

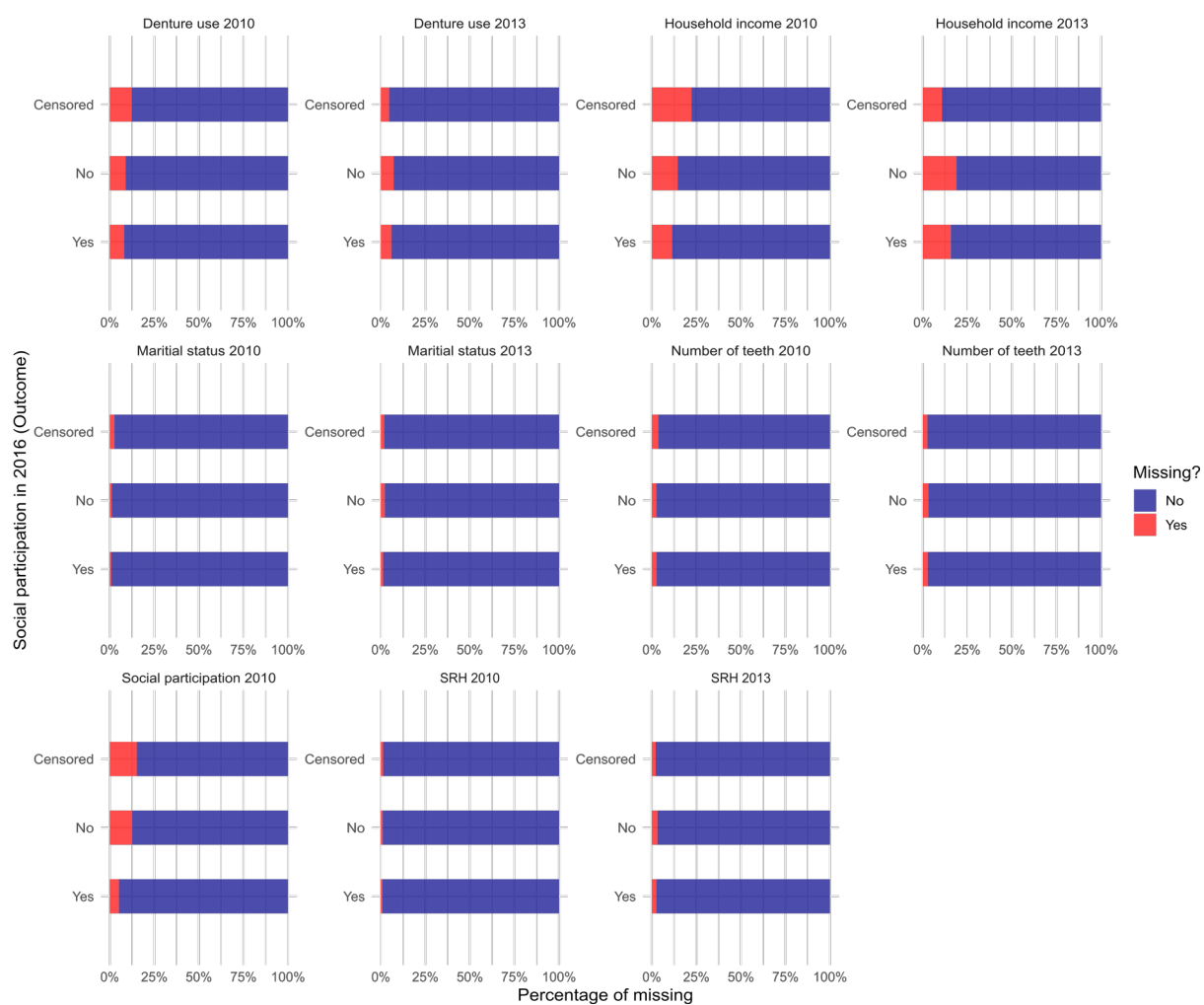
Supplementary Table S1: Baseline characteristics of participants by follow-up status

Characteristics ^a	Follow-up Status			
	Remained (N=24872)	Became ineligible (N=8099)	Died (N=4611)	Lost to follow-up (N=14471)
Number of teeth (2010)				
>= 20 teeth	9717 (57.4%)	1732 (10.2%)	927 (5.5%)	4558 (26.9%)
10-19 teeth	6616 (50.6%)	1748 (13.4%)	995 (7.6%)	3726 (28.5%)
1-9 teeth	5561 (42.3%)	2477 (18.8%)	1340 (10.2%)	3777 (28.7%)
Edentate	2309 (32.0%)	1781 (24.7%)	1184 (16.4%)	1943 (26.9%)
Age (Years)				
Mean (SD)	72.3 (5.0)	78.9 (6.2)	77.7 (6.7)	73.0 (5.4)
Sex				
Female	13717 (49.2%)	5013 (18.0%)	1578 (5.7%)	7574 (27.2%)
Male	11155 (46.2%)	3086 (12.8%)	3033 (12.5%)	6897 (28.5%)
Household income (2010)				
Mean (SD)	2.5 (1.6)	2.2 (1.5)	2.2 (1.6)	2.3 (1.7)
(Missing)	3377	2085	937	3031
Social participation 2010				
Everyday	1098 (56.5%)	199 (10.2%)	112 (5.8%)	533 (27.4%)
2-3 times a week	3912 (57.8%)	779 (11.5%)	382 (5.6%)	1691 (25.0%)
Once a week	3636 (55.7%)	869 (13.3%)	402 (6.2%)	1622 (24.8%)
1-2 times a month	4163 (51.7%)	1096 (13.6%)	656 (8.1%)	2141 (26.6%)
Few times a year	4426 (48.3%)	1251 (13.7%)	889 (9.7%)	2595 (28.3%)
Never	5182 (39.8%)	2529 (19.4%)	1462 (11.2%)	3838 (29.5%)
Denture status (2010)				
Do not wear dentures	10809 (50.9%)	2815 (13.3%)	1637 (7.7%)	5978 (28.1%)

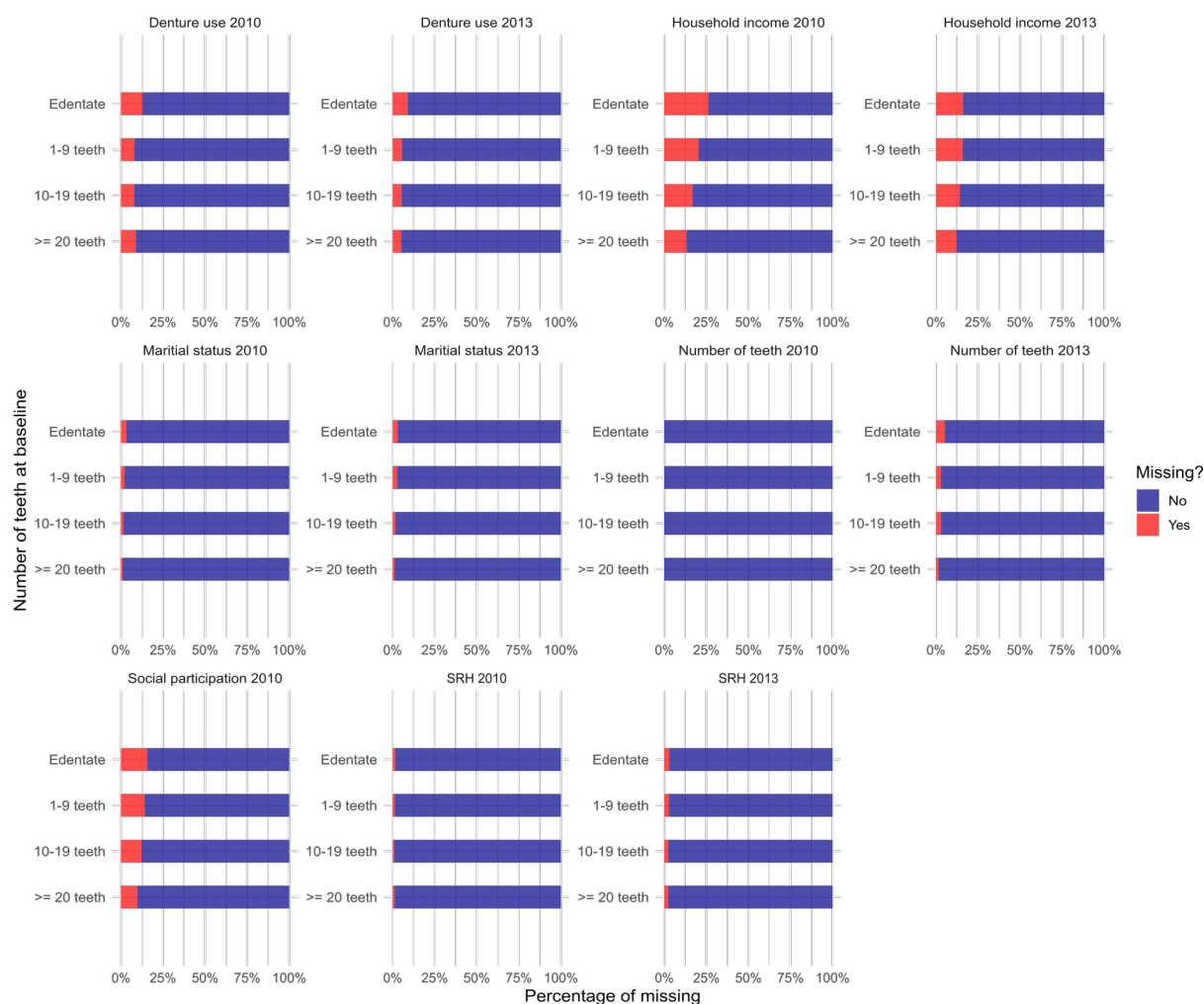
Wear dentures	11883 (47.1%)	4090 (16.2%)	2437 (9.7%)	6833 (27.1%)
Marital status (2010)				
Widowed,divorced, or unmarried	5947 (41.0%)	3364 (23.2%)	1380 (9.5%)	3806 (26.3%)
Married	18673 (51.0%)	4505 (12.3%)	3102 (8.5%)	10335 (28.2%)
Self-rated health (2010)				
Very good	3364 (55.8%)	529 (8.8%)	349 (5.8%)	1790 (29.7%)
Good	18082 (50.7%)	4823 (13.5%)	2697 (7.6%)	10053 (28.2%)
Fair	2900 (34.1%)	2227 (26.2%)	1177 (13.8%)	2207 (25.9%)
Poor	296 (23.5%)	379 (30.1%)	328 (26.1%)	256 (20.3%)

^aMean (SD) for continuous variables; Frequency (%) for categorical variables

Supplementary Figure S1: Distribution of missingness among baseline variables in relation to the outcome



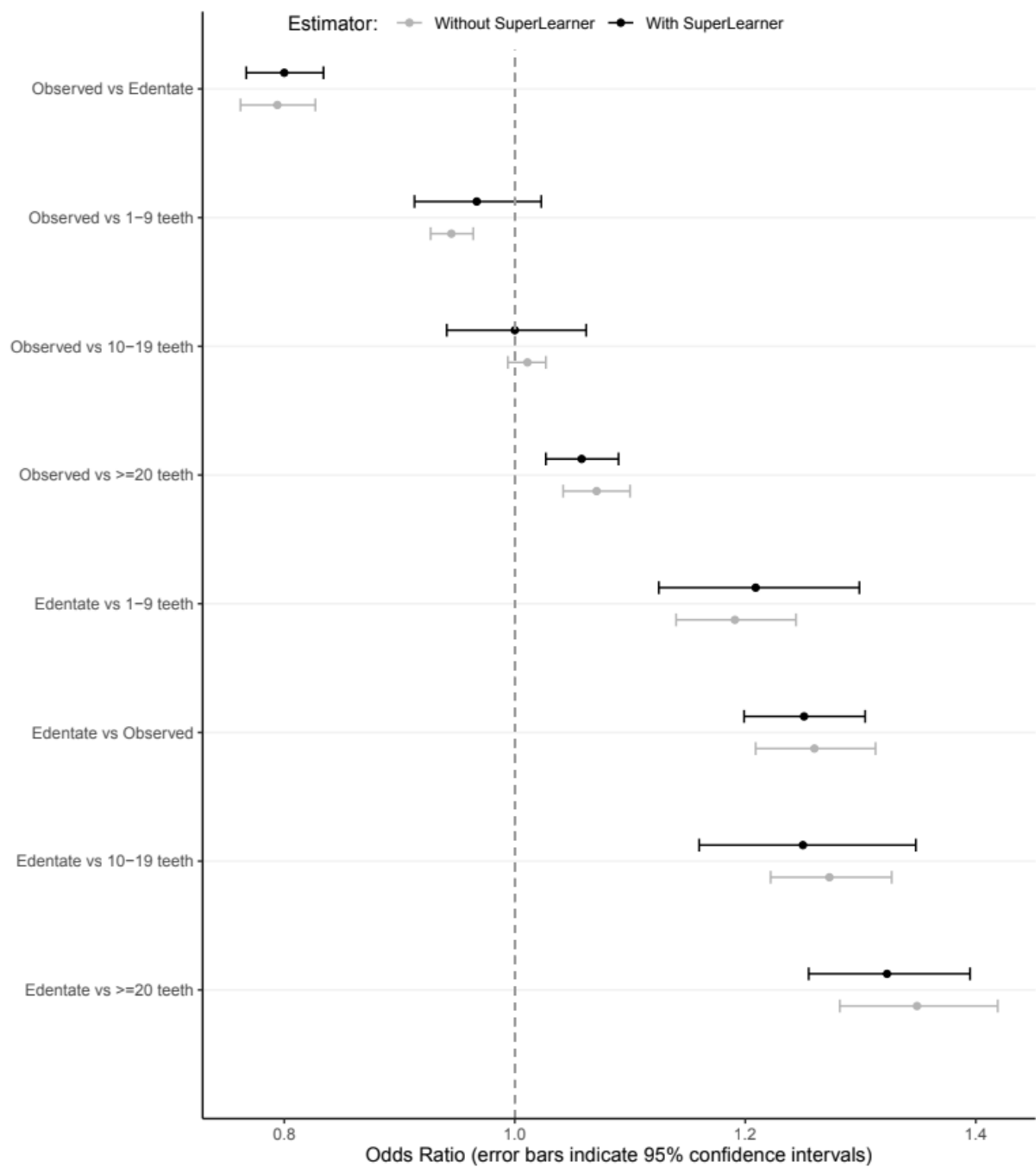
Supplementary Figure S2: Distribution of missingness among baseline variables in relation to the exposure



Supplementary Table S2: Results of TMLE analysis with complete cases

Contrast	OR [95% CI]	P value	E-value
Observed vs Edentate	0.80 [0.77-0.83]	< 0.000	1.81
Observed vs 1-9 teeth	0.97 [0.91-1.02]	0.242	1.22
Observed vs 10-19 teeth	1.00 [0.94-1.06]	0.998	1.00
Observed vs >=20 teeth	1.06 [1.03-1.09]	< 0.000	1.31
Edentate vs 1-9 teeth	1.21 [1.12-1.30]	< 0.000	1.71
Edentate vs 10-19 teeth	1.25 [1.16-1.35]	< 0.000	1.81
Edentate vs Observed	1.25 [1.20-1.30]	< 0.000	1.81
Edentate vs >=20 teeth	1.32 [1.25-1.40]	< 0.000	1.98

Supplementary Figure S3: Comparison of estimates with and without Super Learner (complete case analysis)



Supplementary R script: All results were generated from following R codes. All the functions used in this script can be found at <https://github.com/upulcooray/Social-participation/tree/main/R>

```
library(targets)
library(upulR) # personal R package for creating Table-1

library(future)
library(future.callr)

# Define custom functions and other global objects -----
source("R/functions.R")
source("R/helper_functions.R")

base_cov <- c("Age", "Sex", "LO_inc", "YO_any", "LO_den", "LO_mari", "LO_srh")
l1_cov <- c("L1_inc", "L1_den", "L1_mari", "L1_srh")
expo <- c("A0_teeth", "A1_teeth")
out <- "Y2"

d0 <- NULL

d1 <- function(data, trt) {
  (data[[trt]]==1)*data[[trt]]+ (data[[trt]]!=1)* 1
}

d2 <- function(data, trt) {
  (data[[trt]]==2)*data[[trt]]+ (data[[trt]]!=2)* 2
}

d3 <- function(data, trt) {
  (data[[trt]]==3)*data[[trt]]+ (data[[trt]]!=3)* 3
}

d4 <- function(data, trt) {
  (data[[trt]]==4)*data[[trt]]+ (data[[trt]]!=4)* 4
}
```

```

# Set target-specific options such as packages-----
tar_option_set(packages = c("tidyverse", "haven",
                             "Gmisc", "htmlTable",
                             "flextable", "EValue",
                             "lmtpr", "mice", "upulR"))

plan(callr)

# Starting the list of targets-----
list(
  tar_target(df_file,
             "data/selected",
             format = "file")
  ,
  # Working data -----
  tar_target(working_df,
             readRDS(file=df_file))
  ,
  # create a dataset for descriptive analysis-----
  tar_target(descriptive_data,
             get_descriptive_data(working_df))
  ,
  tar_target(imp_data,
             get_mice_data(descriptive_data ,
                          mice_cars= c(expo[1],base_cov),
                          imp_only_vars= c(expo[2],l1_cov, out),
                          m=5))
  ,

  # plot distribution of missing covariates -----
  tar_target(mis_by_outcome2,
             plot_missing(descriptive_data,
                          by_var = "Y2",
                          x_lab = "Social participation in 2016 (Outcome)") %>%
             ggplot2::ggsave(filename = "figures/missing_outcome.svg",
                              width = 12,height = 10 ),
             format= "file")

```

```

,
tar_target(mis_by_exposure2,
  plot_missing(descriptive_data,
    by_var = "AO_teeth",
    x_lab = "Number of teeth at baseline") %>%
  ggplot2::ggsave(filename = "figures/missing_exposure.svg",
    width = 12,height = 10 ),
  format= "file")

,
# Flow of participants (Add connecting arrows using Inkscape)
tar_target(sample_flowchart2,
  flow_chart_imp(df= descriptive_data ,
    expo,out,base_cov,l1_cov))

,

# Table 1 -----
tar_target(tab1_data2,
  get_tab1_data(descriptive_data),
  format= "rds")

,

tar_target(table_1_2,
  upulR::create_table1(df = tab1_data2,
    headvar = out,
    rowvars = c(expo[1],base_cov),
    headvar_na_level = "Censored",
    file_name = "tables/table_1",
    header = "Social participation in 2016"))

,

tar_target(dropouts_comparison2,
  get_dropout_comparison(df=tab1_data2,
    rowvars= c(expo[1],base_cov)))

,

# get a tmle ready data set-----
# dummify all categorical variables/ all variables as numeric

tar_target(tmle_data2,
  get_tmle_data(imp_data),
  format= "rds")

```



```

,

# Set-up TMLE -----
tar_target(a, expo) # time varying exposure (2010 & 2013)
,

tar_target(y, out) # Outcome (2016)
,
# Time-invariant covariates
tar_target(w, colnames(tmle_data2 %>% select(Age,Sex,contains("Y0"))))

,
# time-varying covariates
tar_target(tv, list(colnames(tmle_data2 %>% select(contains("L0"))),
                    colnames(tmle_data2 %>% select(contains("L1")))))
,

tar_target(cens, c("c1","c2"))
,

tar_target(sl_lib, c("SL.glm", "SL.xgboost", "SL.nnet"))
,

tar_target(params,
  list(trt = a,
       outcome = y ,
       baseline = w ,
       time_vary=tv,
       outcome_type = "binomial",
       cens = cens,
       k=0
       # ,
       # learners_outcome = sl_lib,
       # learners_trt = sl_lib
  ))
,

# Run TMLE -----

tar_target(tmle_res_m1_noSL,
  lapply(paste0("d",0:4) ,
    function (x) do.call(run_lmtp,
      c(params, list(data= tmle_data2 %>% filter(.imp==1),

```

```

                                shift= eval(as.symbol(x)))
        )))
,
tar_target(tmle_res_m1_SL,
  lapply(paste0("d",0:4) ,
    function (x) do.call(run_lmtp,
      c(params, list(data= tmle_data2 %>% filter(.imp==1),
        learners_outcome = sl_lib,
        learners_trt = sl_lib,
        shift= eval(as.symbol(x))))
    )))
,
tar_target(tmle_res_m2_noSL,
  lapply(paste0("d",0:4) ,
    function (x) do.call(run_lmtp,
      c(params, list(data= tmle_data2 %>% filter(.imp==2),
        shift= eval(as.symbol(x))))
    )))
,
tar_target(tmle_res_m2_SL,
  lapply(paste0("d",0:4) ,
    function (x) do.call(run_lmtp,
      c(params, list(data= tmle_data2 %>% filter(.imp==2),
        learners_outcome = sl_lib,
        learners_trt = sl_lib,
        shift= eval(as.symbol(x))))
    )))
,
tar_target(tmle_res_m3_noSL,
  lapply(paste0("d",0:4) ,
    function (x) do.call(run_lmtp,
      c(params, list(data= tmle_data2 %>% filter(.imp==3),
        shift= eval(as.symbol(x))))
    )))
,
tar_target(tmle_res_m3_SL,
  lapply(paste0("d",0:4) ,
    function (x) do.call(run_lmtp,
      c(params, list(data= tmle_data2 %>% filter(.imp==3),
        learners_outcome = sl_lib,
        learners_trt = sl_lib,
        shift= eval(as.symbol(x))))
    )))

```

```

    )))
,
tar_target(tmle_res_m4_noSL,
  lapply(paste0("d",0:4) ,
    function (x) do.call(run_lmtp,
      c(params, list(data= tmle_data2 %>% filter(.imp==4),
        shift= eval(as.symbol(x))))
    )))
,
tar_target(tmle_res_m4_SL,
  lapply(paste0("d",0:4) ,
    function (x) do.call(run_lmtp,
      c(params, list(data= tmle_data2 %>% filter(.imp==4),
        learners_outcome = sl_lib,
        learners_trt = sl_lib,
        shift= eval(as.symbol(x))))
    )))
,
tar_target(tmle_res_m5_noSL,
  lapply(paste0("d",0:4) ,
    function (x) do.call(run_lmtp,
      c(params, list(data= tmle_data2 %>% filter(.imp==5),
        shift= eval(as.symbol(x))))
    )))
,
tar_target(tmle_res_m5_SL,
  lapply(paste0("d",0:4) ,
    function (x) do.call(run_lmtp,
      c(params, list(data= tmle_data2 %>% filter(.imp==5),
        learners_outcome = sl_lib,
        learners_trt = sl_lib,
        shift= eval(as.symbol(x))))
    )))
,

# contrast & pool ref=d0, est= "sl" -----

tar_target(res_d0_sl,
  get_pooled_results(tmle_res_m1_SL,
    tmle_res_m2_SL,
    tmle_res_m3_SL,
    tmle_res_m4_SL,
    tmle_res_m5_SL,

```

```

                                est= "sl",
                                ref_d = 0L))

,

# contrast @ pool ref=d1, est= "sl" -----

tar_target(res_d1_sl,
            get_pooled_results(tmle_res_m1_SL,
                                tmle_res_m2_SL,
                                tmle_res_m3_SL,
                                tmle_res_m4_SL,
                                tmle_res_m5_SL,
                                est= "sl",
                                ref_d = 1L))

,

# contrast @ pool ref=d0, est= "glm" -----

tar_target(res_d0_glm,
            get_pooled_results(tmle_res_m1_noSL,
                                tmle_res_m2_noSL,
                                tmle_res_m3_noSL,
                                tmle_res_m4_noSL,
                                tmle_res_m5_noSL,
                                est= "glm",
                                ref_d = 0L))

,

# contrast @ pool ref=d1, est= "glm" -----

tar_target(res_d1_glm,
            get_pooled_results(tmle_res_m1_noSL,
                                tmle_res_m2_noSL,
                                tmle_res_m3_noSL,
                                tmle_res_m4_noSL,
                                tmle_res_m5_noSL,
                                est= "glm",
                                ref_d = 1L))

,

```

```

# combine pooled results -----

tar_target(pooled_estimates,
            rbind(res_d0_sl,res_d1_sl,
                  res_d0_glm,res_d1_glm))
,

tar_target(table_2,
            get_table2(pooled_estimates),
            format= "file")
,

tar_target(figure_2,
            get_figure2(pooled_estimates),
            format = "file")

)

```

The R session information (including the OS info, R version and all packages used):

```
sessionInfo()
```

```

R version 4.1.2 (2021-11-01)
Platform: x86_64-pc-linux-gnu (64-bit)
Running under: Ubuntu 21.10

```

```

Matrix products: default
BLAS:   /usr/lib/x86_64-linux-gnu/blas/libblas.so.3.9.0
LAPACK: /usr/lib/x86_64-linux-gnu/lapack/liblapack.so.3.9.0

```

```

locale:
 [1] LC_CTYPE=en_US.UTF-8      LC_NUMERIC=C               LC_TIME=en_US.UTF-8
 [4] LC_COLLATE=en_US.UTF-8    LC_MONETARY=ja_JP.UTF-8    LC_MESSAGES=en_US.UTF-8
 [7] LC_PAPER=ja_JP.UTF-8      LC_NAME=C                  LC_ADDRESS=C
[10] LC_TELEPHONE=C            LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C

```

```

attached base packages:
[1] stats      graphics  grDevices  utils      datasets  methods   base

```

```
other attached packages:
```

```
[1] future.callr_0.7.0 future_1.24.0      upulR_0.1.0      stringr_1.4.0
[5] targets_0.10.0      devtools_2.4.3      usethis_2.1.5
```

loaded via a namespace (and not attached):

```
[1] colorspace_2.0-2      ellipsis_0.3.2      visdat_0.5.3      lmtpl_1.0.0
[5] rprojroot_2.0.2      ggstance_0.3.5      flextable_0.6.10  htmlTable_2.4.0
[9] base64enc_0.1-3      fs_1.5.2            rstudioapi_0.13   mice_3.14.0
[13] listenv_0.8.0        remotes_2.4.2      fansi_1.0.2      lubridate_1.8.0
[17] xml2_1.3.3           codetools_0.2-18    splines_4.1.2     cachem_1.0.6
[21] knitr_1.37           pkgload_1.2.4      Formula_1.2-4     nanianr_0.6.1
[25] broom_0.7.11         cluster_2.1.2      png_0.1-7         compiler_4.1.2
[29] backports_1.4.1      assertthat_0.2.1    Matrix_1.4-0      fastmap_1.1.0
[33] survey_4.1-1         cli_3.2.0          htmltools_0.5.2   prettyunits_1.1.1
[37] tools_4.1.2          igraph_1.2.11      gtable_0.3.0      glue_1.6.2
[41] dplyr_1.0.7          tinytex_0.36       Rcpp_1.0.8        SuperLearner_2.0-28
[45] vctr_0.3.8           progressr_0.10.0    iterators_1.0.13  xfun_0.30
[49] globals_0.14.0       ps_1.6.0           brio_1.1.3        testthat_3.1.2
[53] lifecycle_1.0.1      XML_3.99-0.9       scales_1.1.1      parallel_4.1.2
[57] RColorBrewer_1.1-2   yaml_2.3.5         memoise_2.0.1     gridExtra_2.3
[61] ggplot2_3.3.5        gdtools_0.2.3      rpart_4.1-15      glueformula_0.1.0
[65] gam_1.20             latticeExtra_0.6-29 stringi_1.7.6      highr_0.9
[69] desc_1.4.1           fastDummies_1.6.3  foreach_1.5.1     checkmate_2.0.0
[73] pkgbuild_1.3.1       zip_2.2.0          rlang_1.0.2       pkgconfig_2.0.3
[77] systemfonts_1.0.3    evaluate_0.15      lattice_0.20-45    purrr_0.3.4
[81] htmlwidgets_1.5.4    processx_3.5.2     tidyselect_1.1.1  parallelly_1.30.0
[85] magrittr_2.0.2       R6_2.5.1           generics_0.1.1    nnls_1.4
[89] Hmisc_4.6-0          base64url_1.4       DBI_1.1.2         pillar_1.7.0
[93] foreign_0.8-82       withr_2.5.0        Gmisc_3.0.0       forestplot_2.0.1
[97] survival_3.2-13      abind_1.4-5        nnet_7.3-17       tibble_3.1.6
[101] crayon_1.5.0         uuid_1.0-3         utf8_1.2.2        rmarkdown_2.11
[105] officer_0.4.1        jpeg_0.1-9         arsenal_3.6.3      grid_4.1.2
[109] data.table_1.14.2    callr_3.7.0        forcats_0.5.1     digest_0.6.29
[113] tidyr_1.1.4          munsell_0.5.0      mitools_2.4       sessioninfo_1.2.2
```