COVID-19 vaccine effectiveness against mortality methodology

1.1. This supplement provides additional details regarding the analytical methods used.

1.2. Outcomes

Using death registry data:

Outcome	Derivation
All cause mortality (secondary outcome)	DEATH_DATE
COVID-19 underlying cause of death (primary outcome) [1]	ICD-10 U07.1 or U07.2
COVID-19 contributing cause of death	Underlying or contributing ICD-10 U07.1 or U07.2
Circulatory death	ICD-10 beginning with "I"
Ischaemic heart disease	ICD-10 begins with any of I20 to I25
Cerebrovascular disease	ICD-10 begins with any of I60 to I69
Respiratory death	ICD-10 begins with "J"
Dementia death	ICD-10 begins with any of F01, F03, G30

1.3. Exposure

1.3.1. Primary exposure

To estimate COVID-19 vaccine effectiveness according to the number of doses of vaccine and the time since vaccine receipt, a vaccination status variable has been derived based on that in the linked Australian Immunisation Registry record.

In the model, vaccination status has been accounted for as a time-varying covariate based on the date of vaccination. Both dose received and time since receipt have been considered and as an individual moves through the period of interest, their vaccination status changes over time.

The vaccination status categories are: No vaccine; Dose 1; Dose 2 0-7 days; Dose 2 8-90 days; Dose 2 91-180 days; Dose 2 > 180 days; Dose 3 0-7 days; Dose 3 8-90 days; Dose 3 91-180 days; Dose 3 > 180 days; Dose 4 0-7 days; Dose 4 8-90 days; Dose 4 > 90 days for JAN-MAY22 period of interest; Dose 4 91-180 days and Dose 4 > 180 days for JUN-NOV22 period of interest. Dose 1 and Dose 2 0-7 days are included in models but not of exposure categories of interest for investigation.

We used the counting process to set the dataset structure in order to adequately use coxph function from survival package in R as well as PHREG in SAS. This is described in this article: https://www.pharmasug.org/proceedings/china2017/SP/PharmaSUG-China-2017-SP01.pdf on page 4.

It is a two step process. The first step is to construct a dataset with multiple records per patient, one record for each period during which all the covariates remain constant. The second step is to use a special syntax to obtain the model estimates.

Appendix 4.1 shows some examples of how these multiple records are derived.

1.4. Confounders

The following tables outline the confounders that have been considered in the analyses as adjustment variables based on evidence regarding potential to influence COVID-19 vaccine uptake [2] and mortality risk:

Confounder	Source and classification
Age	MADIP spine data. Used as categorical variable with 2 year band from 65 yo to 100yo and anyone above 100yo grouped together (we run a sensitivity analysis using age as a numeric variable and modelling it using a penalised spline with 5 degrees of freedom and results were similar)
Sex	MADIP spine data Female or Male; Individuals who identified as "Other" have been excluded from the model as the group was too small
Socioeconomic measures	Census data This confounder was only included for the general population analysis Weekly household equivalised income (AUD) using the following categories: 0 - 499; 500 - 999; 1,000 - 1,499; 1,500 - 1,999; 2,000+; Other (incl not stated)
Jurisdiction of residence	Census data One of 8 Australian jurisdictions for the general population excluding "Unknown"; NSW, Victoria, Queensland, Others, including Unknown into Others for the aged care population;
Comorbidity	PBS data Use a standard validated measure: RxRisk using the number of conditions in the 6 months prior to the start of the period of follow-up as the adjustment factor; RxRisk has been categorised as follow: 0; 1; 2; 3+ [3]
Health service access	MBS data Number of GP visits in the year prior to the start of the period of follow-up; has been categorised: 0; 1 - 2; 3 - 6; 7 - 12; 13+
Influenza vaccine	AIR data Flu vaccine in 2021 (Yes/No)

1.5. Statistical methods

1.5.1. Cause-specific mortality outcome

The COVID-19 cause specific mortality outcome has been analysed using a cause-specific competing risk model where death due to any reason other than COVID-19 is a competing event and vaccination status is the time-varying exposure variable [4]. Due to the short follow-up (6 months) and the low event rate (of interest or competing), hazard ratio estimates were no different from a Cox model.

In case of no event of interest occurring during follow-up, the time to event data have been censored at date of receipt of a 5th dose of vaccine or end of follow up time whichever occurred first.

Exposure Intervals (vaccination status categories) have been produced (see Appendix) and the number of events and number of person-years per interval were used to assess the stability/reliability of the predefined

intervals. To extract the 95% CI around the crude rates we used an unadjusted Quasi-Poisson regression on the number of deaths due to COVID with the log value of the number of person-year as an offset.

Statistical measures derived from the competing risk model include adjusted hazard ratios (aHR) and vaccine effectiveness (VE) as well as relative VE (rVE) have been calculated using the formula= (1-aHR)*100%. Adjustment factors in the models were pre-specified and model fit was not routinely conducted following inclusion.

R code available in Appendix 4.2.

1.5.2. All-cause mortality outcome

A similar approach as above has been used on the outcome of all-cause mortality but a proportional hazard Cox model with vaccination status as a time-varying covariate has been used instead of a competing risk analysis.

1.5.3. Model assumptions

Due to the size of the cohort, traditional inferential methods to test for proportionality could not be used.

Proportionality assumptions have been examined using:

1. visual examination of the Scaled Schoenfeld residuals as well as use of a permutation exact test based on the residuals versus time looking at the regression coefficient between the permuted residuals and time and comparing it to the observed coefficient. Repeating this permutation 9,999 times and looking at how many times the coefficients resulting from permutation were different from the observed coefficient, we can compute an exact p-value. The null hypothesis is the slope is null.
2. piecewise Cox model. We spilt the JUN-NOV22 into 2 periods: 01JUN to 31AUG22 and 01SEP to 31NOV22. This variable has been added to the model as well the interaction factor between vaccination status and the split period variable and we examined the HRs for the interaction term.

R code available in Appendix 4.2.

1.5.4. Aged care population

Residence in aged care has been ascertained using Commonwealth Aged care data and have been defined as those in permanent aged care based on admission in an aged care facility at the start of the follow-up period.

1.5.5. Exploratory analyses

The impact of receipt of oral antivirals has been explored. Using PBS data on dispensing of Paxlovid (nirmatrelvir and ritonavir) and Lagevrio (molnupiravir), a period of six weeks (42 days) after receipt of oral antivirals has been flagged as antiviral use equals “Yes”, otherwise antiviral use is “No”.

The dataset structure for this analysis is built in two steps:

1. create the structure allowing vaccination status to be accounted for as a time-varying covariate
2. allowing for “antiviral use” to be time-varying as well by building additional records for individuals who received antiviral, creating one record for each interval during which all the covariates remain constant.

The “antiviral use” covariate is added to the model as an adjustment factor and the hazard ratio estimates for vaccination status were extracted and compared to those estimated using the model without the “antiviral use” covariate.

2. . Contributors

National Centre for Immunisation Research and Surveillance: Sandrine Stepien, Bette Liu, Heather Gidding, Kristine Macartney

NHMRC Centre for Research Excellence in Medicines Intelligence: Sallie Pearson, Nicole Pratt, David Henry, Timothy Dobbins, Claire Vajdic

Australian National University: Rosemary Korda, Jennifer Welsh

Australian Government Department of Health and Aged Care: Lucas Mills

3. References

[1] Australian Bureau of Statistics (ABS), "COVID-19 Mortality in Australia: Deaths registered until 31 May 2022," ABS, Canberra, Australia, 2022.

[2] N. Biddle, J. Welsh, P. Butterworth, B. Edwards and R. Korda, "Socioeconomic determinants of vaccine uptake: July 2021 to January 2022," Australian National University, Canberra, Australia, 2022.

[3] N. Pratt, M. Kerr, J. Barratt, A. Kemp-Casey, L. Kalisch Ellett, E. Ramsay and E. Roughead, "The validity of the Rx-Risk Comorbidity Index using medicines mapped to the Anatomical Therapeutic Chemical (ATC) Classification System," *BMJ Open,* vol. 8, p. e021122, 2018.

[4] T. Therneau , C. Crowson and E. Atkinson, "Using Time Dependent Covariates and Time Dependent," pp. 1-31, 5 August 2022.

4. Appendices

4.1. Appendix 1: Examples of classifying vaccination status categories - number of doses of vaccine and the time since vaccine receipt

individual	VAX1	VAX2	VAX3	VAX4
INDIV1	22DEC21	22JAN22		
INDIV2	15SEP21	13OCT21	18MAY22	
INDIV3	18OCT21	20DEC21	15MAY22	11SEP22

Start and stop date by vaccination status categories

	INDIV1		INDIV2		INDIV3	
Vaccination Status	**Start date**	**Stop date**	**Start date**	**Stop date**	**Start date**	**Stop date**
No Vaccine		21DEC21		14SEP22		17OCT21
Dose 1	22DEC21	21JAN22	15SEP21	12OCT21	18OCT21	19DEC21
Dose 2 (0 - 7 days)	22JAN22	29JAN22	13OCT21	20OCT21	20DEC21	27DEC21
Dose 2 (8 - 90 days)	30JAN22	30APR22	21OCT21	11JAN22	28DEC22	20MAR22
Dose 2 (91 - 180 days)	01MAY22	21JUL22	12JAN22	11APR22	21MAR22	14MAY22
Dose 2 (> 180 days)	22JUL22		12APR22	17MAY22	none	none
Dose 3 (0 - 7 days)			18MAY222	25MAY22	15MAY22	22MAY22
Dose 3 (8 - 90 days)			26MAY22	16AUG22	23MAY22	13AUG22
Dose 3 (91 - 180 days)			17AUG22	14NOV22	14AUG22	10SEP22
Dose 3 (> 180 days)			15NOV22		none	none
Dose 4 (0 - 7 days)					11SEP22	18SEP22

Dose 4 (8 - 90 days)					19SEP22	10DEC22
Dose 4 (> 90 days)					11DEC22	

For period 1 (01JAN22 to 30MAY22), vaccination status that will be contributing to analysis

	INDIV1			INDIV2			INDIV3		
Vaccination Status	Start date	Stop date	Person-Y	Start date	Stop date	Person-Y	Start date	Stop date	Person-Y
No Vaccine									
Dose 1	**01JAN22**	21JAN22	0.057 (21 days)						
Dose 2 (0 - 7 days)	22JAN22	29JAN22	0.019 (7 days)						
Dose 2 (8 - 90 days)	30JAN22	30APR22	0.227 (83 days)	**01JAN22**	11JAN22	0.030 (11 days)	**01JAN22**	20MAR22	0.216 (79 days)
Dose 2 (91 - 180 days)	01MAY22	**31MAY22**	0.085 (31 days)	12JAN22	11APR22	0.227 (83 days)	21MAR22	14MAY22	0.151 (55 days)
Dose 2 (> 180 days)				12APR22	17MAY22	(36 days)	none	none	none
Dose 3 (0 - 7 days)				18MAY222	25MAY22	0.019 (7 days)	15MAY22	22MAY22	0.019 (7 days)
Dose 3 (8 - 90 days)				26MAY22	**31MAY22**	0.014 (5 days)	23MAY22	**31MAY22**	0.022 (8 days)
Dose 3 (91 - 180 days)									
Dose 3 (> 180 days)									
Dose 4 (0 - 7 days)									

Dose 4 (8 - 90 days)									
Dose 4 (> 90 days)									

All individuals included in our analysis cohort (Australian 65+ yo alive and in the country) contributed to the model unless they received a 5th dose of vaccine before the start of the period of interest, they have identified as “Other” for gender (general and aged care population) and have an unknown jurisdiction of residence (general population only). However depending of their vaccination schedule, individuals contributed to different intervals.

4.2. Appendix 2: Examples of R code and SAS code for the different models used for analysis

```
#name:CSC_model.r
#date: 06APR2022
#description: Analyse of mortality with cause specific competing risk analysis using R,
#time efficiency purpose

gc(verbose=TRUE)

library(devtools)
library(readr)

library(tidyverse)
library(broom)
library(plyr)
library(dplyr)

library(survival) #For the time varying model
library(survminer)
library(epiR)
library(KMsurv)

library(ggplot2)
library(scales)

setwd("P:\\NCIRS\\20220915 COVID19 VE against mortality")

Rbind <- function(object) do.call(rbind,object)



##############################################################################################
##
##
##                     P E R I O D   1:   JAN to MAY 2022
##
##############################################################################################
##


#Cause specific.   cohort           agegp                      brand                  outcome
#competing risk   naive vaccinated  65+                   all combined             infection
#============================================================================================


COVDS <- read_csv("01_Data\\Intervals\\JANMAY22_O65_UCOVDTH.csv")


# remove unvaccinated as it is biased group
COVDS  <-subset(COVDS,COVDS$SEXCD!="O")
COVDS2 <-subset(COVDS,COVDS$STEUCPc!="09")%>%
  mutate(VAXSTATUSC = ifelse(VAXSTATUSC == '013', '012', VAXSTATUSC),
         SEXCD      = factor(SEXCD,
                             levels = c('F',      'M'),
                             labels = c('Female', "Male")),
         HIEDcatc   = ifelse(HIEDcatc == '02', '00', HIEDcatc)
  )
```

```
COVDS3 <- COVDS2 %>% select(SPINE_ID, t1, t2, F_UCOVDTHINTV, VAXSTATUSC, SEXCD, agecatc, AGE_AT_01JAN,
STEUCPc,GPviscatc,comorbcatc,HIEDcatc, F_FLU21C )
rm(COVDS)
rm(COVDS2)

#-----------------------------------------------------------------
#---            using age as a categorical variable:
#---             2yr band + one category for age >=100


#check processing time: start time
start_time <- Sys.time()
start_time

#fit the model for competing risk analysis
compRisk0 <- coxph(Surv(t1, t2, F_UCOVDTHINTV==1) ~ VAXSTATUSC + SEXCD + agecatc + GPviscatc +
comorbcatc +STEUCPc + HIEDcatc + F_FLU21C   ,
                   ties="efron",
                   data=COVDS3)

#check processing time: stop time
end_time <- Sys.time()
#check processing time: difference
end_time - start_time

#save results in an object
res.model0 <- summary(compRisk0)

#results of interest: exponentiated coefficients and 95% CI
res.model0$conf.int
write.csv(res.model0$conf.int, file = "03_Outputs\\model coefficients\\JANMAY_O65_UCOVDTH.csv")

gc()


#-----------------------------------------------------------------
#---            using age as a numeric variable:
#---             applying a penalised spline of 5 degrees of freedom


#check processing time: start time
start_time <- Sys.time()
start_time

#fit the model
compRisk1 <- coxph(Surv(t1, t2, F_UCOVDTHINTV==1) ~ VAXSTATUSC + SEXCD + GPviscatc + comorbcatc +STEUCPc
+ HIEDcatc + F_FLU21C + pspline(AGE_AT_01JAN,df=5),
                   ties="efron",
                   data=COVDS3)

#check processing time: stop time
end_time <- Sys.time()
#check processing time: difference
end_time - start_time

#save results in an object
res.model1 <- summary(compRisk1)

tiff(filename="03_Outputs\\model coefficients\\AGE_penalised_spline_curve_O65.tiff")
termplot(compRisk1,term=8,se=TRUE,col.term=1,col.se=1)
dev.off()

#results of interest: exponentiated coefficients and 95% CI
res.model1$conf.int
write.csv(res.model1$conf.int, file = "03_Outputs\\model coefficients\\JANMAY_O65_adjpsplineAGEdf5.csv")

gc()


#-----------------------------------------------------------------
#---          COX model and no competing risk
#---          using age as a categorical variable:


COVDS4 <-  mutate(COVDS3,
                  F_UCOVDTHINTV = ifelse(F_UCOVDTHINTV == 2,
                                         0,
                                         F_UCOVDTHINTV))

sort(unique(COVDS4$F_UCOVDTHINTV,incomparables = FALSE))

#check processing time: start time
```

```
start_time <- Sys.time()
start_time

#fit the model for cox model
Cox <- coxph(Surv(t1, t2, F_UCOVDTHINTV) ~ VAXSTATUSC + SEXCD + agecatc + GPviscatc + comorbcatc
+STEUCPc + HIEDcatc + F_FLU21C,
                  ties="efron",
                  data=COVDS4)

#check processing time: stop time
end_time <- Sys.time()
#check processing time: difference
end_time - start_time

#save results in an object
res.model2 <- summary(Cox)

#results of interest: exponentiated coefficients and 95% CI
res.model2$conf.int
write.csv(res.model2$conf.int, file = "03_Outputs\\model coefficients\\JANMAY_O65_COX.csv")

gc()


#############################################################################################################
##
##
##                     P E R I O D   2:   JUN to NOV 2022
##
#############################################################################################################
##


#Cause specific.   cohort           agegp                       brand                    outcome
#competing risk   naive vaccinated  65+                    all combined             infection
#============================================================================================================

COVDS <- read_csv("01_Data\\Intervals\\JUNNOV22_O65_UCOVDTH.csv")


# remove unvaccinated as it is biased group
COVDS  <-subset(COVDS,COVDS$SEXCD!="O")
COVDS2 <-subset(COVDS,COVDS$STEUCPc!="09")%>%
  mutate(SEXCD      = factor(SEXCD,
                             levels = c('F',      'M'),
                             labels = c('Female', "Male")),
         HIEDcatc   = ifelse(HIEDcatc == '02', '00', HIEDcatc)
  )


COVDS3 <- COVDS2 %>% select(SPINE_ID, t1, t2, F_UCOVDTHINTV, VAXSTATUSC, SEXCD, agecatc, AGE_AT_01JUN,
STEUCPc,GPviscatc,comorbcatc,HIEDcatc, F_FLU21C )

COVDS4 <-  mutate(COVDS3,
                  F_UCOVDTHINTV = ifelse(F_UCOVDTHINTV == 2,
                                         0,
                                         F_UCOVDTHINTV))


#-------------------------------------------------------------------
#---   weighted shoenfeld residuals
#---       hazard of vaccination status over time


#check processing time: start time
start_time <- Sys.time()
start_time
#fit the model
coxfit1 <- coxph(Surv(t1, t2, F_UCOVDTHINTV==1) ~ VAXSTATUSC + SEXCD + agecatc + GPviscatc + comorbcatc
+STEUCPc + HIEDcatc + F_FLU21C  ,
                  ties="efron",
                  data=COVDS4)

#check processing time: stop time
end_time <- Sys.time()
#check processing time: difference
end_time - start_time

resid.Shoenfeld <- residuals(coxfit1, type="scaledsch")

#extract survival times
survtime <- as.numeric(rownames(resid.Shoenfeld))
```

```
##transform to a data frame
resid.Shoenfeld<- data.frame(resid.Shoenfeld,row.names = NULL)

##name variables
names(resid.Shoenfeld) <- names(coef(coxfit1))

##add survival time
resid.Shoenfeld$survtime <- survtime

##Melt
resid.Shoenfeld.melt <- melt(resid.Shoenfeld,id.vars=c("survtime"))

tiff("03_outputs\\02_period start JUN22\\Cox assumptions\\shoenfeld_residuals.tiff", units = "in", width
= 9, height = 6, res = 300)
ggplot(data=resid.Shoenfeld.melt, mapping= aes(x=survtime,y=value))+
    layer(data=resid.Shoenfeld.melt ,mapping= aes(x=survtime,y=value), geom="point",
          stat="identity", position="identity", params=list(na.rm=FALSE))+
    facet_wrap(~variable, scales="free")+
    geom_smooth(method='loess', linewidth=0.75,linetype=1, se=FALSE, span=0.20)+
    scale_y_continuous(name="weighted Schoenfeld residuals")
dev.off()

VAX <- subset(resid.Shoenfeld.melt, str_sub(resid.Shoenfeld.melt$variable,1,3)=="VAX")%>%
       mutate( variable = factor(variable,
                                 levels = c('VAXSTATUSC001', "VAXSTATUSC002","VAXSTATUSC003",
"VAXSTATUSC004", "VAXSTATUSC005",
                                            "VAXSTATUSC006", "VAXSTATUSC007","VAXSTATUSC008",
"VAXSTATUSC009", "VAXSTATUSC010",
                                            "VAXSTATUSC011", "VAXSTATUSC012", "VAXSTATUSC013"),
                                 labels = c('Dose1',"Dose2 0-7days","Dose2 8-90days","Dose2 91-180days",
"Dose2 >180days",
                                            "Dose3 0-7days","Dose3 8-90days","Dose3 91-180days", "Dose3
>180days",
                                            "Dose4 0-7days","Dose4 8-90days","Dose4 91-180days", "Dose4
>180days")))

tiff("03_outputs\\02_period start JUN22\\Cox assumptions\\shoenfeld_residuals_vaccination status.tiff",
units = "in", width = 9, height = 6, res = 300)
ggplot(data=VAX, mapping= aes(x=survtime,y=value))+
  ylim(-5,5)+
  layer(data=VAX,mapping= aes(x=survtime,y=value), geom="point",
        stat="identity", position="identity", params=list(na.rm=FALSE))+
  facet_wrap(~variable, scales="free")+
  geom_smooth (aes ( color = variable), method='loess', linewidth=0.75,linetype=1, se=FALSE, span=0.20)+
  scale_y_continuous(name="weighted Schoenfeld residuals")
dev.off()

VAX3.day8   <- subset(VAX, VAX$variable=="Dose3 8-90days")
VAX3.day91  <- subset(VAX, VAX$variable=="Dose3 91-180days")
VAX3.day181 <- subset(VAX, VAX$variable=="Dose3 >180days")
VAX4.day8   <- subset(VAX, VAX$variable=="Dose4 8-90days")
VAX4.day91  <- subset(VAX, VAX$variable=="Dose4 91-180days")
VAX4.day181 <- subset(VAX, VAX$variable=="Dose4 >180days")

#tiff("03_outputs\\02_period start JUN22\\Cox assumptions\\shoenfeld_dose3_91_180.tiff", units = "in",
width = 9, height = 6, res = 300)
#ggplot(data=VAX3.day91, mapping= aes(x=survtime,y=value))+
#  layer(data=VAX3.day91,mapping= aes(x=survtime,y=value), geom="point",
#        stat="identity", position="identity", params=list(na.rm=FALSE))+
#  facet_wrap(~variable, scales="free")+
#  scale_y_continuous(name="weighted Schoenfeld residuals")
#dev.off()

tiff("03_outputs\\02_period start JUN22\\Cox assumptions\\shoenfeld_dose3_8_90.tiff", units = "in",
width = 9, height = 6, res = 300)
ggplot(data=VAX3.day8, mapping= aes(x=survtime,y=value))+
  geom_point()+
  geom_smooth (method='loess', linewidth=0.75,linetype=1, span=0.20)+
  scale_y_continuous(name="weighted Schoenfeld residuals")
dev.off()

tiff("03_outputs\\02_period start JUN22\\Cox assumptions\\shoenfeld_dose3_91_180.tiff", units = "in",
width = 9, height = 6, res = 300)
ggplot(data=VAX3.day91, mapping= aes(x=survtime,y=value))+
  geom_point()+
  geom_smooth(span=0.5)+
  scale_y_continuous(name="weighted Schoenfeld residuals")
dev.off()

tiff("03_outputs\\02_period start JUN22\\Cox assumptions\\shoenfeld_dose3_181.tiff", units = "in", width
= 9, height = 6, res = 300)
```

```
ggplot(data=VAX3.day181, mapping= aes(x=survtime,y=value))+
  geom_point()+
  geom_smooth (method='loess', linewidth=0.75,linetype=1, span=0.20)+
  scale_y_continuous(name="weighted Schoenfeld residuals")
dev.off()

tiff("03_outputs\\02_period start JUN22\\Cox assumptions\\shoenfeld_dose4_8_90.tiff", units = "in",
width = 9, height = 6, res = 300)
ggplot(data=VAX4.day8, mapping= aes(x=survtime,y=value))+
  geom_point()+
  geom_smooth (method='loess', linewidth=0.75,linetype=1, span=0.20)+
  scale_y_continuous(name="weighted Schoenfeld residuals")
dev.off()


tiff("03_outputs\\02_period start JUN22\\Cox assumptions\\shoenfeld_dose4_91_180.tiff", units = "in",
width = 9, height = 6, res = 300)
ggplot(data=VAX4.day91, mapping= aes(x=survtime,y=value))+
  geom_point()+
  geom_smooth (method='loess', linewidth=0.75,linetype=1, span=0.20)+
  scale_y_continuous(name="weighted Schoenfeld residuals")
dev.off()


tiff("03_outputs\\02_period start JUN22\\Cox assumptions\\shoenfeld_dose4_181.tiff", units = "in", width
= 9, height = 6, res = 300)
ggplot(data=VAX4.day181, mapping= aes(x=survtime,y=value))+
  geom_point()+
  geom_smooth (method='loess', linewidth=0.75,linetype=1, span=0.20)+
  scale_y_continuous(name="weighted Schoenfeld residuals")
dev.off()


#test the slope between scaled shoenfeld and time
#pearson correlation or slope of regression line

#a permutation test may be used to obtain an exact pvalue, regardless of the form of the
#distribution of epsilon(i)'s
#Under H0: beta(1)=0 or H0: r=0, X does not affect the value of Ym so an observed Y is just
#as likely to occur with any X. Thus, the permutation distribution is derived from all
#possible assignments of observed Ys to the Observed Xs

set.seed(5678)

#strength of correlation -1 to 1
#r=0 means no correlation

r.obs <- cor(VAX3.day8$value,VAX3.day8$survtime)
r.obs


slope.obs <- lm(VAX3.day8$value~VAX3.day8$survtime)$coef[2]
slope.obs

n <- length(VAX3.day8$value)
nperms <- 9999

result.r <-numeric(nperms)
result.slope <-numeric(nperms)

for(i in 1:nperms)
{
  index <- sample(n, size=n, replace = FALSE)
  result.r[i] <- cor(VAX3.day8$survtime,VAX3.day8$value[index])
  result.slope[i] <- lm(VAX3.day8$value[index]~VAX3.day8$survtime)$coef[2]
}
'permutation test p-value for dose 3 day 8-90 days'
(sum(result.r >= r.obs)+1)/(nperms + 1)
(sum(result.slope >= slope.obs)+1)/(nperms + 1)


set.seed(89633)

#strength of correlation -1 to 1
#r=0 means no correlation

r.obs <- cor(VAX3.day91$value,VAX3.day91$survtime)
r.obs


slope.obs <- lm(VAX3.day91$value~VAX3.day91$survtime)$coef[2]
slope.obs
```

```
n <- length(VAX3.day91$value)
nperms <- 9999

result.r <-numeric(nperms)
result.slope <-numeric(nperms)

for(i in 1:nperms)
{
  index <- sample(n, size=n, replace = FALSE)
  result.r[i] <- cor(VAX3.day91$survtime,VAX3.day91$value[index])
  result.slope[i] <- lm(VAX3.day91$value[index]~VAX3.day91$survtime)$coef[2]
}
'permutation test p-value for dose 3 day 91-180 days'
(sum(result.r >= r.obs)+1)/(nperms + 1)
(sum(result.slope >= slope.obs)+1)/(nperms + 1)


set.seed(4557)

#strength of correlation -1 to 1
#r=0 means no correlation

r.obs <- cor(VAX3.day181$value,VAX3.day181$survtime)
r.obs


slope.obs <- lm(VAX3.day181$value~VAX3.day181$survtime)$coef[2]
slope.obs

n <- length(VAX3.day181$value)
nperms <- 9999

result.r <-numeric(nperms)
result.slope <-numeric(nperms)

for(i in 1:nperms)
{
  index <- sample(n, size=n, replace = FALSE)
  result.r[i] <- cor(VAX3.day181$survtime,VAX3.day181$value[index])
  result.slope[i] <- lm(VAX3.day181$value[index]~VAX3.day181$survtime)$coef[2]
}
'permutation test p-value for dose 3 day >180 days'
(sum(result.r >= r.obs)+1)/(nperms + 1)
(sum(result.slope >= slope.obs)+1)/(nperms + 1)


set.seed(66891)

#strength of correlation -1 to 1
#r=0 means no correlation

slope.obs <- lm(VAX4.day8$value~VAX4.day8$survtime)$coef[2]
slope.obs

n <- length(VAX4.day8$value)
nperms <- 9999

result.slope <-numeric(nperms)

for(i in 1:nperms)
{
  index <- sample(n, size=n, replace = FALSE)
  result.slope[i] <- lm(VAX4.day8$value[index]~VAX4.day8$survtime)$coef[2]
}
'permutation test p-value for dose 4 day 8-90 days'
(sum(result.slope >= slope.obs)+1)/(nperms + 1)


set.seed(57933)

#strength of correlation -1 to 1
#r=0 means no correlation

slope.obs <- lm(VAX4.day91$value~VAX4.day91$survtime)$coef[2]
slope.obs

n <- length(VAX4.day91$value)
nperms <- 9999

result.slope <-numeric(nperms)
```

```
for(i in 1:nperms)
{
  index <- sample(n, size=n, replace = FALSE)
  result.slope[i] <- lm(VAX4.day91$value[index]~VAX4.day91$survtime)$coef[2]
}
'permutation test p-value for dose 4 day 91-180 days'
(sum(result.slope >= slope.obs)+1)/(nperms + 1)


set.seed(1289)

#strength of correlation -1 to 1
#r=0 means no correlation


slope.obs <- lm(VAX4.day181$value~VAX4.day181$survtime)$coef[2]
slope.obs

n <- length(VAX4.day181$value)
nperms <- 9999

result.slope <-numeric(nperms)

for(i in 1:nperms)
{
  index <- sample(n, size=n, replace = FALSE)
  result.slope[i] <- lm(VAX4.day181$value[index]~VAX4.day181$survtime)$coef[2]
}
'permutation test p-value for dose 4 day >180 days'
(sum(result.slope >= slope.obs)+1)/(nperms + 1)



#*************************************************************************************************
#                                                          END OF PROGRAM
#*************************************************************************************************


*=================================================================================

Project Name:            20220915 COVID19 VE against mortality

Program Name:            Cox_piecewise_model_SASshortcut

Purpose:                 run piecewise cox model using base SAS instead of EG
                         use shortcut to redirect WORK temporary folder
                    for greater memory so the model could run

Author/date:             Sandrine STEPIEN/14APR2023

Platform/SAS version:    0121010Desktop /SAS base v9.4

Input datasets:          NA

Outputs/Saved datasets:  NA

Revisions:
Programmer/Date:        Reason:
-----------------       -------
SS /DDMMMYYYY           add description of reasons for modifications

=================================================================================;


%global RUNTIME RUNDATE ;

%let RUNTIME = %sysfunc(putn(%sysfunc(TIME()),TIME5.));
%let RUNDATE = %sysfunc(putn(%sysfunc(TODAY()),DATE9.));

option source source2 nodate nonumber mprint;

* Define macro variables for ease of use and smmother changes if required;
%global ROOT;

%let ROOT=P:\NCIRS\20220915 COVID19 VE against mortality;

title1 "COVID19 VE against mortality";


* Define output libraries ;
*======================;
libname DERIVED  "&ROOT.\01_Data";
```

```
libname PER2_    "&ROOT.\01_Data\JUNSEP22";
libname MAINDATA "&ROOT.\01_Data\JUNSEP22\Main Analysis";

%global period;
%let Period  = 2;

libname RESULT   "&ROOT.\03_Outputs";
libname RATES    "&ROOT.\01_Data\Rates";
libname INTV     "&ROOT.\01_Data\Intervals";


* Define output libraries for perturbed outputs;
*=========================================;
libname RKEY     "&ROOT.\01_Data\rkeyDATA";

* Define output libraries for formats;
*=========================================;
libname LIBRARY  "&ROOT.\01_Data";

*--Macros and utility programs;
FILENAME macros ("&ROOT.\02_Program_Code_Scripts\Macros");

options MSGLEVEL=I
        MERGENOBY=WARN
        MPRINT
        SOURCE
        SOURCE2
        ORIENTATION=LANDSCAPE
        MRECALL MAUTOSOURCE SASAUTOS=(MACROS SASAUTOS)
        NOCENTER
        MISSING="."
        NODATE
        NONUMBER
        FORMDLIM="";


* create seed codes for all perturbed analysis
*=========================================;
*just run once at start of project;

title2;
title3;
title4;
*ODS LISTING CLOSE;

proc options option=work;
run;

proc datasets memtype=data library=work nolist kill;
run;
quit;

%global STDYPER
        REFSTDTN
        REFENDTN
        VAXCENS
        Period
        EVNTofINTEREST
        OUTCOMEsel
        AgeVar
    AgeLower
    AgeHigher
    AgeByIntv
        AgeAbove;

*---study period and references---*;
%let STDYPER             = JUNNOV22_O65PW_;      *----part of label for result datasets: eg
rates.&STDYPER.&EVNTofINTEREST._CrudeRates;
%let REFSTDTN            = 01JUN2022;                    *----start date of period of interest or study
start date or outbreak start date = reference start date--*;
%let REFENDTN            = 30NOV2022;                    *----stop date of period of interest or study
stop date or outbreak stop date = reference stop date--*;
%let SPLITDTN         = 01SEP2022;
%let VAXCENS          = 5;                            *----only for waning analysis: vaccination
number used as censoring for waning---*;

*---outcome---*;
%let Period           = 2;
%let EVNTofINTEREST   = UCOVDTH;          *--- outcome of interest;
                                                *---- UCOVDTH;*---CCOVDTH;*------CIRCDTH;
                                                *---- ISCHDTH;*---CRBRVASCDTH;*--RESPDTH;
```

```
                                                *---- DEMDTH;*---CANCERDTH;
*---age interval for analysis---*;
%let AgeVar           = AGE_AT_01JUN;       *----AGE_AT_01JAN or AGE_AT_01JUN;
%let AgeLower         = 65;                      *--- low bound for age interval (lower bound included)-
--*;
%let AgeHigher        = 130;                         *--- high bound for age interval : 130 to include
all elderly - (upper bound included)---*;
%let AgeByIntv        = 2;                              *--- number of years in each age interval to
create age category---;
%let AgeAbove         = 100;             *--- age after which we should create only one category---;

libname work list;

*SPLITperc +  VAXSTATUSC:SPLITperc + SEXCD +
agecatc + GPviscatc + comorbcatc +STEUCPc + HIEDcatc + F_FLU21C;

*the default method for Ties is Breslow however Efron approximation is more accurate when
dealing with tied death times, and is as efficient computationally. The exact method compute
the exact partial likelihood, which is equivalent to a conditional logistic model.
if there are a large number of ties the computational time will be excessive;
*=> using EFRON;

proc phreg data=INTV.&STDYPER.&EVNTofINTEREST;

  class  VAXSTATUSC(ref="000") SPLITstart SEXCD(ref="F") agecatc(ref="01")
         GPviscatc comorbcatc(ref="00") STEUCPc HIEDcatc(ref="02")
         F_FLU21C /param=glm; ;

  model (t1,t2)*F_&EVNTofINTEREST.INTV(0)= VAXSTATUSC SPLITstart VAXSTATUSC*SPLITstart
                                           SEXCD agecatc  GPviscatc comorbcatc STEUCPc
                                           HIEDcatc F_FLU21C / TIES = EFRON RL type3 ;
  lsmeans VAXSTATUSC*SPLITstart / exp cl diff;
run;


*************************************************************************************************
                                      E N D     O F    P R O G R A M
*************************************************************************************************;
```