T2D

2024-10-18

Differences in T2D diagnosed people

We look into the differences between the Treatment group and the placebo group in different scenarios. The effect of T2D on Death is in all simulations 1. For each setting we simulate where the effect of the drug on T2D is respectively 0, -0.5, -1.

We define a summary function

```
my_summary <- function(data) {
  # T2D events
  T2D_events <- data[Delta == 3]
  # T2D people
  T2D_peeps <- data[ID %in% T2D_events$ID]
  # Mean baseline L0
  L0Mean <- T2D_events[, .(L0_mean = mean(L0)), by = A0][order(A0)]
  # Mean Time of T2D diagnosis
  T_Mean <- T2D_events[, .(T_mean = mean(Time)), by = A0][order(A0)]
  # Number of T2D events in the two groups
  num_events <- T2D_events[,  .N, by = A0][order(A0)]

  # 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

  table_output <- data.table("A0" = L0Mean$A0, "L0 mean" = L0Mean$L0_mean,
                             "T2D Time mean" = T_Mean$T_mean,
                             "Number of Events" = num_events$N,
                            "Prop dead 2 years after T2D" = c(prop_plac, prop_treat))


  return(table_output)
}
```

And histogram function

```
my_hist <- function(data0, data0.5, data1) {
  # T2D events
  T2D_events0 <- data0[Delta == 3]; T2D_events0.5 <- data0.5[Delta == 3]
  T2D_events1 <- data1[Delta == 3]

  my_hist <- gridExtra::grid.arrange(
```

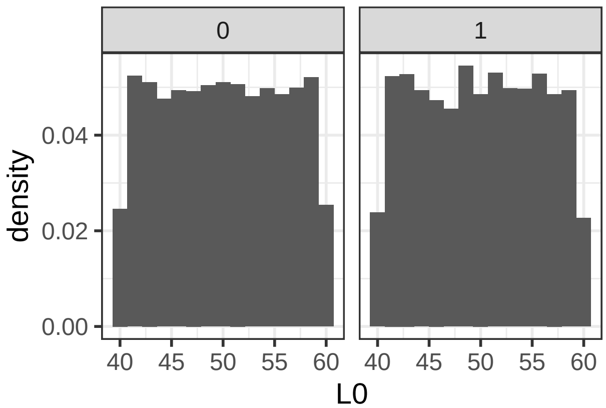
```
  ggplot(T2D_events0)+
    geom_histogram(aes(x = L0, y =..density..), bins = 15)+
    facet_grid(~A0),
  ggplot(T2D_events0.5)+
    geom_histogram(aes(x = L0, y =..density..), bins = 15)+
    facet_grid(~A0),
  ggplot(T2D_events1)+
    geom_histogram(aes(x = L0, y =..density..), bins = 15)+
    facet_grid(~A0),
  nrow = 2
)

  return(my_hist)
}
```

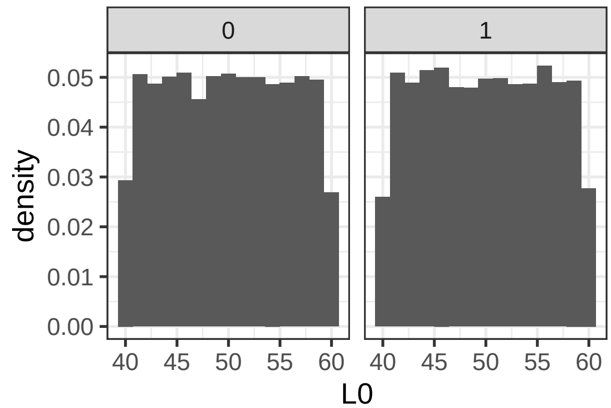
Scenario A: No effect of drug on death, L0 has no effect on neither T2D nor on Death.

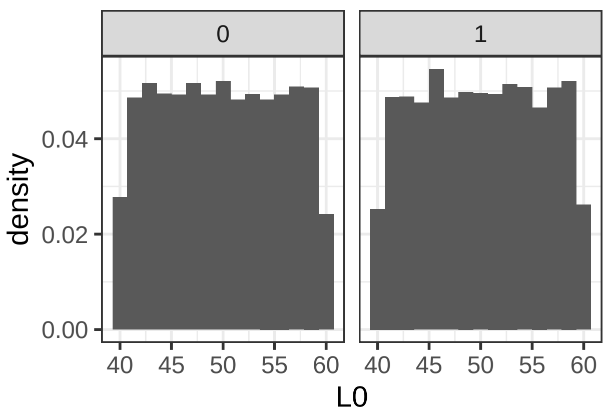
```
N <- 4*10^4
data0_a <- sim_data_setting2(N = N, beta_L_D = 1, beta_A0_D = 0,
                             beta_L0_L = 0, beta_A0_L = 0,
                             cens = 0, beta_L0_D = 0)
data0.5_a <- sim_data_setting2(N = N, beta_L_D = 1, beta_A0_D = 0,
                             beta_L0_L = 0, beta_A0_L = -0.5,
                             cens = 0, beta_L0_D = 0)
data1_a <- sim_data_setting2(N = N, beta_L_D = 1, beta_A0_D = 0,
                             beta_L0_L = 0, beta_A0_L = -1, beta_L0_D = 0,
                             cens = 0)
```

First we look at the distribution of the covariate L_0:

```
## Warning: The dot-dot notation (`..density..`) was deprecated in ggplot2 3.4.0.
## i Please use `after_stat(density)` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```







Looks pretty evenly distributed.

```
my_summary(data0_a) |> kable()
```

A0	L0 mean	T2D Time mean	Number of Events	Prop dead 2 years after T2D
0	49.99361	4.157685	9959	0.2893865
1	49.95676	4.188004	10031	0.2850164

```
my_summary(data0.5_a) |> kable()
```

A0	L0 mean	T2D Time mean	Number of Events	Prop dead 2 years after T2D
0	49.94841	4.167603	10048	0.2920979
1	49.98874	5.169815	7548	0.2844462

```
my_summary(data1_a) |> kable()
```

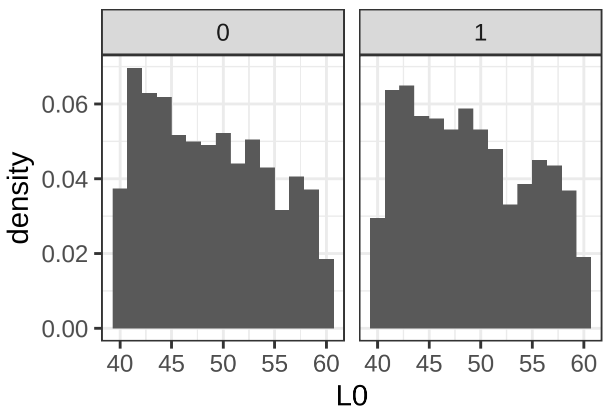
A0	L0 mean	T2D Time mean	Number of Events	Prop dead 2 years after T2D
0	49.94647	4.143611	10048	0.2904061
1	50.04802	5.868896	5401	0.2882799

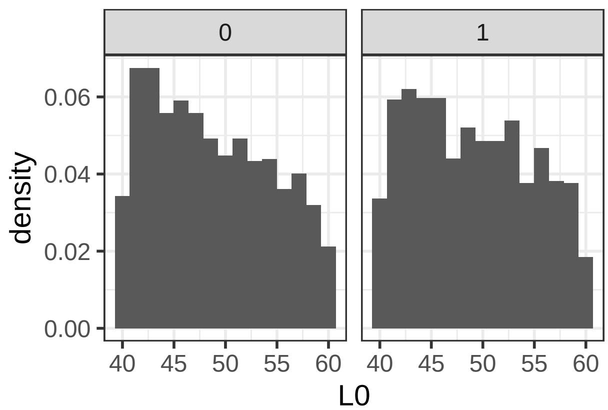
The L0 mean is roughly 50, the larger the effect of the drug, the later the T2D diagnose. The number of T2D diagnosed patients is smaller when the effect of there is an effect of the drug. The proportion of T2D patients dead two years after T2D diagnose is increasing slightly as the effect of the drug increases (in the placebo group a constant proportion   0.29 are dead).

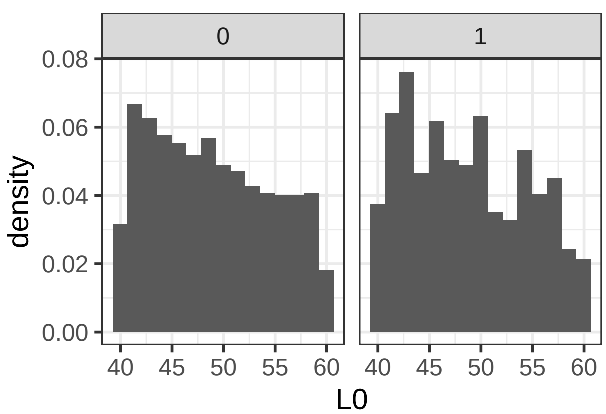
Scenario B: No effect of drug on death, L0 has no effect on T2D but an (very large) effect on Death.

```
data0_b <- sim_data_setting2(N = N, beta_L_D = 1, beta_A0_D = 0,
                             beta_L0_L = 0, beta_A0_L = 0,
                             cens = 0, beta_L0_D = 2)
data0.5_b <- sim_data_setting2(N = N, beta_L_D = 1, beta_A0_D = 0,
                             beta_L0_L = 0, beta_A0_L = -0.5,
                             cens = 0, beta_L0_D = 2)
data1_b <- sim_data_setting2(N = N, beta_L_D = 1, beta_A0_D = 0,
                             beta_L0_L = 0, beta_A0_L = -1, beta_L0_D = 2,
                             cens = 0)
```

First we look at the distribution of the covariate L_0:







Now the distribution is uneven, with the T2D patients having a smaller $L0$. This might be due to the people with high L0 values being dead. But the effect is the same for placebo and treatment patients.

```
my_summary(data0_b) |> kable()
```

A0	L0 mean	T2D Time mean	Number of Events	Prop dead 2 years after T2D
0	48.79310	1.245843	2413	0.8777455
1	48.98386	1.162180	2395	0.8826722

```r
my_summary(data0.5_b) |> kable()
```

A0	L0 mean	T2D Time mean	Number of Events	Prop dead 2 years after T2D
0	48.78299	1.192034	2406	0.8877805
1	49.04895	1.233452	1559	0.8755613

```r
my_summary(data1_b) |> kable()
```

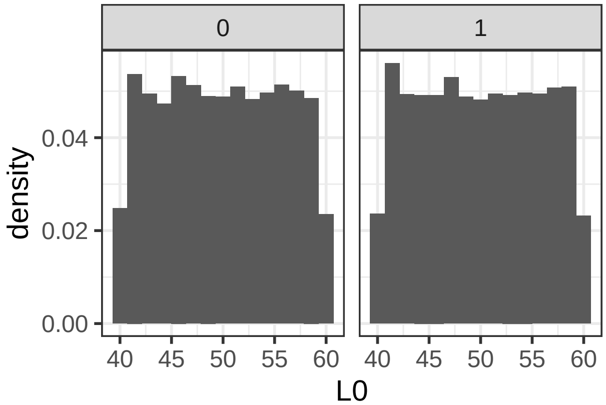
A0	L0 mean	T2D Time mean	Number of Events	Prop dead 2 years after T2D
0	48.93639	1.218034	2486	0.8857603
1	48.72778	1.231720	919	0.8748640

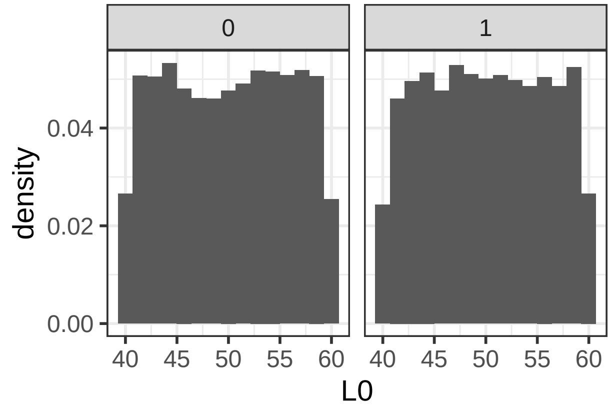
The mean of $L0$ is below 50, but not by much. The T2D time is again larger for the treatment group. The number of T2D events is decreasing for the T2D group as before. Prop dead 2 years after T2D is roughly the same in the two groups.

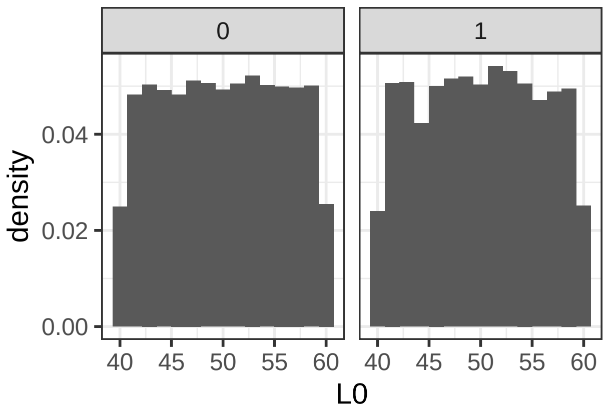
Scenario C: No effect of drug on death, L0 has an effect on T2D and also an effect on Death (of equal size).

```r
data0_c <- sim_data_setting2(N = N, beta_L_D = 1, beta_A0_D = 0,
                             beta_L0_L = 1, beta_A0_L = 0,
                             cens = 0, beta_L0_D = 1)
data0.5_c <- sim_data_setting2(N = N, beta_L_D = 1, beta_A0_D = 0,
                             beta_L0_L = 1, beta_A0_L = -0.5,
                             cens = 0, beta_L0_D = 1)
data1_c <- sim_data_setting2(N = N, beta_L_D = 1, beta_A0_D = 0,
                             beta_L0_L = 1, beta_A0_L = -1, beta_L0_D = 1,
                             cens = 0)
```

First we look at the distribution of the covariate L_0:







Now the distribution seems to be even. The two effects are cancelling!

```
my_summary(data0_c) |> kable()
```

A0	L0 mean	T2D Time mean	Number of Events	Prop dead 2 years after T2D
0	49.92320	1.693743	9924	0.5867594
1	49.93998	1.672889	10009	0.5779798

```
my_summary(data0.5_c) |> kable()
```

A0	L0 mean	T2D Time mean	Number of Events	Prop dead 2 years after T2D
0	50.01462	1.697590	9897	0.5849247
1	50.06515	2.042732	7632	0.5845126

```
my_summary(data1_c) |> kable()
```

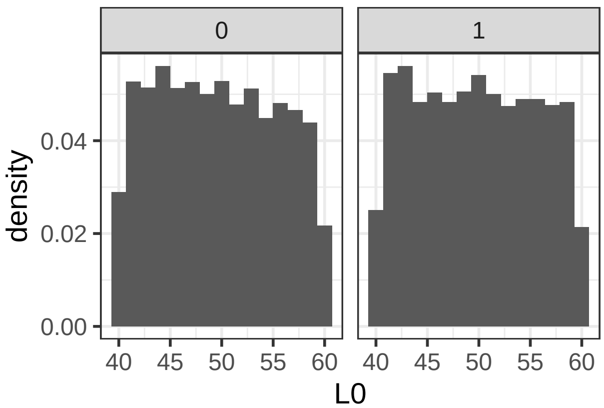
A0	L0 mean	T2D Time mean	Number of Events	Prop dead 2 years after T2D
0	50.04532	1.681428	10109	0.5753289
1	50.02834	2.366642	5429	0.5822435

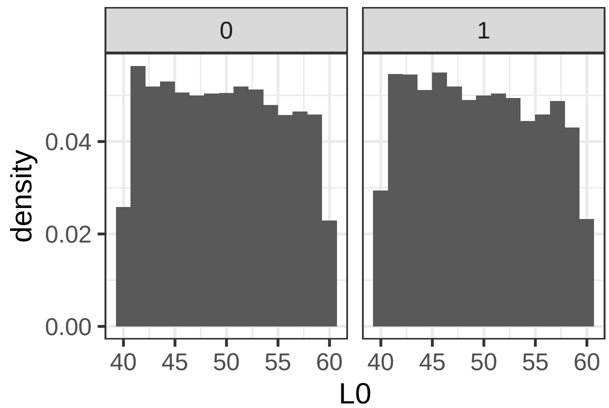
The same results as we found in scenario A.

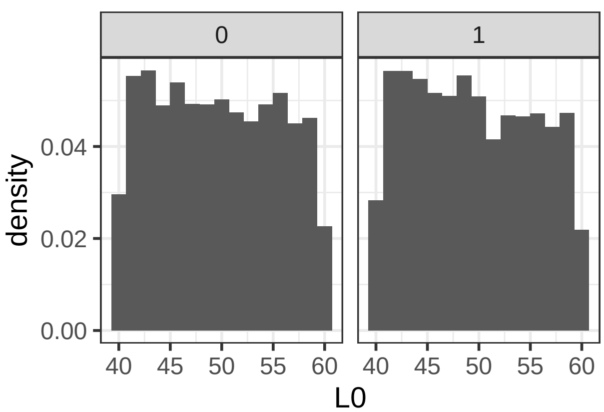
Scenario D: Small effect of drug on Death (-0.3), L0 has a medium effect on T2D (0.7) and and a large effect on Death (1.5).

```
data0_d <- sim_data_setting2(N = N, beta_L_D = 1, beta_A0_D = -0.3,
                             beta_L0_L = 0.7, beta_A0_L = 0,
                             cens = 0, beta_L0_D = 1.5)
data0.5_d <- sim_data_setting2(N = N, beta_L_D = 1, beta_A0_D = -0.3,
                             beta_L0_L = 0.7, beta_A0_L = -0.5,
                             cens = 0, beta_L0_D = 1.5)
data1_d <- sim_data_setting2(N = N, beta_L_D = 1, beta_A0_D = -0.3,
                             beta_L0_L = 0.7, beta_A0_L = -1, beta_L0_D = 1.5,
                             cens = 0)
```

First we look at the distribution of the covariate L_0:







Since the effects are not symmetric, we see again find an uneven distribution, allbeit a less extreme one then the one from scenario B.

```
my_summary(data0_d) |> kable()
```

A0	L0 mean	T2D Time mean	Number of Events	Prop dead 2 years after T2D
0	49.62421	1.446573	6123	0.7440797
1	49.77489	1.740644	7689	0.6466381

```
my_summary(data0.5_d) |> kable()
```

A0	L0 mean	T2D Time mean	Number of Events	Prop dead 2 years after T2D
0	49.69634	1.433143	6189	0.7479399
1	49.58135	2.008021	5583	0.6546660

```
my_summary(data1_d) |> kable()
```

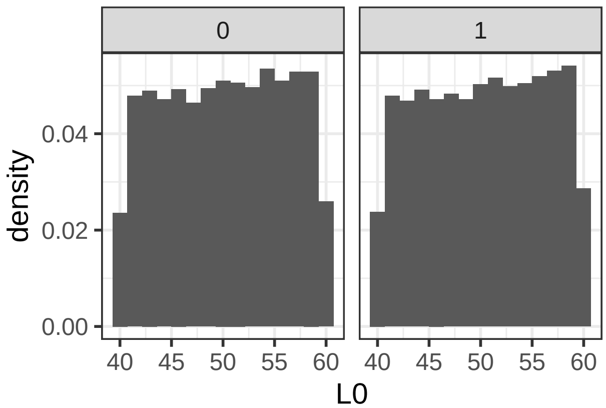
A0	L0 mean	T2D Time mean	Number of Events	Prop dead 2 years after T2D
0	49.63281	1.474355	6217	0.7465015
1	49.51728	2.244914	3610	0.6526316

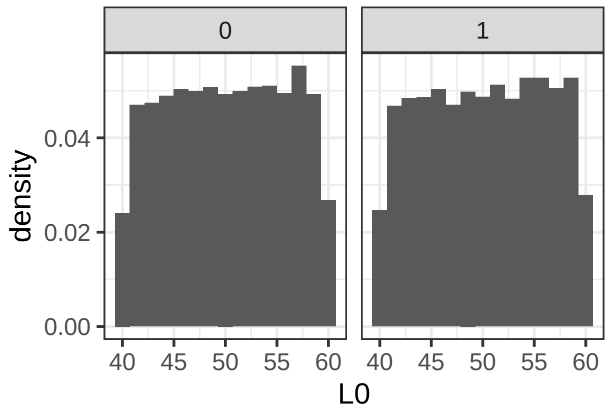
Pretty much the same results as in B. But this time with the proportion of dead after T2D diagnose smaller for the treatment group. But with the proportion of dead incresing as the effect of the drug increases. This must be due to the later T2D diagnose (?).

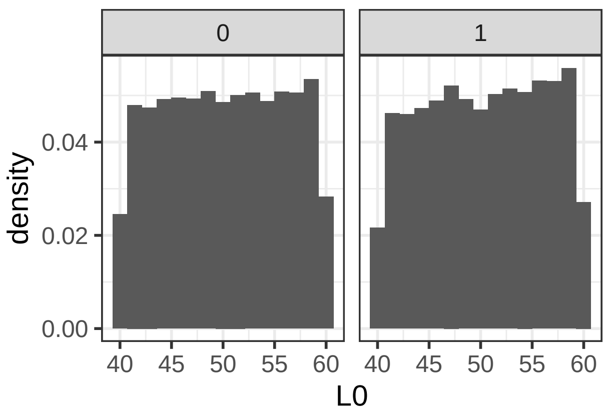
Scenarie E: L0 har en effekt KUN på T2D, No effect of drug on death, Effect of drug on T2D: 0, -0.5, -1.

```
data0_e <- sim_data_setting2(N = N, beta_L_D = 1, beta_A0_D = 0,
                             beta_L0_L = 1.5, beta_A0_L = 0,
                             cens = 0, beta_L0_D = 0)
data0.5_e <- sim_data_setting2(N = N, beta_L_D = 1, beta_A0_D = 0,
                             beta_L0_L = 1.5, beta_A0_L = -0.5,
                             cens = 0, beta_L0_D = 0)
data1_e <- sim_data_setting2(N = N, beta_L_D = 1, beta_A0_D = 0,
                             beta_L0_L = 1.5, beta_A0_L = -1, beta_L0_D = 0,
                             cens = 0)
```

First we look at the distribution of the covariate L_0:







The distribution looks a bit uneven, but it is hard to tell. And hard to see the differences between the scenarios.

```
my_summary(data0_e) |> kable()
```

A0	L0 mean	T2D Time mean	Number of Events	Prop dead 2 years after T2D
0	50.20562	1.690670	16408	0.2674915
1	50.26445	1.663321	16289	0.2708576

```
my_summary(data0.5_e) |> kable()
```

A0	L0 mean	T2D Time mean	Number of Events	Prop dead 2 years after T2D
0	50.14497	1.676359	16170	0.2671614
1	50.18608	2.386365	14587	0.2740111

```
my_summary(data1_e) |> kable()
```

A0	L0 mean	T2D Time mean	Number of Events	Prop dead 2 years after T2D
0	50.16431	1.683541	16403	0.2696458
1	50.31669	3.213476	12483	0.2851879

The L0 mean appears to be increasing for both groups, with the Treatment group having a larger L0 mean consistently. Albeit very close to 50 for all. For Treatment patients the Time of T2D increases substantially and the number of T2D patients decreases as the effect of the drug increases. The proportion of dead in the treatment group increases slightly as the effect of the drug increases.

Larger effect of time on intensity of death

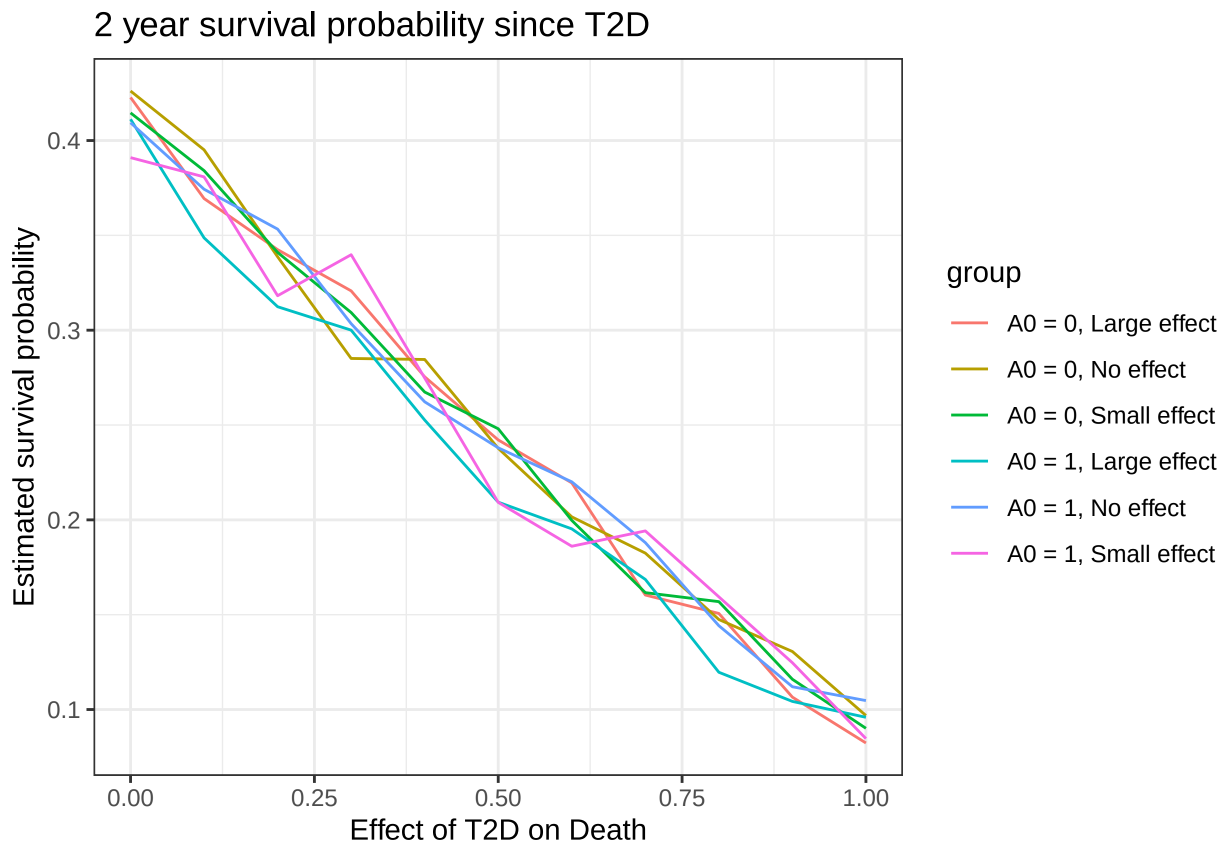
We now investigate how the estimates change if we increase the time effect on the intensity. Remember that we used the Weibull hazard as the baseline hazard:

$$\alpha(t) = \eta\nu t^{nu-1}$$

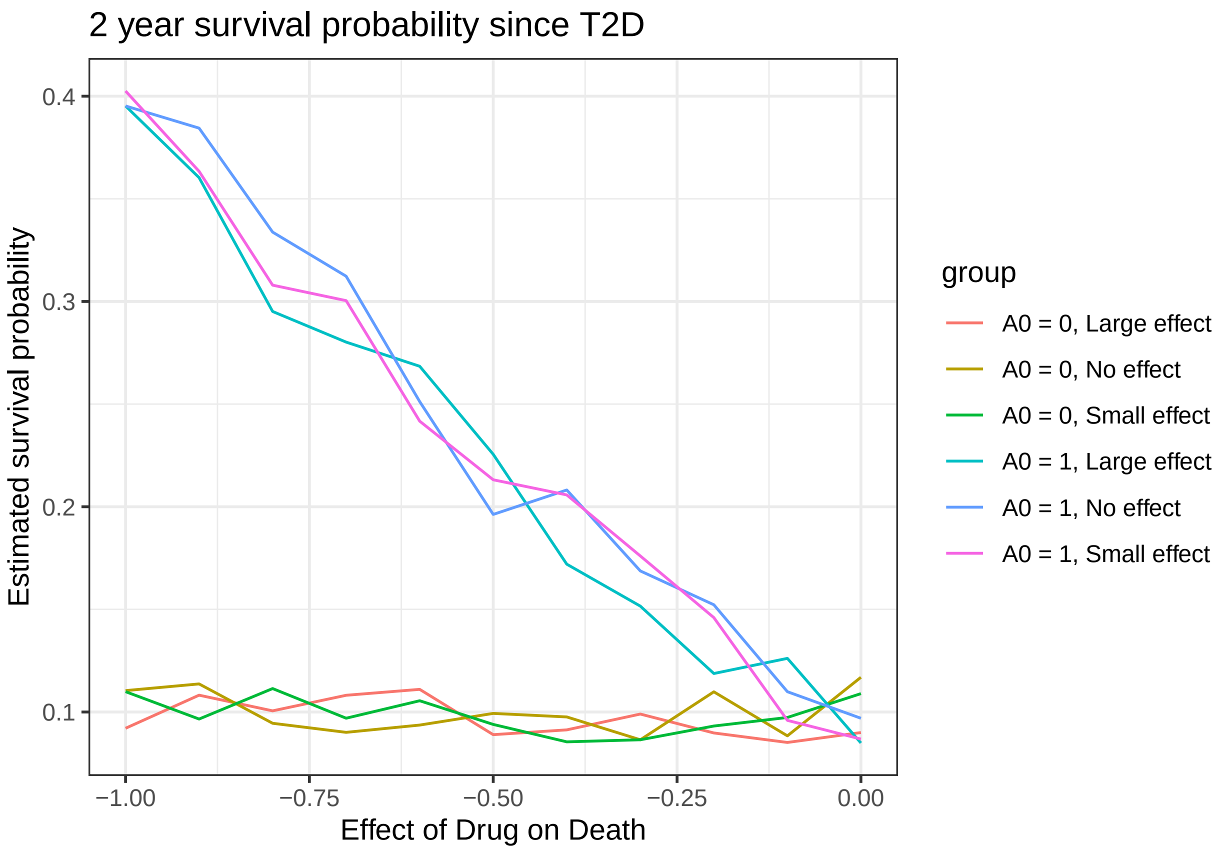
We now increase ν from 1.1 to 1.3.

```r
N <- 8000

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 = 2, newdata = data.frame(A0 = c(0,1)))
    return(c(preds$`A0=0`, preds$`A0=1`))
}

res5 <- compare_effects(estimator = estimator2, N = N, beta_L_D = seq(0,1,by = 0.1), nu = rep(1.3,4))
res6 <- compare_effects(estimator = estimator2, N = N, beta_A0_D = seq(-1,0,by = 0.1), nu = rep(1.3,4))
```

```r
plot_compare(res5, diff_betas= seq(0,1,by = 0.1))+
      ylab("Estimated survival probability")+
      xlab("Effect of T2D on Death")+
      labs(title = "2 year survival probability since T2D")
```




```
plot_compare(res6, diff_betas= seq(-1,0,by = 0.1))+
      ylab("Estimated survival probability")+
      xlab("Effect of Drug on Death")+
      labs(title = "2 year survival probability since T2D")
```



Might there be a slight difference between the three curves? With the estimated survival probability being smallest when the drug has larger effect?