Danish registry analyses LTMLE specifications - treatment gap

Andrew Mertens

### 1) The Research Question

What is the effect of the 5 year usage of the 2nd line diabetes drug GLP1 (with imperfect adherence) on dementia cumulative incidence, compared to the 5 year usage of the 2nd line diabetes drug SGLT2 (with imperfect adherence) on dementia cumulative incidence

### 2) The Causal Model

* Baseline covariates : age, sex, wealth
* Longitudinal covariates (subset of : Insulin, hypertension, and 2nd line diabetes drugs not a part of the treatment regime, death
* Treatment : GlP1 usage (A1\_t=1) with one-node gaps allowed (treatment holiday) vs SGLT2 usage (A2\_t=1) with one-node gaps allowed (treatment holiday)
* Outcome : dementia incidence

### 3) The Causal Parameter

The relative 5-year risk in dementia between GLP1 users and SGLT2 users:

### 4) Observed Data and Statistical Model

Data from this cohort can be described by a general longitudinal data structure, denoted as:

where W denotes baseline (time-invariant) covariate information, and is exposure status at the beginning of the first node or time interval (i.e. index date). nodes also include an administrative censoring mechanism, C, which denotes censoring during the time interval. For ease of notation above, we refer to the treatment and censoring processes together as . We refer to the covariate and outcome nodes collectively as , which includes both time-varying covariate information and outcome information . Death will be considered a competing risk in the primary analyses and is thus also included in . Index date and baseline measurements are measured at the date of switch to second-line regimen. Here we use a 6-month time scale for time increments in the main analyses. is defined as the maximum follow up time.

#### Glimpse of simulated data:

Note in the real analysis, k=10 for the full 5-year followup with 6-month time nodes, but here we use 3 time nodes for simplicity.

age\_base sex ie.type quartile\_income insulin\_0 hypertension\_0 glp1\_0  
1: 63.73899 0 -0.3260365 0.7032476 0 0 1  
2: 85.62075 1 0.5524619 -0.7747945 0 1 0  
3: 64.44215 1 -0.6749438 -1.0302495 0 0 0  
 sglt2\_inhib\_0 other\_drugs\_0 event\_dementia\_0 censor\_0 event\_death\_0  
1: 0 0 0 0 0  
2: 0 1 0 0 0  
3: 0 0 0 0 0  
 insulin\_1 hypertension\_1 renal.disease\_1 glp1\_1 sglt2\_inhib\_1 other\_drugs\_1  
1: 0 0 0 1 0 0  
2: 0 1 0 0 0 1  
3: 0 0 0 0 0 0  
 event\_dementia\_1 censor\_1 event\_death\_1 insulin\_2 hypertension\_2 glp1\_2  
1: 0 0 0 0 0 0  
2: 0 0 0 0 1 0  
3: 0 0 0 0 0 0  
 sglt2\_inhib\_2 other\_drugs\_2 event\_dementia\_2 censor\_2 event\_death\_2  
1: 0 0 0 1 0  
2: 0 1 0 0 0  
3: 0 0 0 0 0  
 insulin\_3 hypertension\_3 glp1\_3 sglt2\_inhib\_3 other\_drugs\_3 event\_dementia\_3  
1: 0 0 0 0 0 0  
2: 0 1 0 0 1 0  
3: 0 0 0 0 0 0  
 censor\_3 event\_death\_3 event\_dementia\_4  
1: 1 0 0  
2: 0 0 0  
3: 1 0 0

### 5) Identification assumptions

* temporality
* no interference
* consistency

### 6) Statistical Estimation and Inference

* GLM for estimation

#### Specification of the nodes passed to the LMTLE call:

baseline\_vars = c("age\_base","sex", "quartile\_income")  
  
long\_covariates = c("insulin\_","hypertension\_","other\_drugs\_", "event\_death\_")  
  
spec\_ltmle <- spec\_analysis(data=d, long\_covariates, baseline\_vars,  
 N\_time=4,  
 Avars=c("glp1\_","sglt2\_inhib\_"),  
 Yvars=c("event\_dementia\_"),  
 Cvars=c("censor\_"))

**Specifying {A} to allow 1-node treatment gaps:**

Below is an R function to add treatment gaps in {A} specification, where (the intervention arm of GLP1) is intervened upon to start on GLP1 () and individuals remain on the drug but allowing for small treatment gaps and remaining n In the reference arm, the specification is reversed, where is set to allow small gaps and is set to 0.

spec\_Abar\_with\_treatment\_holiday <- function(data=spec\_ltmle$data, A\_nodes=spec\_ltmle$Anodes, intervention\_nodes=c(1,3,5), control\_nodes=c(2,4,6)){  
   
 A\_df <- data %>% select(!!(A\_nodes)) %>% as.data.frame()  
   
 #Make matrices of size A\_nodes X N\_subjects for the intervention and control regimes  
 abar\_gap1 = abar\_gap2 = matrix(NA,nrow=nrow(A\_df),ncol=ncol(A\_df))  
   
 #set first intervention node to one for all: hypothetical trial where all start on drug  
 abar\_gap1[,intervention\_nodes[1]] <- 1  
 #set all control arm drug nodes to 0  
 abar\_gap1[,control\_nodes] <- 0  
   
 #set 2nd intervention node to the observed: hypothetical trial where all start on drug (1st node)  
 #but gaps up to one node are allowed so the 2nd node can be the observed value of A  
 abar\_gap1[,3] = A\_df[,3]  
   
 #for the third node onward, set to:   
 for(i in 3:length(intervention\_nodes)){  
 # 1) observed if previously treated (A\_t=a|A\_(t-1)=1)  
 abar\_gap1[abar\_gap1[,intervention\_nodes[i-1]]==1, intervention\_nodes[i]] = A\_df[abar\_gap1[,intervention\_nodes[i-1]]==1, intervention\_nodes[i]]  
 # 2) treated if previously not treated (A\_t=1|A\_(t-1)=0)   
 abar\_gap1[abar\_gap1[,intervention\_nodes[i-1]]==0, intervention\_nodes[i]] = 1  
 }  
   
 #now repeat the process for the control regime  
 abar\_gap2[,control\_nodes[1]] <- 1  
 abar\_gap2[,intervention\_nodes] <- 0  
 abar\_gap2[,control\_nodes[2]] = A\_df[,control\_nodes[2]]  
 for(i in 3:length(control\_nodes)){  
 # 1) observed if previously treated (A\_t=a|A\_(t-1)=1)  
 abar\_gap2[abar\_gap2[,control\_nodes[i-1]]==1, control\_nodes[i]] = A\_df[abar\_gap2[,control\_nodes[i-1]]==1, control\_nodes[i]]  
 # 2) treated if previously not treated (A\_t=1|A\_(t-1)=0)   
 abar\_gap2[abar\_gap2[,control\_nodes[i-1]]==0, control\_nodes[i]] = 1  
 }  
   
 abar\_gap=list(abar\_gap1, abar\_gap2)  
 return(abar\_gap)  
}

#### Set up contrast between GLP1 usage with gaps and no SGLT2 usage VS. SGLT2 usage with gaps and no GLP1 usage

abar\_gap= spec\_Abar\_with\_treatment\_holiday(data=spec\_ltmle$data,   
 A\_nodes=spec\_ltmle$Anodes,   
 intervention\_nodes=c(1,3,5),   
 control\_nodes=c(2,4,6))

Examine examples of 2 different intervention regimes, comparing observed data and treatment regimes that match the observed data as close as possible while intervening to start on the GLP1 drug of interest, intervening to avoid gaps in treatment longer than 1 node (6 months), and intervening to prevent SGLT2 use:

spec\_ltmle$data %>% select(!!(spec\_ltmle$Anodes)) %>% slice(c(1,2,4,9))

glp1\_1 sglt2\_inhib\_1 glp1\_2 sglt2\_inhib\_2 glp1\_3 sglt2\_inhib\_3  
1: 1 0 0 0 0 0  
2: 0 0 0 0 0 0  
3: 0 0 1 0 0 0  
4: 1 0 1 0 1 0

abar\_gap[[1]][c(1,2,4,9),]

[,1] [,2] [,3] [,4] [,5] [,6]  
[1,] 1 0 0 0 1 0  
[2,] 1 0 0 0 1 0  
[3,] 1 0 1 0 0 0  
[4,] 1 0 1 0 1 0

Example reference regime (SGLT2 use with gaps, no GLP1 use):

head(abar\_gap[[2]],3)

[,1] [,2] [,3] [,4] [,5] [,6]  
[1,] 0 1 0 0 0 1  
[2,] 0 1 0 0 0 1  
[3,] 0 1 0 0 0 1

#### Run analysis

res\_RR <- ltmle(data=spec\_ltmle$data,  
 Anodes = spec\_ltmle$Anodes,  
 Cnodes = spec\_ltmle$Cnodes,  
 Lnodes = spec\_ltmle$Lnodes,  
 Ynodes = spec\_ltmle$Ynodes,  
 survivalOutcome = T,  
 abar = abar\_gap,  
 SL.library = "glm",  
 variance.method = "ic" )  
  
res<-summary(res\_RR)  
res$effect.measures$RR$estimate

[1] 0.5837859

res$effect.measures$RR$CI

2.5% 97.5%  
[1,] 0.009310536 36.60434

#### Alternative 1: Setting A\_bar to continious usage, except for when the observed data has small, allowable gaps

Above, we set as close as possible to the observed data for

If observed GLP1 is for t=1,2,3 is 0,0,0, then is set to 1,0,1

In the real analysis (10 time nodes), where the majority of people are not using GLP1, that means the majority of

Does it make more sense to set for all individuals whose observed data does not match a regime with allowable (1-node) gaps in GLP1 usage, and set to the observed when the observed fits an allowable regime.

This would then use more GLP1 users’ data in estimating dementia risk, while still emulating an ideal trial where individuals continiously use the drug but adherence is imperfect.

I’ve implemented this below to compare effect of the estimated relative risk.

But maybe I’m not understanding what LTMLE is doing when each individual has a different ?

spec\_Abar\_with\_treatment\_holiday\_alt <- function(data=spec\_ltmle$data, A\_nodes=spec\_ltmle$Anodes, intervention\_nodes=c(1,3,5), control\_nodes=c(2,4,6)){  
   
 A\_df <- data %>% select(!!(A\_nodes)) %>% as.data.frame()  
  
 #Make matrices of size A\_nodes X N\_subjects for the intervention and control regimes using observed data  
 abar\_gap1 = abar\_gap2 = A\_df  
   
 #set first intervention node to one for all: hypothetical trial where all start on drug  
 abar\_gap1[,intervention\_nodes[1]] <- 1  
 abar\_gap2[,control\_nodes[1]] <- 1  
   
 #set all control arm drug nodes to 0  
 abar\_gap1[,control\_nodes] <- 0  
 abar\_gap2[,intervention\_nodes] <- 0  
   
 #Set intervention nodes to 1, unless observed data matches the ideal treatment regimes (continious usage with short gaps)  
 is\_following\_regime=rep(TRUE, nrow(abar\_gap1))  
 for(i in 2:length(intervention\_nodes)){  
 is\_following\_regime=ifelse(A\_df[intervention\_nodes[1]]==1 & !(A\_df[,i-1]==0 & A\_df[,i]==0), is\_following\_regime,F)  
 }  
 abar\_gap1[is\_following\_regime, intervention\_nodes] <- A\_df[is\_following\_regime, intervention\_nodes]  
 abar\_gap1[!is\_following\_regime, intervention\_nodes] <- 1  
   
 is\_following\_control=rep(TRUE, nrow(abar\_gap2))  
 for(i in 2:length(control\_nodes)){  
 is\_following\_control=ifelse(A\_df[control\_nodes[1]]==1 & !(A\_df[,i-1]==0 & A\_df[,i]==0), is\_following\_control,F)  
 }  
 abar\_gap2[is\_following\_control, control\_nodes] <- A\_df[is\_following\_control, control\_nodes]  
 abar\_gap2[!is\_following\_control, control\_nodes] <- 1  
  
 abar\_gap=list(as.matrix(abar\_gap1), as.matrix(abar\_gap2))  
 return(abar\_gap)  
}  
  
  
abar\_spec\_alt = spec\_Abar\_with\_treatment\_holiday\_alt()  
  
  
head(abar\_spec\_alt[[1]],5)

glp1\_1 sglt2\_inhib\_1 glp1\_2 sglt2\_inhib\_2 glp1\_3 sglt2\_inhib\_3  
[1,] 1 0 1 0 1 0  
[2,] 1 0 1 0 1 0  
[3,] 1 0 1 0 1 0  
[4,] 1 0 1 0 1 0  
[5,] 1 0 1 0 1 0

head(abar\_spec\_alt[[2]],5)

glp1\_1 sglt2\_inhib\_1 glp1\_2 sglt2\_inhib\_2 glp1\_3 sglt2\_inhib\_3  
[1,] 0 1 0 1 0 1  
[2,] 0 1 0 1 0 1  
[3,] 0 1 0 1 0 1  
[4,] 0 1 0 1 0 1  
[5,] 0 1 0 1 0 1

#### Run analysis

res\_RR <- ltmle(data=spec\_ltmle$data,  
 Anodes = spec\_ltmle$Anodes,  
 Cnodes = spec\_ltmle$Cnodes,  
 Lnodes = spec\_ltmle$Lnodes,  
 Ynodes = spec\_ltmle$Ynodes,  
 survivalOutcome = T,  
 abar = abar\_spec\_alt,  
 SL.library = "glm",  
 variance.method = "ic" )  
  
res<-summary(res\_RR)  
res$effect.measures$RR$estimate

[1] 0.6163843

res$effect.measures$RR$CI

2.5% 97.5%  
[1,] 0.009440081 40.24643

#### Alternative 2: deterministic g function

#specify the intervened treatment  
abar\_spec = list(rep(c(1,0),3),rep(c(0,1),3))  
abar\_spec

[[1]]  
[1] 1 0 1 0 1 0  
  
[[2]]  
[1] 0 1 0 1 0 1

This deterministic g function is the equivalent logic of the spec\_Abar\_with\_treatment\_holiday() function, but applied as determistic knowledge (which feels wrong to me, but was the approach recommended by Zeyi):

SI\_function1 <- function(data, current.node, nodes) {  
 # if not an A node, then skip  
 # if there is no previous node, then treat as usual  
 Anodes <- nodes$A  
 if (!(current.node %in% Anodes)) return(NULL)  
 if (!(any(Anodes < current.node))) return(NULL)  
  
 # if t-1 for the specific A node is treated, then deterministically set to be the observed value; (when observed value is not NA)  
 # prob1 is the prob of being 1, equal to obs value here: observed 1, set a^d\_t = obs value with prob 1; observed 0, set a^d\_t = obs value with prob 0  
 # if t-1 is not treated, then treat as usual (not deterministic)  
 #note that t-1 for a specific type of A is t-4 when there are 4 types of treatments (-3 from the max index of previous Anodes)  
  
 #get Anodes prior to current.node  
 prev.a.node <- max(Anodes[Anodes < current.node])  
 is.deterministic <- ifelse(!is.na(data[, current.node]) & !is.na(prev.a.node), data[, prev.a.node] == 1, F)  
 prob1 <- data[, current.node][is.deterministic]  
  
 return(list(is.deterministic=is.deterministic, prob1=prob1))  
}  
  
 res\_RR <- ltmle(data=spec\_ltmle$data,  
 Anodes = spec\_ltmle$Anodes,  
 Cnodes = spec\_ltmle$Cnodes,  
 Lnodes = spec\_ltmle$Lnodes,  
 Ynodes = spec\_ltmle$Ynodes,  
 survivalOutcome = T,  
 abar = abar\_spec,  
 deterministic.g.function = SI\_function1,  
 SL.library = "glm",  
 variance.method = "ic" )  
  
res<-summary(res\_RR)  
res$effect.measures$RR$estimate

[1] 0.6693092

res$effect.measures$RR$CI

2.5% 97.5%  
[1,] 0.01614784 27.74208

#### Sensitivity analysis: continious usage (no treatment holidays)

res\_RR <- ltmle(data=spec\_ltmle$data,  
 Anodes = spec\_ltmle$Anodes,  
 Cnodes = spec\_ltmle$Cnodes,  
 Lnodes = spec\_ltmle$Lnodes,  
 Ynodes = spec\_ltmle$Ynodes,  
 survivalOutcome = T,  
 abar = abar\_spec,  
 SL.library = "glm",  
 variance.method = "ic" )  
  
res<-summary(res\_RR)  
res$effect.measures$RR$estimate

[1] 0.57511

res$effect.measures$RR$CI

2.5% 97.5%  
[1,] 0.009630054 34.34576

### 7) Interpretation of Results

The deterministic g function leads to an estimate further from the “no treatment holiday” analysis than by manually specifying to allow gaps.

Is the correct interpretation of the regime allowing gaps the following?

“The cumulative risk of dementia in GLP1 users, allowing for small gaps reflective of real world usage, is 0.575 times the risk in SGLT2 users with the same realistic usage”

### Appendix: More complicated comparison of dual treatment regimes.

I think this is how to code the analysis Thomas is working on, contrasting the dual usage of two 2nd-line diabetes drugs (with treatment holidays allowed) against the dual usage of two different 2nd-line diabetes drugs (with treatment holidays allowed).

long\_covariates = c("insulin\_","hypertension\_", "event\_death\_")  
  
#Make mock DPP4 drug for purposes of contrasting 4 drugs  
set.seed(123456)  
d$dpp4\_0 <- d$dpp4\_1 <- d$dpp4\_2 <- d$dpp4\_3 <- rbinom(nrow(d),1, 0.1)  
  
#specify LTMLE analysis  
spec\_ltmle <- spec\_analysis(data=d, long\_covariates, baseline\_vars,   
 N\_time=4, Avars=c("glp1\_","sglt2\_inhib\_","dpp4\_","other\_drugs\_"),  
 Yvars=c("event\_dementia\_"),  
 Cvars=c("censor\_"))  
  
  
#update function to allow for multiple drugs in each arm (new argument `n\_drugs\_per\_arm`)  
spec\_dual\_Abar\_with\_treatment\_holiday <- function(data=spec\_ltmle$data, A\_nodes=spec\_ltmle$Anodes, intervention\_nodes=c(1,2,5,6,9,10), control\_nodes=c(3,4,7,8,11,12), n\_drugs\_per\_arm=2){  
  
 A\_df <- data %>% select(!!(A\_nodes)) %>% as.data.frame()  
   
 #Make matrices of size A\_nodes X N\_subjects for the intervention and control regimes  
 abar\_gap1 = abar\_gap2 = matrix(NA,nrow=nrow(A\_df),ncol=ncol(A\_df))  
   
 #set first intervention node to one for all: hypothetical trial where all start on drug  
 abar\_gap1[,intervention\_nodes[1:n\_drugs\_per\_arm]] <- 1  
 #set all control arm drug nodes to 0  
 abar\_gap1[,control\_nodes] <- 0  
   
 #set 2nd intervention node to the observed: hypothetical trial where all start on drug (1st node)  
 #but gaps up to one node are allowed so the 2nd node can be the observed value of A  
 for(i in 1:n\_drugs\_per\_arm){  
 abar\_gap1[,intervention\_nodes[i+n\_drugs\_per\_arm]] = A\_df[,intervention\_nodes[i+n\_drugs\_per\_arm]]  
 }  
   
   
 #for the third node onward, set to:   
 for(i in (2\*n\_drugs\_per\_arm+1):length(intervention\_nodes)){  
 # 1) observed if previously treated (Aj\_t=a|Aj\_(t-1)=1)  
 abar\_gap1[abar\_gap1[,intervention\_nodes[i-n\_drugs\_per\_arm]]==1, intervention\_nodes[i]] = A\_df[abar\_gap1[,intervention\_nodes[i-n\_drugs\_per\_arm]]==1, intervention\_nodes[i]]  
 # 2) treated if previously not treated (Aj\_t=1|Aj\_(t-1)=0)   
 abar\_gap1[abar\_gap1[,intervention\_nodes[i-n\_drugs\_per\_arm]]==0, intervention\_nodes[i]] = 1  
 }  
   
 #now repeat the process for the control regime  
 abar\_gap2[,control\_nodes[1:n\_drugs\_per\_arm]] <- 1  
 abar\_gap2[,intervention\_nodes] <- 0  
   
 for(i in 1:n\_drugs\_per\_arm){  
 abar\_gap2[,control\_nodes[i+n\_drugs\_per\_arm]] = A\_df[,control\_nodes[i+n\_drugs\_per\_arm]]  
 }  
   
 for(i in (2\*n\_drugs\_per\_arm+1):length(control\_nodes)){  
 # 1) observed if previously treated (A\_t=a|A\_(t-1)=1)  
 abar\_gap2[abar\_gap2[,control\_nodes[i-n\_drugs\_per\_arm]]==1, control\_nodes[i]] = A\_df[abar\_gap2[,control\_nodes[i-n\_drugs\_per\_arm]]==1, control\_nodes[i]]  
 # 2) treated if previously not treated (A\_t=1|A\_(t-1)=0)   
 abar\_gap2[abar\_gap2[,control\_nodes[i-n\_drugs\_per\_arm]]==0, control\_nodes[i]] = 1  
 }  
   
 abar\_gap=list(abar\_gap1, abar\_gap2)  
 return(abar\_gap)  
   
}

#### Specify dual treatments:

dual\_abar\_gap= spec\_dual\_Abar\_with\_treatment\_holiday(data=spec\_ltmle$data,   
 A\_nodes=spec\_ltmle$Anodes,   
 intervention\_nodes=c(1,2,5,6,9,10),   
 control\_nodes=c(3,4,7,8,11,12),   
 n\_drugs\_per\_arm=2)

Examine examples of 2 different regimes

#intervention arm  
head(dual\_abar\_gap[[1]],5)

[,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]  
[1,] 1 1 0 0 0 0 0 0 1 1 0 0  
[2,] 1 1 0 0 0 0 0 0 1 1 0 0  
[3,] 1 1 0 0 0 0 0 0 1 1 0 0  
[4,] 1 1 0 0 1 0 0 0 0 1 0 0  
[5,] 1 1 0 0 0 0 0 0 1 1 0 0

#reference arm  
head(dual\_abar\_gap[[2]],5)

[,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]  
[1,] 0 0 1 1 0 0 0 0 0 0 1 1  
[2,] 0 0 1 1 0 0 0 1 0 0 1 1  
[3,] 0 0 1 1 0 0 0 0 0 0 1 1  
[4,] 0 0 1 1 0 0 0 1 0 0 1 1  
[5,] 0 0 1 1 0 0 0 0 0 0 1 1

res\_dual <- ltmle(data=spec\_ltmle$data,  
 Anodes = spec\_ltmle$Anodes,  
 Cnodes = spec\_ltmle$Cnodes,  
 Lnodes = spec\_ltmle$Lnodes,  
 Ynodes = spec\_ltmle$Ynodes,  
 survivalOutcome = T,  
 abar = dual\_abar\_gap,  
 SL.library = "glm",  
 variance.method = "ic" )  
  
res<-summary(res\_dual)  
res$effect.measures$RR$estimate

[1] 6.721185

res$effect.measures$RR$CI

2.5% 97.5%  
[1,] 7.295275e-05 619227.2

Hopefully, extreme CI’s are from positivity issues or unrealistic data plus the rare outcome in the small (n=1000) simulated dataset used here.