HERU Mathematical Modeler Application Technical Skills Assessment

Lia Humphrey

20/10/2021

HERU Mathematical Modeler Application - Technical Skills Assessment

PART 1 – Running the model and replicating cost-effectiveness outputs

```
# Run model and accumulate outcomes (deterministic)
# This module allows for city selection, loads list of intervention combinations, loads ordinary differential equation functions,
# analysis scenario (deterministic or sensitivity analysis), and loads input parameters and comparators.
# It also provides code for running parallel estimations.

##############################################################################
# 1. SET directory and workspace
##############################################################################

rm(list=ls())
library(rstudioapi)
```

```
## Warning: package 'rstudioapi' was built under R version 4.0.5
```

```
library(LEMHIVpack)
library(rlist)
```

```
## Warning: package 'rlist' was built under R version 4.0.5
```

```
library(foreach)
```

```
## Warning: package 'foreach' was built under R version 4.0.5
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.0.5
```

```r
library(plyr)
```

```
## Warning: package 'plyr' was built under R version 4.0.3
```

```r
#setwd(dirname(rstudioapi::getActiveDocumentContext()$path))

technical_assessment_part <- 1
source("01_Setup/CascadeCEA-Interventions-1-LoadBaselineWorkspace.R")
```

```
## Warning: package 'openxlsx' was built under R version 4.0.5
```

```
## Warning: package 'abind' was built under R version 4.0.3
```

```
## Warning: package 'reshape2' was built under R version 4.0.3
```

```
## Warning: package 'gridExtra' was built under R version 4.0.5
```

```
## Warning: package 'zoo' was built under R version 4.0.5
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
## Warning: package 'tictoc' was built under R version 4.0.5
```

```
## Warning: package 'rmarkdown' was built under R version 4.0.5
```

```
# SELECT city ##
#* CITY <- select.list(all.cities, multiple = FALSE, title = 'Select city', graphics = FALSE)

#************************* Setting city as LA for Part 1
ww <- 3;
CITY <- all.cities[[ww]] # Otherwise you can set city by its index

## LOAD list of all combinations, interventions indexed by number in each combination
# if the list does not exist, source("CascadeCEA-Combination-0-Setup-combination.list.R")
combination.list <- readRDS("Combination/Combination.list.rds")

#************************* Storing the 10 frontier combination indices
combination.frontier.list <- as.vector(readRDS(paste0("Combination/ProductionFunction-Frontier-"
, CITY, ".rds"))$frontier)

## LOAD ODE function
#source("ModelCoreModules/CascadeCEA-Model-0-Function-ode_model-Combination.R") # Function autom
atically loaded with LEMHIVpack

## LOAD analysis scenario
case = "DM"  # DM for deterministic, SA for sensitivity analysis

## LOAD all input parameters and comparators
source("01_Setup/CascadeCEA-Interventions-1-LoadParameterWorkspace-Combination.R")

#* total.comb <- length(combination.list)
#************************* Total combinations for just the frontier
total.comb <- length(combination.frontier.list)

tic("Model run")

#outcome.comb.mx <- matrix(0, nrow = total.comb, ncol = 44)    ##Initialize combination outcome
 matrix (to save results)
outcome.comb.mx <- foreach(cc=1:total.comb, .combine=rbind, .export = export.int.model.names
) %dopar% {
  comb.eva.func(input.parameters = all.params, current.int = interventions[combination.list[[com
bination.frontier.list[cc]]]])
}
```

```
## Warning: executing %dopar% sequentially: no parallel backend registered
```

```
#future:::ClusterRegistry("stop")   ##Stop parallel estimations and free cores

#colnames(outcome.comb.mx)        <- rep("FILL", 44)
colnames(outcome.comb.mx)[1:6]   <- c("Infections.total-20Y", "SuscPY-over20Y", "Infections.tota
l-10Y", "SuscPY-over10Y", "Infections.total-5Y", "SuscPY-over5Y")
colnames(outcome.comb.mx)[7:32]  <- paste0("Year", c(2015:2040))
colnames(outcome.comb.mx)[33:44] <- c("QALYs.sum", "costs.total.sum", "costs.hru.sum", "costs.ar
t.sum", "costs.art.ini.sum",
                                      "costs.oat.sum", "costs.prep.sum", "costs.prep.tests.sum",
"costs.test.sum", "int.costs.sum",
                                      "int.impl.costs.sum", "int.sust.costs.sum")

saveRDS(object = outcome.comb.mx, file = paste0("Combination/Outcome-All-Combination-", CITY, "-
DM.rds"))
toc()
```

```
## Model run: 41.39 sec elapsed
```

```
QALYs_LA <- outcome.comb.mx[,"QALYs.sum"]
costs_total_LA <- outcome.comb.mx[,"costs.total.sum"]

#* Importing status quo values
status_quo_outcomes <- unlist(readRDS(paste0("Inputs/Combination-DM-", CITY, "-refcase-outcomes.
rds")))

#* Recreating table for Supplementary Table 3 Panel A -- with new 70% increase on psi.m between
 2021-2030
Supp_Table_3_Panel_A <- matrix(0, nrow=10, ncol=4)
colnames(Supp_Table_3_Panel_A) <- c("Strategy Index","Incremental Cost: $M", "Incremental QALYs"
,"ICER: $/ QALY")
Supp_Table_3_Panel_A[,1] <- c(combination.frontier.list)
Supp_Table_3_Panel_A[,2] <- round(c((costs_total_LA - status_quo_outcomes[[2]])/1000000),digits=
1)
Supp_Table_3_Panel_A[,3] <- round(c(QALYs_LA - status_quo_outcomes[[1]]),digits=0)

for (i in 1:10) {
  if (Supp_Table_3_Panel_A[i,2] <= 0) {
    Supp_Table_3_Panel_A[i,4] <- c("CS")
  }

  else {
    Supp_Table_3_Panel_A[i,4] <- round((outcome.comb.mx[i,"costs.total.sum"]-outcome.comb.mx[i-1
,"costs.total.sum"])/(outcome.comb.mx[i,"QALYs.sum"]-outcome.comb.mx[i-1,"QALYs.sum"]), digits=0
)
  }
}

Supp_Table_3_Panel_A
```

```
##       Strategy Index Incremental Cost: $M Incremental QALYs ICER: $/ QALY
##  [1,] "297"         "-197.2"             "12855"           "CS"
##  [2,] "1241"        "-158.9"             "14240"           "CS"
##  [3,] "3593"        "-122.1"             "15532"           "CS"
##  [4,] "7618"        "-88.1"              "16206"           "CS"
##  [5,] "12742"       "-75.1"              "16423"           "CS"
##  [6,] "22397"       "205.7"              "20948"           "62066"
##  [7,] "22923"       "206.6"              "20961"           "66546"
##  [8,] "22905"       "594.1"              "26631"           "68343"
##  [9,] "23023"       "595"                "26644"           "69822"
## [10,] "23039"       "604.1"              "26644"           "Inf"
```

```
saveRDS(object = Supp_Table_3_Panel_A, file = paste0("Combination/Outcome-All-Combination-", CIT
Y, "-DM-Supp-Table-3-Panel-A.rds"))
```

I realize that the ICER of Frontier strategy set 10 introduces a division by zero (set 9 and 10 offer the same QALYs) thus producing an “Inf” result. I’m not familiar with how to classify this scenario further (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3955019/ (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3955019/)), but logically I would assume this strategy set would not be cost-effective as strategy set 9 provides the same utility for a lower cost (no marginal utility gained from the additional intervention).

PART 2 – Modeling a new intervention scenario for Los Angeles

For the same city (Los Angeles), modify code to construct a new intervention scenario where monthly HIV testing rates (parameter psi.m in all.params) are increased by 70% from baseline for 2021-2030, and return to the pre-intervention level – baseline - from 2030-2040. Please determine the impact of this intervention as measured by the reduction in the total number of HIV infections in 20 years (2021-2040) in comparison to status quo (current.int = “No interventions”).

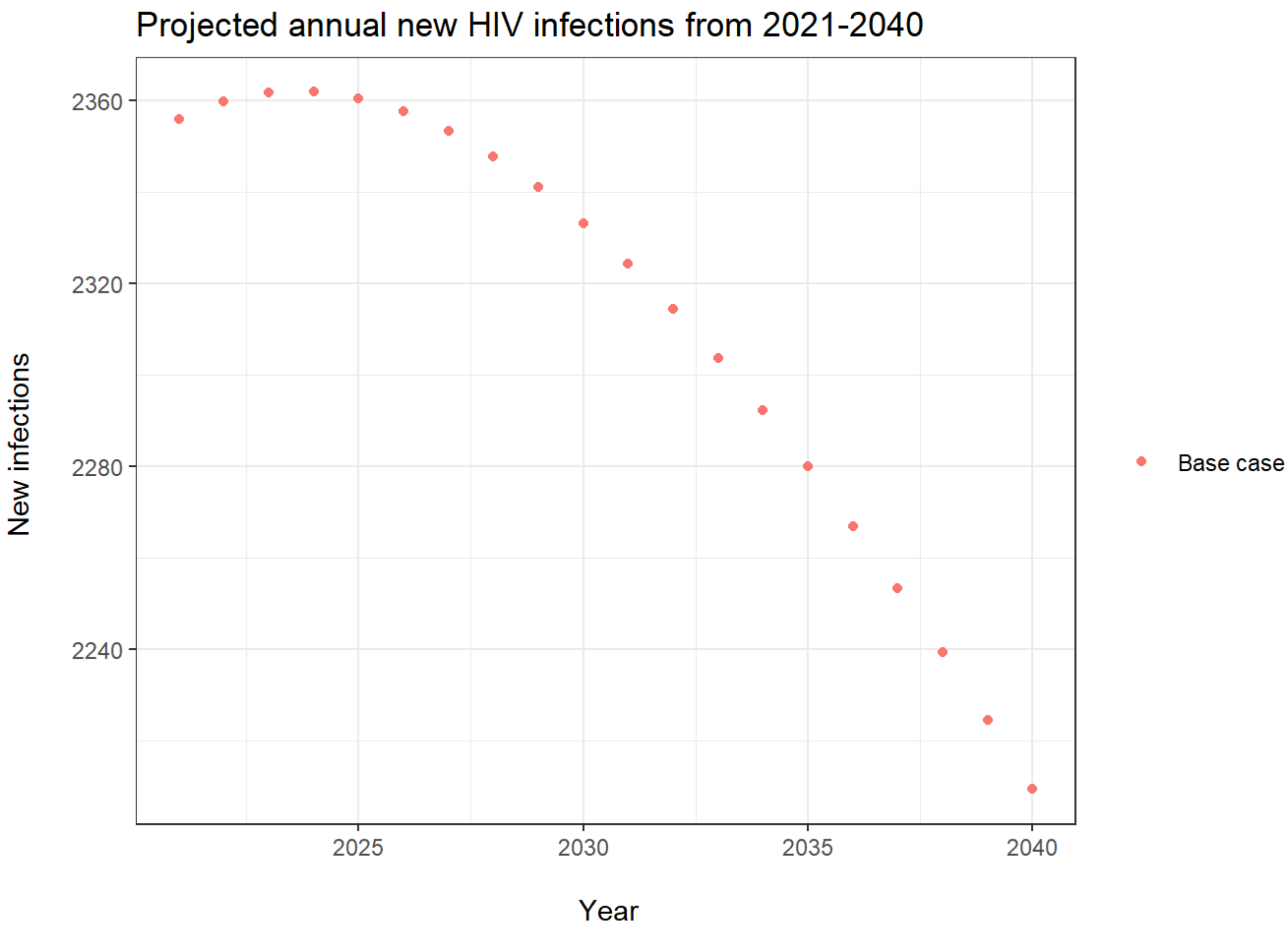
```
#* Importing reference case infections for 2021-2040
status_quo_infections <- readRDS(paste0("Inputs/Combination-DM-", CITY, "-refcase-infections.rd
s"))
status_quo_infections_2021_to_2040 <- as.data.frame(status_quo_infections[["out.inf.yr"]][7:26])

status_quo_infections_2021_to_2040 <- cbind(status_quo_infections_2021_to_2040,c(2021:2040))
colnames(status_quo_infections_2021_to_2040) <- c("Annual_reference_case_infections","Year")
status_quo_infections_2021_to_2040
```

```
##    Annual_reference_case_infections Year
## 1                          2355.881 2021
## 2                          2359.693 2022
## 3                          2361.640 2023
## 4                          2361.862 2024
## 5                          2360.481 2025
## 6                          2357.605 2026
## 7                          2353.336 2027
## 8                          2347.772 2028
## 9                          2341.009 2029
## 10                         2333.136 2030
## 11                         2324.237 2031
## 12                         2314.393 2032
## 13                         2303.679 2033
## 14                         2292.163 2034
## 15                         2279.910 2035
## 16                         2266.977 2036
## 17                         2253.417 2037
## 18                         2239.276 2038
## 19                         2224.597 2039
## 20                         2209.414 2040
```

```
cat("Total number of new HIV infections from 2021 to 2040 with NO interventions:", sum(status_qu
o_infections_2021_to_2040[1:20,]), "\n")
```

```
## Total number of new HIV infections from 2021 to 2040 with NO interventions: 86850.48
```

```
plot1 <- ggplot(status_quo_infections_2021_to_2040, aes(x=Year,y=Annual_reference_case_infection
s, col="Base case")) +
    geom_point() +
    theme_bw() +
    labs(color=" ", title="Projected annual new HIV infections from 2021-2040", x="\n Year", y=
"New infections \n")
plot1
```




```
technical_assessment_part <- 2

## LOAD ODE function
#source("R/CascadeCEA-Model-0-Function-ode_model-Combination.R") # Function automatically loaded
with LEMHIVpack -- importing anyway to make sure psi update works
source("01_Setup/CascadeCEA-Interventions-1-LoadBaselineWorkspace.R") # reinitializing int.proj
 and end.proj
AllParams_LA <- readRDS("C:/Users/Lia/Desktop/CHEOS/Code/LEMHIVpack/Inputs/AllParams-Combination
-DM-LA.rds")

tic("Model run")

outcome.comb.mx.new <- matrix(0, nrow = total.comb, ncol = 44)    ##Initialize combination outco
me matrix (to save results)
outcome.comb.mx.new <- foreach(cc=1:total.comb, .combine=rbind, .export = export.int.model.names
) %dopar% {
  comb.eva.func(input.parameters = all.params, current.int = interventions[combination.list[[com
bination.frontier.list[cc]]]])
}

#future:::ClusterRegistry("stop")   ##Stop parallel estimations and free cores
toc()
```

```
## Model run: 42.72 sec elapsed
```

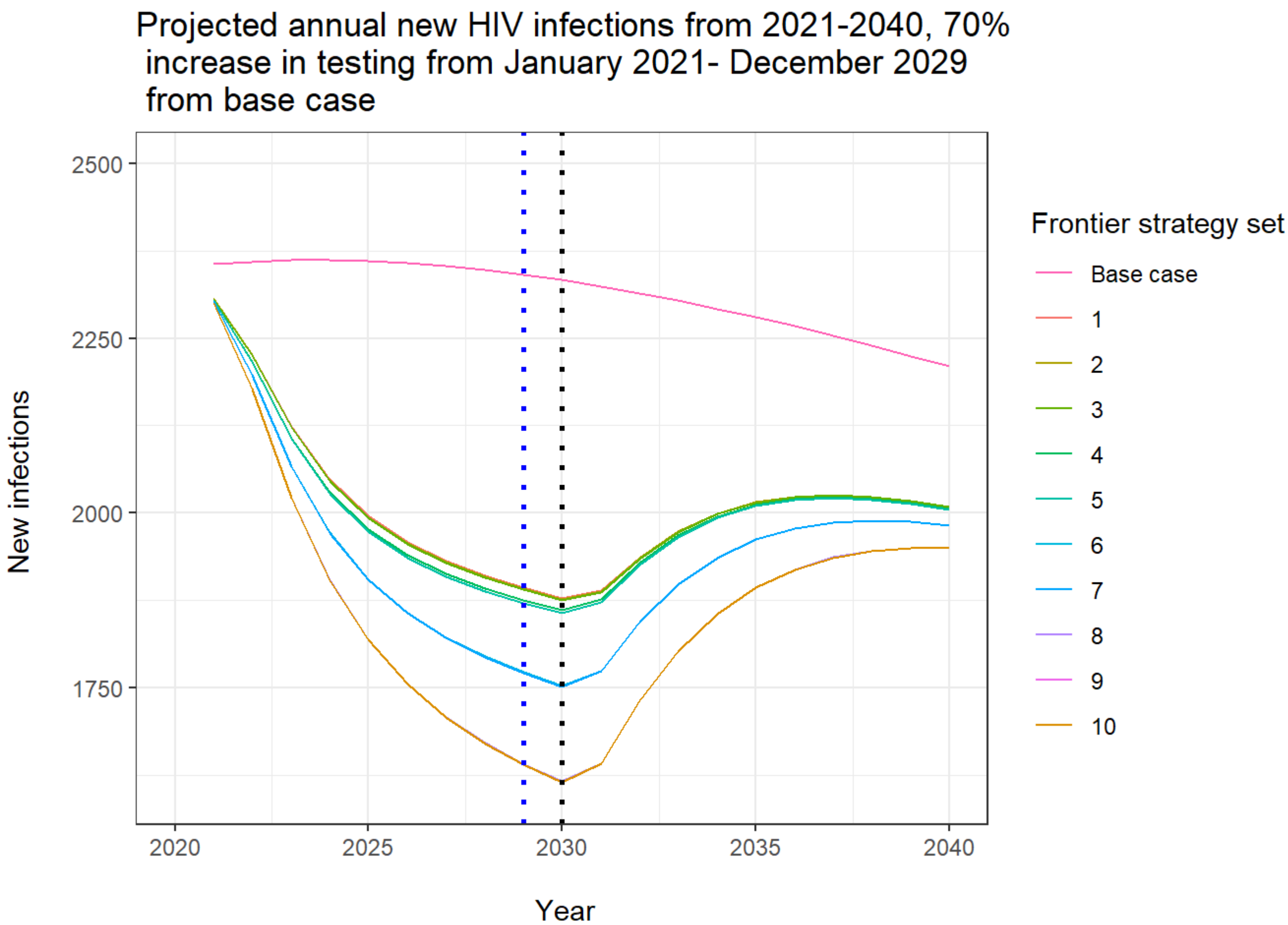
```
#colnames(outcome.comb.mx.new)        <- rep("FILL", 44)
colnames(outcome.comb.mx.new)[1:6]   <- c("Infections.total-20Y", "SuscPY-over20Y", "Infections.
total-10Y", "SuscPY-over10Y", "Infections.total-5Y", "SuscPY-over5Y")
colnames(outcome.comb.mx.new)[7:32]  <- paste0("Year", c(2015:2040))
colnames(outcome.comb.mx.new)[33:44] <- c("QALYs.sum", "costs.total.sum", "costs.hru.sum", "cost
s.art.sum", "costs.art.ini.sum",
                                      "costs.oat.sum", "costs.prep.sum", "costs.prep.tests.sum",
"costs.test.sum", "int.costs.sum",
                                      "int.impl.costs.sum", "int.sust.costs.sum")
rownames(outcome.comb.mx.new)  <- paste0("Frontier_strategy_set_", c(1:10))

outcome.comb.df.new <- as.data.frame(t(outcome.comb.mx.new[,13:32]))
```

After increasing HIV testing by 70% between 2021 and 2030, the final HIV infection rates from 2021-2040 are as follows:

```
cat("Annual new HIV infections per intervention strategy set from 2021 to 2040, having increased
HIV testing by 70% between 2021-2030:", outcome.comb.mx.new[,1], "\n")
```

```
## Annual new HIV infections per intervention strategy set from 2021 to 2040, having increased H
IV testing by 70% between 2021-2030: 40186.65 40168.49 40154.58 39979.6 39931.46 38781.46 38775.
19 37241.86 37235.77 37235.77
```

```
plot2 <- ggplot(outcome.comb.df.new) +
    geom_line(aes(x=c(2021:2040),y=Frontier_strategy_set_1, col="1")) +
    geom_line(aes(x=c(2021:2040),y=Frontier_strategy_set_2, col="2")) +
    geom_line(aes(x=c(2021:2040),y=Frontier_strategy_set_3, col="3")) +
    geom_line(aes(x=c(2021:2040),y=Frontier_strategy_set_4, col="4")) +
    geom_line(aes(x=c(2021:2040),y=Frontier_strategy_set_5, col="5")) +
    geom_line(aes(x=c(2021:2040),y=Frontier_strategy_set_6, col="6")) +
    geom_line(aes(x=c(2021:2040),y=Frontier_strategy_set_7, col="7")) +
    geom_line(aes(x=c(2021:2040),y=Frontier_strategy_set_8, col="8")) +
    geom_line(aes(x=c(2021:2040),y=Frontier_strategy_set_9, col="9")) +
    geom_line(aes(x=c(2021:2040),y=Frontier_strategy_set_10,col="10")) +
    geom_line(data=status_quo_infections_2021_to_2040, aes(x=c(2021:2040),y=Annual_reference_cas
e_infections, col="Base case")) +
      geom_vline(xintercept = 2030, linetype = "dotted", color="black",size=1) +
  geom_vline(xintercept = 2029, linetype = "dotted", color="blue",size=1) +  #included for reaso
ning outlined in email submission
    theme_bw() +
    ylim(1600,2500)+
    xlim(2020,2040)+
    scale_color_discrete(breaks=c("Base case","1", "2", "3","4","5","6","7","8","9","10"))+
    labs(color="Frontier strategy set", title="Projected annual new HIV infections from 2021-204
0, 70% \n increase in testing from January 2021- December 2029\n from base case", x="\n Year", y
="New infections \n")
plot2
```

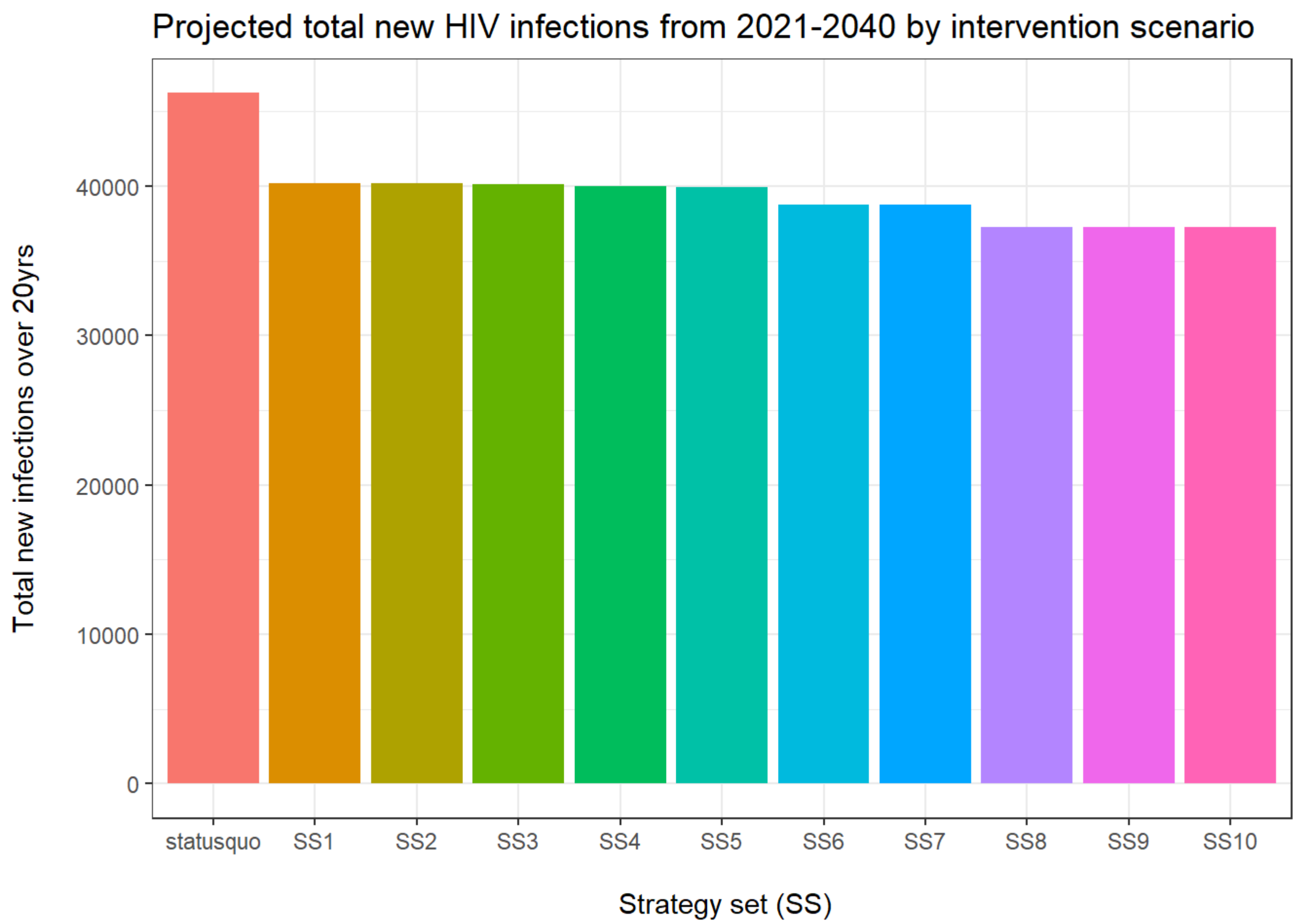



```
total.infections <- as.data.frame(matrix(0, nrow = 11, ncol = 1))
total.infections[1,1] <- sum(status_quo_infections_2021_to_2040$Annual_reference_case_infections
[1:20])

for (i in 2:11) {
  total.infections[i,1] <- sum(outcome.comb.df.new[,i-1])
}

Row.names <- c("statusquo", "SS1", "SS2", "SS3", "SS4", "SS5", "SS6", "SS7", "SS8", "SS9","SS10"
)
total.infections <- cbind(total.infections, Row.names)
colnames(total.infections) <- c("Total_infections","Scenario")

total.infections$Scenario <- factor(total.infections$Scenario, levels = c("statusquo", "SS1", "S
S2", "SS3", "SS4", "SS5", "SS6", "SS7", "SS8", "SS9", "SS10"))
```

```
cat("Total new HIV infections per intervention strategy set from 2021 to 2040, having increased
 HIV testing by 70% between 2021-2030:", unlist(total.infections$Total_infections[2:11]), "\n")
```

```
## Total new HIV infections per intervention strategy set from 2021 to 2040, having increased HI
V testing by 70% between 2021-2030: 40186.65 40168.49 40154.58 39979.6 39931.46 38781.46 38775.1
9 37241.86 37235.77 37235.77
```

```
plot3 <- ggplot(total.infections, aes(Scenario,Total_infections), col = Scenario) +
    geom_bar(aes(fill=Scenario),stat="identity") +
    theme_bw() +
    theme(legend.position = "none")+
    labs(color=" ", title="Projected total new HIV infections from 2021-2040 by intervention sce
nario", x="\n Strategy set (SS)", y="Total new infections over 20yrs\n")
plot3
```



Interpretation

1. What may have contributed to the reduction in new HIV infections from this change in testing rates in the new intervention scenario?

Increased testing rates present the opportunity to catch more infections at different and/or earlier stages and introduce appropriate intervention measures. This not only decreases the number of cases that will reach advanced stages of care, but may also induce behavioural changes in the target communities and reduce contacts between infected and susceptible people (e.g. increased awareness of risk and self-protection measures for the susceptible population, diagnosed infected individuals can take precautions to avoid transmission to others), preventing the spread of HIV to create new cases. In plot 3, one can see that increased testing as a part of the frontier intervention strategies contributes to further decreasing total case counts overall compared to base line.

2. Why do you think there is a difference in the annual number of new infections between the intervention and status quo scenarios after 2030, even though testing parameters have returned to baseline values from 2030-2040?

Prophylactic measures between 2021-2030 inhibit the development of as many annual new HIV cases as compared to base line, as shown in the second plot. While the number of new cases increases after heightened testing is removed, the pool of infected and undiagnosed individuals is comparably smaller by 2030 than it would have been without intervention (dotted line), and thus case rates rise more slowly overall until 2040.

Additional notes

I wanted to add some additional explanation regarding my approach for executing Part 2 due to some uncertainty surrounding simulation and intervention timelines. In first exploring some of the 01_Setup files, I noticed in *CascadeCEA-Interventions-1-LoadBaselineWorkspace.R* Line 53 (Line 60 in my version of the file) that the variable *end.proj* works out to (2040-2012)*12 = 336 months (December 2039), despite my general understanding and the technical assessment document indicating that the end of the projection period is 348 months (December 2040). Furthermore, in line 48 (50 in my version), *lyr* is set to 2040, which to me implies truncation of simulations at the start of 2040 rather than the end. I also noted that Plot 3 Panel A in the **Ending the HIV epidemic in the USA: an economic modelling study in six cities** manuscript shows total incremental costs estimated until the end of 2039, not 2040. Both of these are counterintuitive since the initialization of variable *nyr* specifies the simulation period as 2040-2012**+1**=29 years (348 months).

I also noticed in the *if* statements for the psi.m parameters in *CascadeCEA-Model-0-Function-ode_model-Combination.R* at line 251, psi = psi.m[1,] when t<12 months, implying that 2012's contact rates were not applied to December 2012. Similarly, psi = psi.m[2,] applied to t>=12 & t<24 (December 2012 to November 2013) and so on. This logic is used elsewhere in different files so I didn't want to update it in this instance, but I think this may pose a logical problem with results. It didn't seem to affect the first part of the assessment, but upon tinkering I found using different start and endpoints to a year did influence my Part 2.

I produced plot2 (*Projected annual new HIV infections from 2021-2040, 70% increase in testing from January 2021- December 2029 from base case*) to visualize the effect of the 70% increased testing scenario over time (implemented by introducing a factor of 1.7*psi.m[4,] throughout January 2021 (t=109 months) to December 2029 (t=216 months)). I am fairly confident I implemented this intervention scenario in the code correctly. However, the lowest new case rate visibly occurs BEFORE the end of the increased testing intervention (the red dotted line in the image *plot2_2020-2040* provided in my email). Intuitively, I would expect infections to rise when the intervention was removed in January of 2030 (t>=217, or the black dotted line in *plot2_2020-2040*), leading me to experiment with the simulation start and endpoints.

I found I was only able to produce a plot2 that made logical sense when I adjusted *int.first.year* to 2021 and *lyr* to 2041 (see *plot2_2021-2041*). Now, the minimum of the *ode* result occurs appropriately at 2030 and does not begin to increase prematurely. I imposed *if* statements at lines 48 and 52 of my version of *CascadeCEA-Interventions-1-LoadBaselineWorkspace.R* to reflect which part of the assessment I was running to create this change.

The myriad connected files in this project made me hesitant to tinker any further, but I believe there may be some small discrepancies in the starting and ending points of the simulation timeline that lead to the date inconsistencies in my initial analysis attempt. I would gladly accept an opportunity for discussion and clarification.