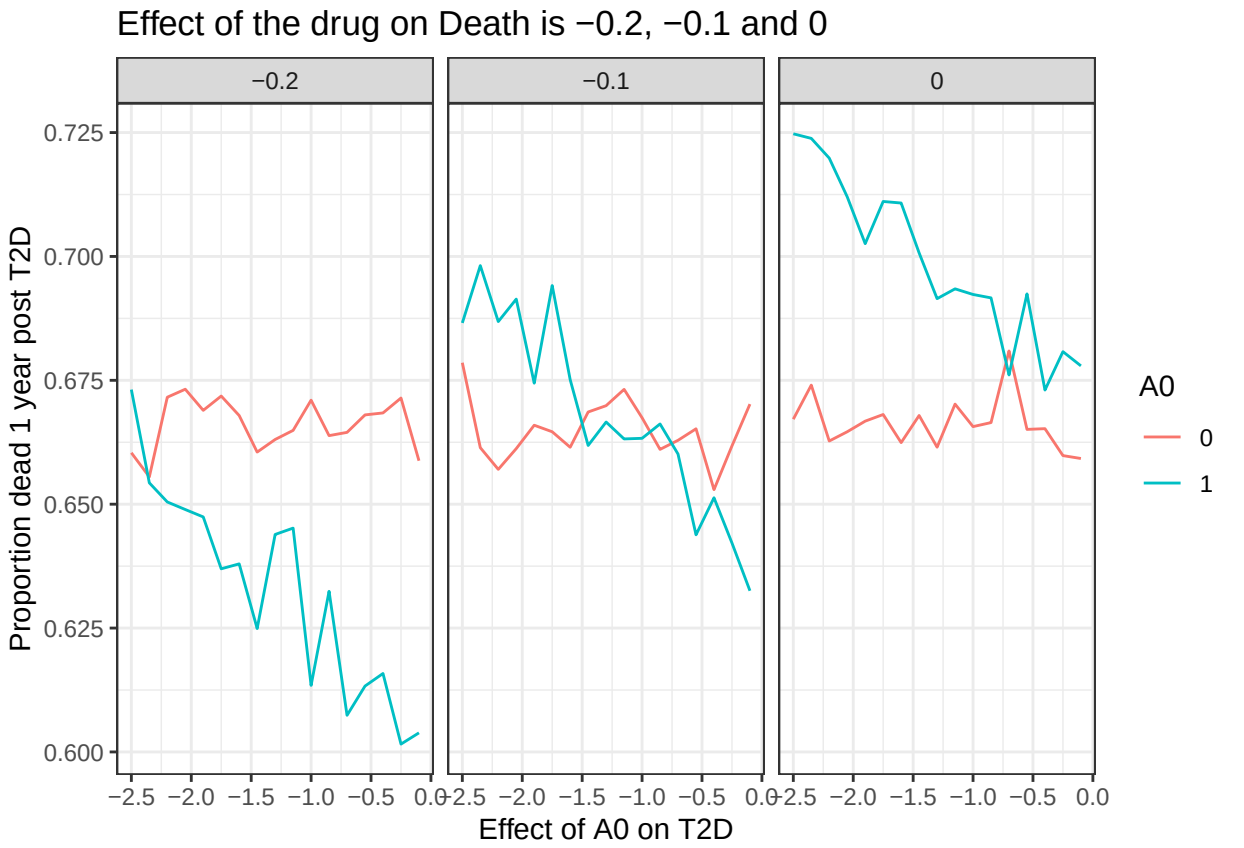
Below we calculate the proportion of dead T2D patients 1 year post T2D diagnose. We do this in three different scenarios: where the effect of the drug on Death is 0, -0.1 and -0.2. The effect of A_0 on T2D is let to vary from -2.5 to 0 by 0.15. The effect of L_0 on T2D is large ($= 2$). And the effect of T2D on death is moderate ($= 1$). There is no effect of L_0 directly on death.

```
N <- 2 * 10^4

estimator1 <- function(data, N) {
  # T2D events
  T2D_events <- data[Delta == 3]
  # T2D people
  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]

  # Proportion of treatment and placebo patients who have died before 1 year after T2D diagnose
  prop_treat <- nrow(T2D_peeps[Time_T2D < 1 & Delta == 1 & A0 == 1]) / length(unique(T2D_peeps[A0 == 1]$
  prop_plac <- nrow(T2D_peeps[Time_T2D < 1 & Delta == 1 & A0 == 0]) / length(unique(T2D_peeps[A0 == 0]$I

  # The proportion dead
  return(c(prop_plac, prop_treat))
}

res1 <- compare_effects(estimator = estimator1,
                        N = N,
                        eta = c(0.1,0.3,0.1,0.1),
                        nu = c(1.1,1.3,1.1,1.1),
                        beta_L0_L = 2,
                        beta_A0_L = seq(-2.5, 0, by = 0.15),
                        beta_L_D = 1,
                        beta_L0_D = 0)
```

Proportion Dead

Below the plots show the proportion dead for different effects of the drug on Death.

```
plot_compare(res1, diff_betas = seq(-2.5, 0, by = 0.15))+
  ylab("Proportion dead 1 year post T2D")+
  xlab("Effect of A0 on T2D")+
  labs(title = "Effect of the drug on Death is -0.2, -0.1 and 0")
```



We see that the proportion of dead is larger in the treatment group than the placebo group, when the drug does not have an effect on Death. When the drug has a small effect the proportions are equally large. And only when the beneficial effect of the drug is large, the proportion of dead is smaller. It seems our hypothesis about the selection of individuals is true. It further seems, in all three scenarios, as if the proportion of dead in the treatment group is decreasing as the effect of A_0 on T2D decreases, further supporting the selection hypothesis.

Survival Probability

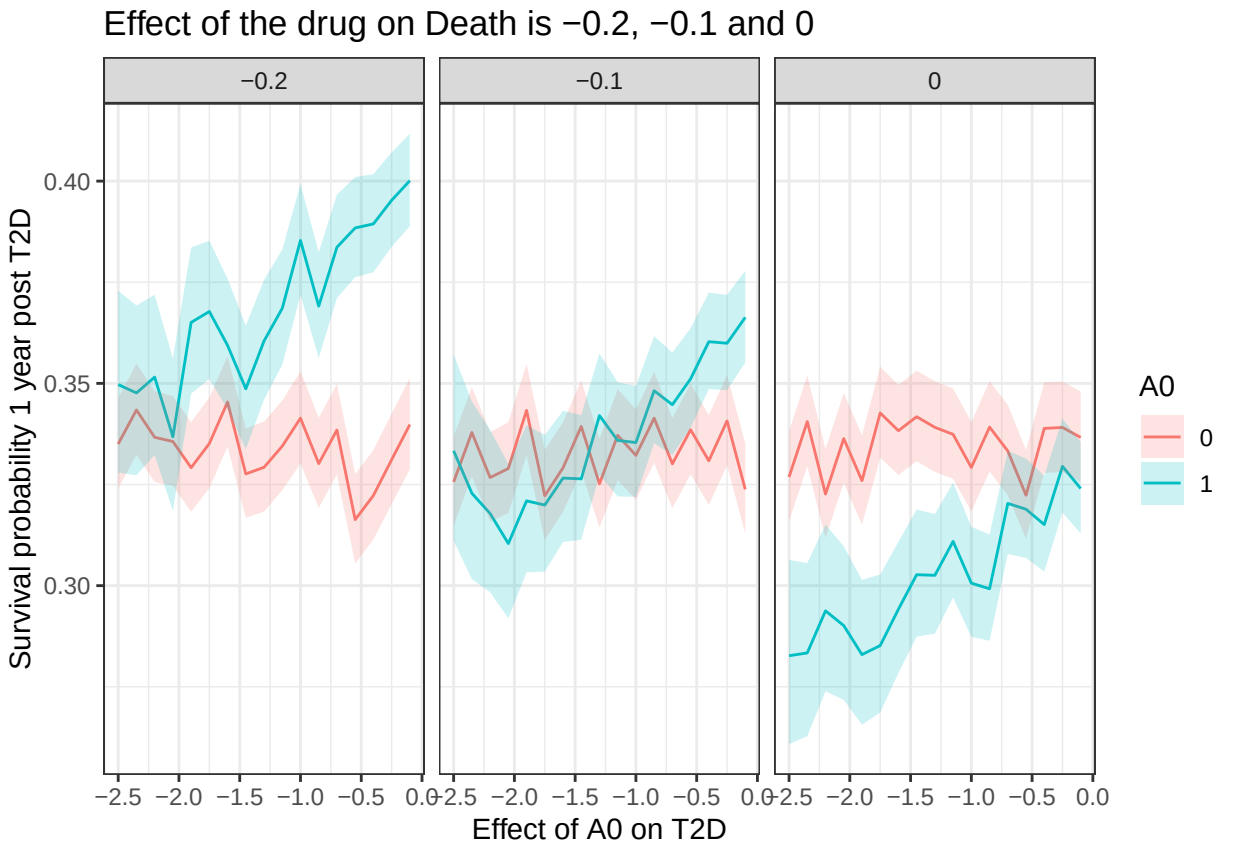
We now instead turn to estimating the survival probability instead of the proportion of dead. The situation is the same. Now data is just censored.

```
estimator2 <- function(data, N) {
  # T2D events
  T2D_events <- data[Delta == 3]
  # T2D people
  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]

  # Kaplan meyer fit
  fit <- survfit(Surv(Time_T2D, Delta == 1) ~ A0, data = T2D_peeps)
```

```
  # Save estimate of survival probability, and confidence limits
  preds <- summary(fit, times = 1, data.frame = TRUE)

  # The KM estimate, SE's, lower CI, and upper CI is returned
  return(c(preds[,5], preds[,7], preds[,9], preds[,10]))
}

res2 <- compare_effects(estimator = estimator2,
                        N = N,
                        cens = 1,
                        eta = c(0.1,0.3,0.1,0.1),
                        nu = c(1.1,1.3,1.1,1.1),
                        beta_L0_L = 2,
                        beta_A0_L = seq(-2.5, 0, by = 0.15),
                        beta_L_D = 1,
                        beta_L0_D = 0)
```

Below the Kaplan-Meier estimates as well as the .95 CI's of the survival probability 1 year after T2D diagnose are plotted.

```
estimates <- res2[,c(1,2,9,10,17,18)]
lower <- res2[,c(5,6,13,14,21,22)]
upper <- res2[,c(7,8,15,16,23,24)]

plot_compare(estimates = estimates, diff_betas = seq(-2.5, 0, by = 0.15),
             lower = lower, upper = upper)+
  ylab("Survival probability 1 year post T2D")+
  xlab("Effect of A0 on T2D")+
  labs(title = "Effect of the drug on Death is -0.2, -0.1 and 0")
```



The Kaplan-Meier estimates tell the same story as the proportions.

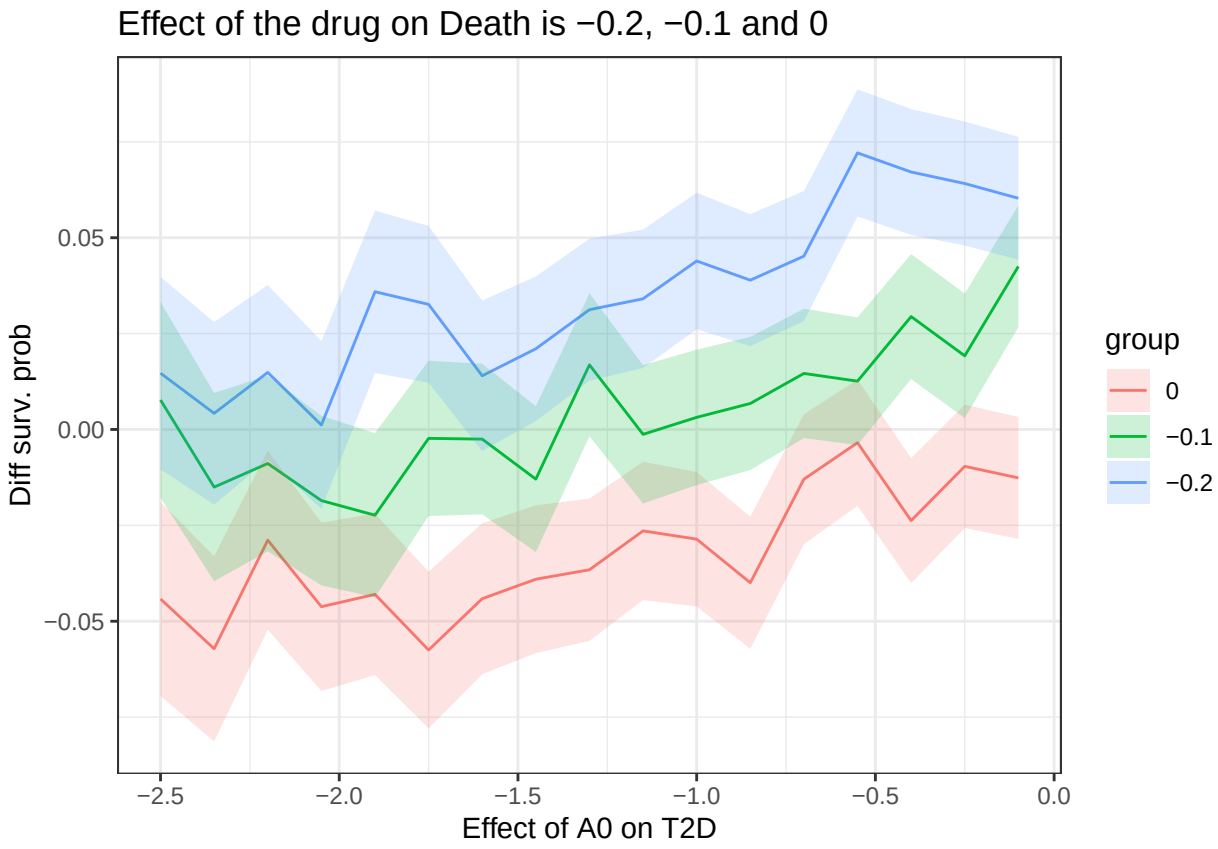
We now plot the estimated differences in survival probabilities between treatment and placebo group. We calculate the standard error of the difference in survival probability between the two groups as $SE_{diff} = \sqrt{Var(\hat{S}(t|A_0 = 1)) + Var(\hat{S}(t|A_0 = 0))}$, where the two standard errors are obtained from the KM fit from the survival package, and calculate the confidence interval as

$$\hat{S}(t|A_0 = 1) - \hat{S}(t|A_0 = 0) \pm \alpha_{0.05} SE_{diff}$$

using a normal approximation.

```
diffs <- res2[,c(2,10,18)] - res2[,c(1,9,17)]
se <- sqrt(res2[,c(3,11,19)]^2 + res2[,c(4,12,20)]^2)
upper <- diffs + qnorm(0.025)*se
lower <- diffs - qnorm(0.025)*se

plot.data <- data.table(ests = c(diffs),
           beta = rep(seq(-2.5, 0, by = 0.15), 3),
           group = factor(c(rep("0", 17), rep("-0.1", 17), rep("-0.2", 17)),
                          levels =c("0", "-0.1", "-0.2")),
           upper = c(upper),
           lower = c(lower))



ggplot(plot.data)+
  geom_line(aes(x = beta, y = ests, color = group))+
  geom_ribbon(aes(x = beta, ymin = lower, ymax = upper, fill = group), alpha = 0.2)+
  xlab("Effect of A0 on T2D")+
```

```
  labs(title = "Effect of the drug on Death is -0.2, -0.1 and 0")+
  ylab("Diff surv. prob")
```



Interventions

We want to see whether we can estimate the effect of the drug on Death more correctly with interventions. The idea is to generate a large data set. Estimate the intensity of the T2D event for the non treatment group. Generate two new data sets: one with the estimated intensity of the T2D event and one simulated as before. And at last, compare the proportion of dead people in the treatment group τ years after beginning of trial. This is done by the function `int_effect`.

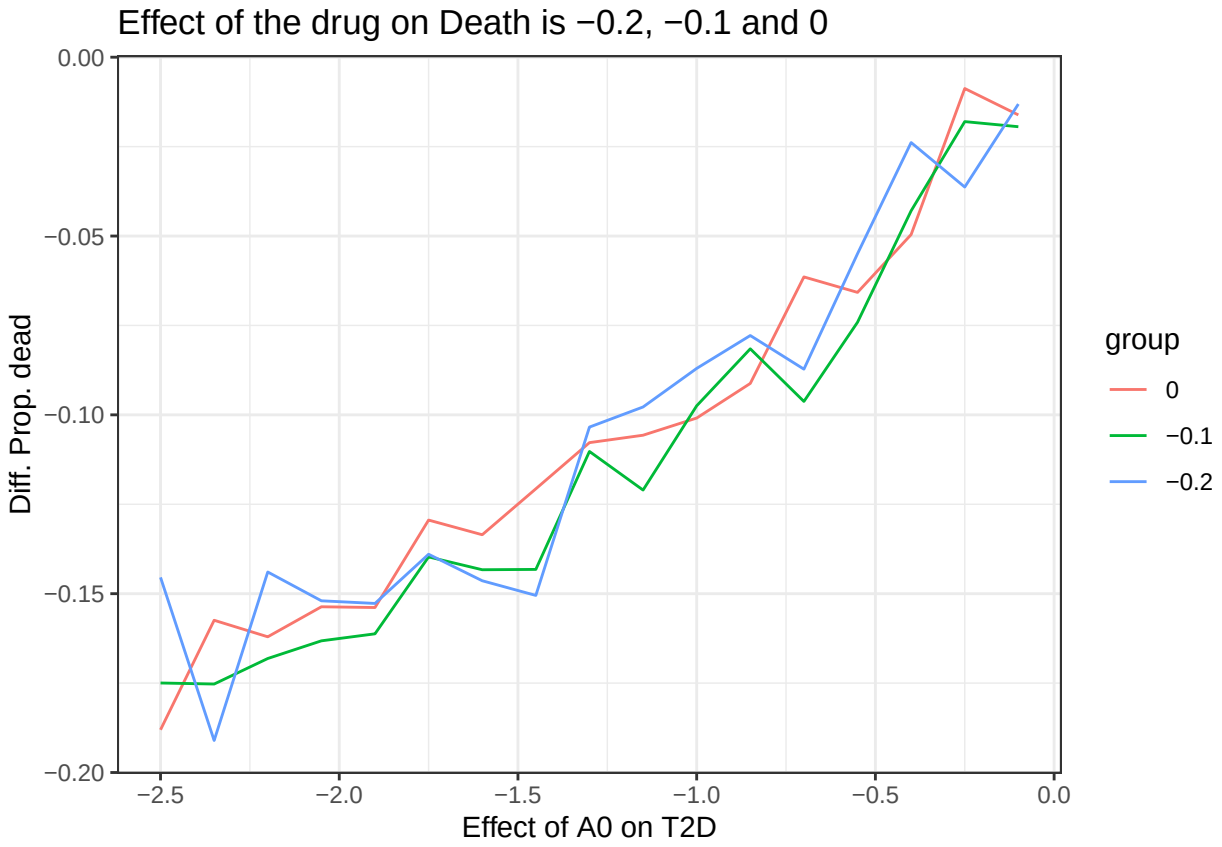
We want, as before, to estimate the intervention effect in three different scenarios: where the effect of the drug on Death is 0, -0.1 and -0.2. And where the effect of A_0 on T2D is let to vary from -2.5 to 0 by 0.15. The effect of L_0 on T2D is large ($= 2$). And the effect of T2D on Death is moderate ($= 1$). There is no effect of L_0 directly on Death. The function `compare_int_effect` estimates the intervention effect, by using the function `int_effect`, for all the different scenarios.

```
N1 <- 5000
N2 <- 5000
res3 <- compare_int_effect(N1 = N1,
           N2 = N2,
           eta = c(0.1,0.3,0.1,0.1),
           nu = c(1.1,1.3,1.1,1.1),
           beta_L0_L = 2,
           beta_A0_L = seq(-2.5, 0, by = 0.15),
           beta_L_D = 1,
           beta_L0_D = 0,
```

```
            tau = 2)
```

We plot the intervention effect as the difference in proportion of dead between the data without intervention and the intervened data:

```
diffs <- res3[,c(2,4,6)] - res3[,c(1,3,5)]

data.table(ests = c(diffs),
           beta = rep(seq(-2.5, 0, by = 0.15), 3),
           group = factor(c(rep("0", 17), rep("-0.1", 17), rep("-0.2", 17)),
                          levels =c("0", "-0.1", "-0.2"))) |>
  ggplot() +
  geom_line(aes(x = beta, y = ests, color = group))+
  xlab("Effect of A0 on T2D")+
  labs(title = "Effect of the drug on Death is -0.2, -0.1 and 0")+
  ylab("Diff. Prop. dead")
```



The difference in proportion of dead differs slightly between the three scenarios it seems, being smallest when there is no effect of drug on Death. The difference is largest when the effect of the drug on T2D is largest, declining up to the point of no difference, as the effect of the drug decreases. The proportion of dead is quite a lot larger in the data with the intervened intensity.

Setting B

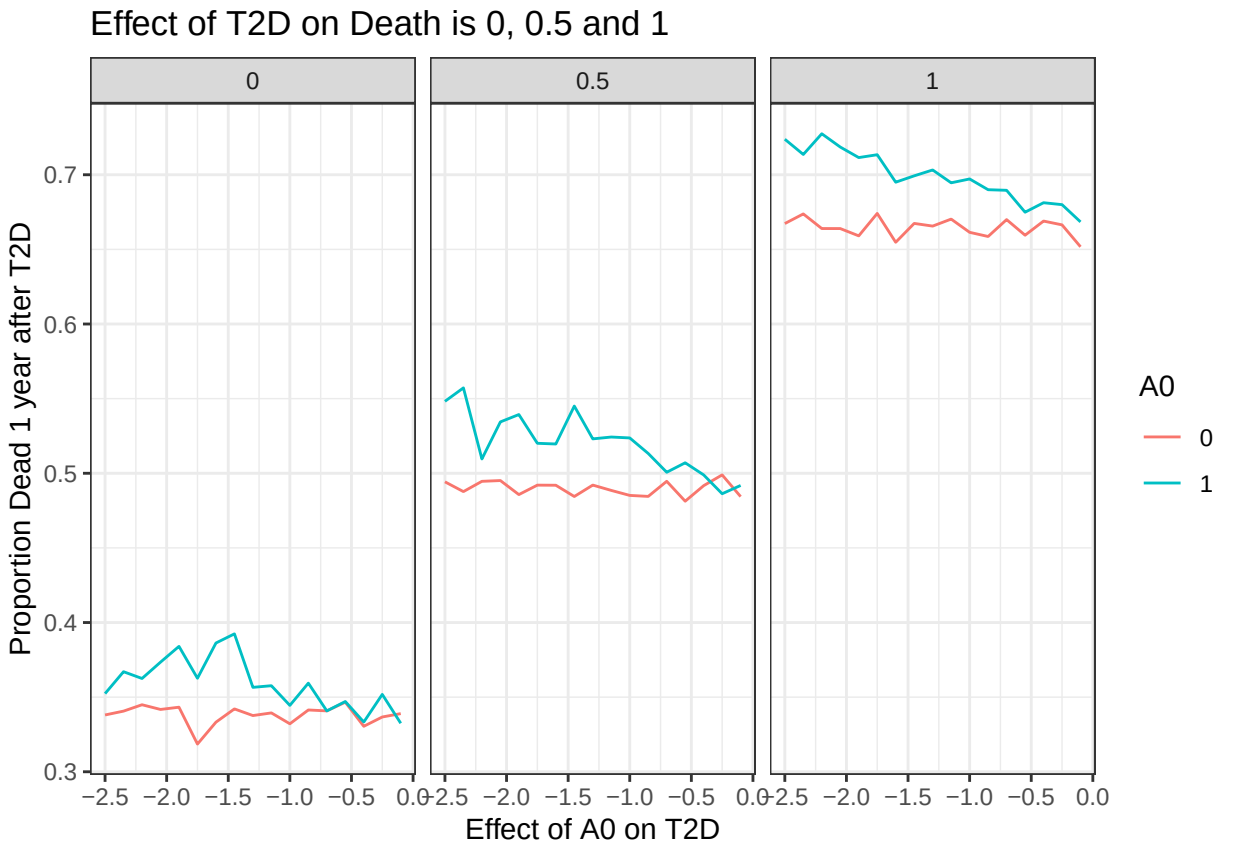
We now turn to a similar setting. Three different scenarios are considered: 1) no effect of T2D on Death (0), 2) moderate effect of T2D on Death (0.5), 3) larger effect of T2D on Death (1). And no effect of drug on Death. All other parameters are the same. The effect of A_0 on T2D is again let to vary from -2.5 to 0 by 0.15.

```
res4 <- compare_effects(estimator = estimator1,
                        N = N,
                        eta = c(0.1,0.3,0.1,0.1),
                        nu = c(1.1,1.3,1.1,1.1),
                        beta_L0_L = 2,
                        beta_A0_L = seq(-2.5, 0, by = 0.15),
                        beta_L0_D = 0,
                        setting = 2)
```

Proportion dead

Below the plots show the proportion dead for different effects of the drug on Death.

```
plot_compare(res4, diff_betas = seq(-2.5, 0, by = 0.15), groups = c(0, 0.5, 1))+
  ylab("Proportion Dead 1 year after T2D")+
  xlab("Effect of A0 on T2D")+
  labs(title = "Effect of T2D on Death is 0, 0.5 and 1")
```



For all three scenarios we see that the proportion of dead is larger in the treatment group than in the placebo group when there is an effect of $A0$ on T2D. This is again due to the selection of individuals. Since $A0$ decreases the risk of T2D, the people in the treatment group that end up diagnosed with T2D are older, and thus more likely to die. It seems the difference is increasing as the effect of T2D on Death increases.

Survival Probability

Below the Kaplan-Meier estimates as well as the .95 CI's of the survival probability 1 year after T2D diagnose are plotted. Again, KM is estimated on censored data.

```
res5 <- compare_effects(estimator = estimator2,
                        N = N,
                        cens = 1,
                        eta = c(0.1,0.3,0.1,0.1),
                        nu = c(1.1,1.3,1.1,1.1),
                        beta_L0_L = 2,
                        beta_A0_L = seq(-2.5, 0, by = 0.15),
                        beta_L0_D = 0,
                        setting = 2)

estimates <- res5[,c(1,2,9,10,17,18)]
lower <- res5[,c(5,6,13,14,21,22)]
upper <- res5[,c(7,8,15,16,23,24)]

plot_compare(estimates = estimates, diff_betas = seq(-2.5, 0, by = 0.15),
             lower = lower, upper = upper, groups = c(0, 0.5, 1)) +
  ylab("Survival probability 1 year post T2D")+
  xlab("Effect of A0 on T2D")+
  labs(title = "Effect of T2D on Death is 0, 0.5 and 1")
```

Again the Kaplan-Meier estimates tell the same story. We now plot the estimated differences in survival probabilities between treatment and placebo group, as well as confidence bands.

```
diffs <- res5[,c(2,10,18)] - res5[,c(1,9,17)]
se <- sqrt(res5[,c(3,11,19)]^2 + res5[,c(4,12,20)]^2)
upper <- diffs + qnorm(0.025)*se
lower <- diffs - qnorm(0.025)*se
```

```
plot.data <- data.table(ests = c(diffs),
            beta = rep(seq(-2.5, 0, by = 0.15), 3),
            group = factor(c(rep("0", 17), rep("0.5", 17), rep("1", 17)),
                           levels =c("0", "0.5", "1")),
            upper = c(upper),
            lower = c(lower))



ggplot(plot.data)+
  geom_line(aes(x = beta, y = ests, color = group))+
  geom_ribbon(aes(x = beta, ymin = lower, ymax = upper, fill = group), alpha = 0.2)+
  ylab("Diff Survival prob 1 year post T2D")+
  xlab("Effect of A0 on T2D")+
  labs(title = "Effect of T2D on Death is 0, 0.5 and 1")
```

Interventions

As before we want to see whether we can estimate the effect of the drug on Death more correctly with interventions. The same procedure as before is applied.

Again three different scenarios are considered: 1) no effect of T2D on Death (0), 2) moderate effect of T2D on Death (0.5), 3) larger effect of T2D on Death (1). And no effect of drug on Death. All other parameters are the same. The effect of A_0 on T2D is again let to vary from -2.5 to 0 by 0.15.

```
res6 <- compare_int_effect(N1 = N1,
            N2 = N2,
```

```
            eta = c(0.1,0.3,0.1,0.1),
            nu = c(1.1,1.3,1.1,1.1),
            beta_A0_D = 0,
            beta_L0_L = 2,
            beta_A0_L = seq(-2.5,0, by = 0.15),
            beta_L0_D = 0,
            tau = 2,
            setting = 2)
```

We plot the intervention effect as the difference in proportion of dead between the intervened data and the data without intervention:

```
diffs <- res6[,c(2,4,6)] - res6[,c(1,3,5)]

data.table(ests = c(diffs),
           beta = rep(seq(-2.5, 0, by = 0.15), 3),
           group = factor(c(rep("0", 17), rep("0.5", 17), rep("1", 17)),
                          levels =c("0", "0.5", "1"))) |>
  ggplot() +
  geom_line(aes(x = beta, y = ests, color = group))+
  xlab("Effect of A0 on T2D")+
  labs(title = "Effect of the T2D on Death is 0, 0.5 and 1")+
  ylab("Diff. Prop. dead")
```

When there is no effect of T2D on Death, there is no difference in proportion of dead. When there is a moderate effect of T2D on Death, there is a difference in proportion of dead. When we do not intervene, fewer people die in the treatment group. This effect declines as the effect of the drug decreases. The same is true when the effect of T2D on Death is large, here the difference is even larger.