Comparison of transient infections across viruses and covariates

Contents

This file does the comparisons analysis of the CMV transient infections between other viruses, infants with primary infection or not, and with breastfeeding.

1) Compare viruses

2) Compare by infection status

3) Compare by breastfeeding

Setup the data

```
blip_rates = arrange(all_blips, Virus, PatientID) %>% group_by(Virus, PatientID) %>%
  dplyr::summarize(blip_total = sum(count > 0),
                   total_swabs = n(),
                   infected = infantInfection[1],
                   total_time = as.numeric(difftime(max(times), min(times), units = "days")))

blip_rates$inf_label = factor(ifelse(blip_rates$infected == 1, "Yes", "No"),
                              levels = rev(c("Yes", "No")))
blip_rates$blip_prop = with(blip_rates, blip_total/total_swabs)
```
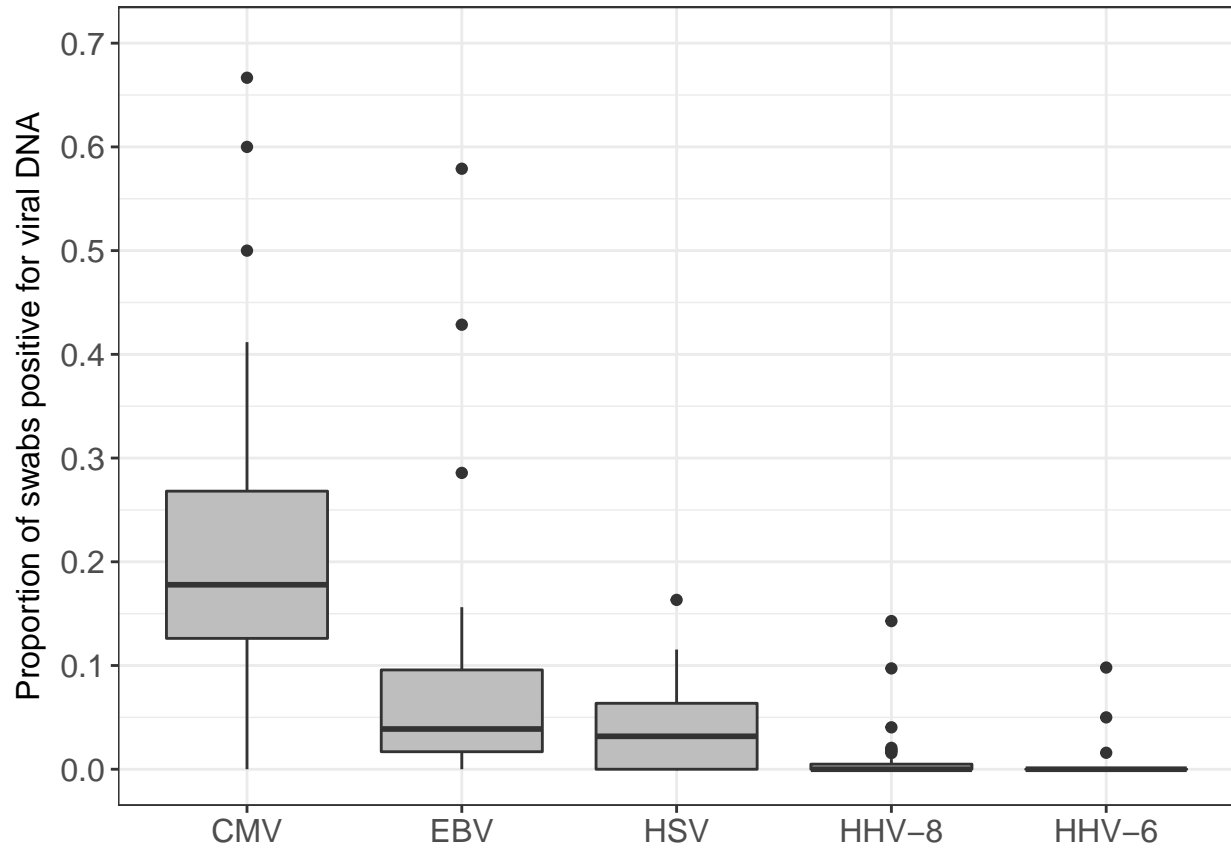
1.) By virus

```
blip_rates$Virus = factor(blip_rates$Virus, ordered = T)
blip_rates$Virus = with(blip_rates, reorder(Virus, -blip_prop))

blip_rates %>% group_by(Virus) %>% summarize(
    total = n(),
    mean = mean(round(blip_prop, 2)),
    median = median(round(blip_prop, 2)),
    IQR = paste(quantile(round(blip_prop, 2), c(0.25, 0.75)), collapse = ", "),
    range = paste(range(round(blip_prop, 2)), collapse = ", ")
) %>% xtable() %>% print()
```

| Virus	total	mean	median	IQR	range
CMV	30	0.21	0.17	0.1225, 0.265	0, 0.67
EBV	32	0.08	0.04	0.02, 0.0925	0, 0.58
HSV	32	0.04	0.03	0, 0.0625	0, 0.16
HHV8	31	0.01	0.00	0, 0.005	0, 0.14
HHV6	32	0.01	0.00	0, 0	0, 0.1

2. By infant infection status

all viruses

```
blip_rates %>% group_by(Virus, inf_label) %>% summarize(
    total = n(),
    mean = mean(round(blip_prop, 2)),
    median = median(round(blip_prop, 2)),
    IQR = paste(quantile(round(blip_prop, 2), c(0.25, 0.75)), collapse = ", "),
    range = paste(range(round(blip_prop, 2)), collapse = ", ")
) %>% xtable() %>% print()
```

Virus	inf_label	total	mean	median	IQR	range
CMV	No	10	0.17	0.18	0.135, 0.205	0, 0.3
CMV	Yes	20	0.23	0.17	0.1175, 0.275	0, 0.67
EBV	No	13	0.05	0.02	0.02, 0.06	0, 0.29
EBV	Yes	19	0.10	0.06	0.02, 0.105	0, 0.58
HSV	No	23	0.03	0.02	0, 0.055	0, 0.12
HSV	Yes	9	0.05	0.04	0.01, 0.07	0, 0.16
HHV8	No	31	0.01	0.00	0, 0.005	0, 0.14
HHV6	No	8	0.02	0.00	0, 0.0275	0, 0.1
HHV6	Yes	24	0.00	0.00	0, 0	0, 0
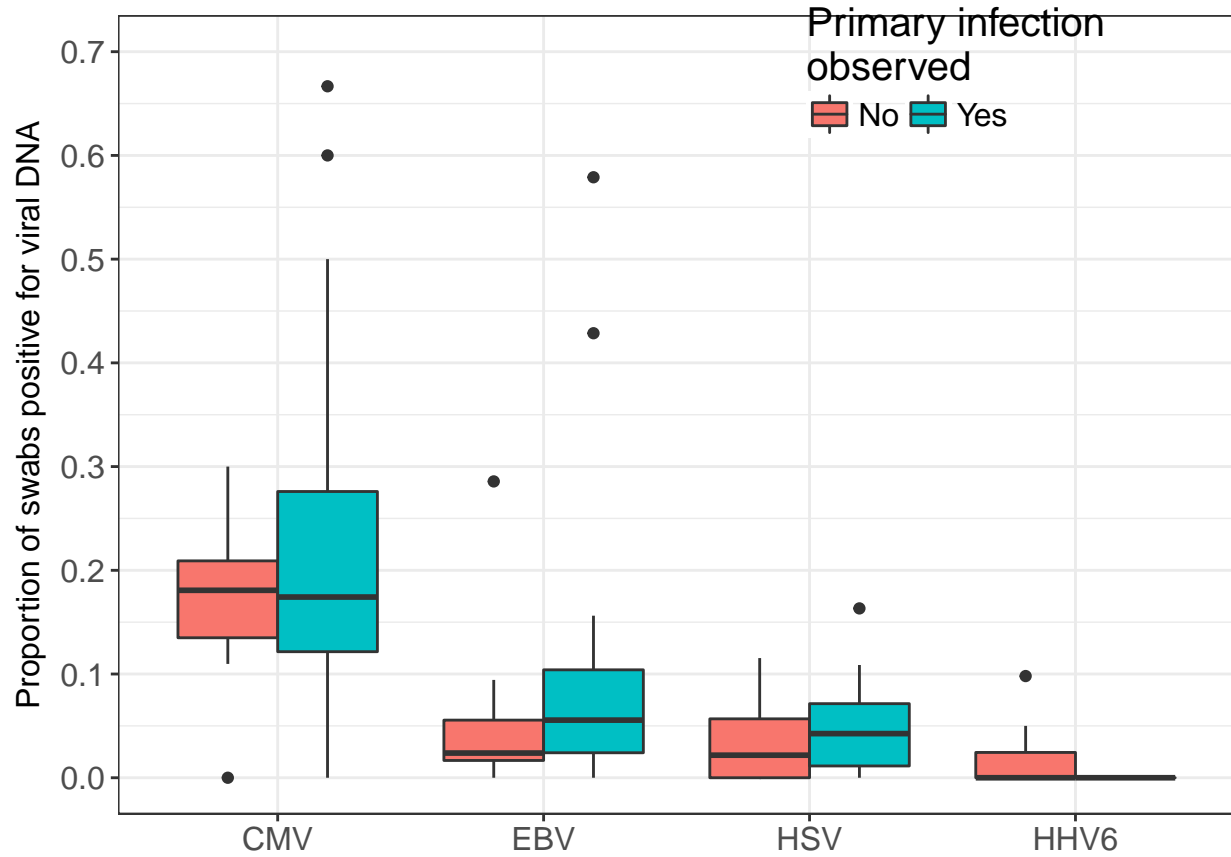
CMV only

```
subset(blip_rates, Virus == "CMV") %>% group_by(inf_label) %>% summarize(
    total = n(),
    mean = mean(round(blip_prop, 2)),
    median = median(round(blip_prop, 2)),
    IQR = paste(quantile(round(blip_prop, 2), c(0.25, 0.75)), collapse = ", "),
    range = paste(range(round(blip_prop, 2)), collapse = ", ")
) %>% xtable() %>% print()
```
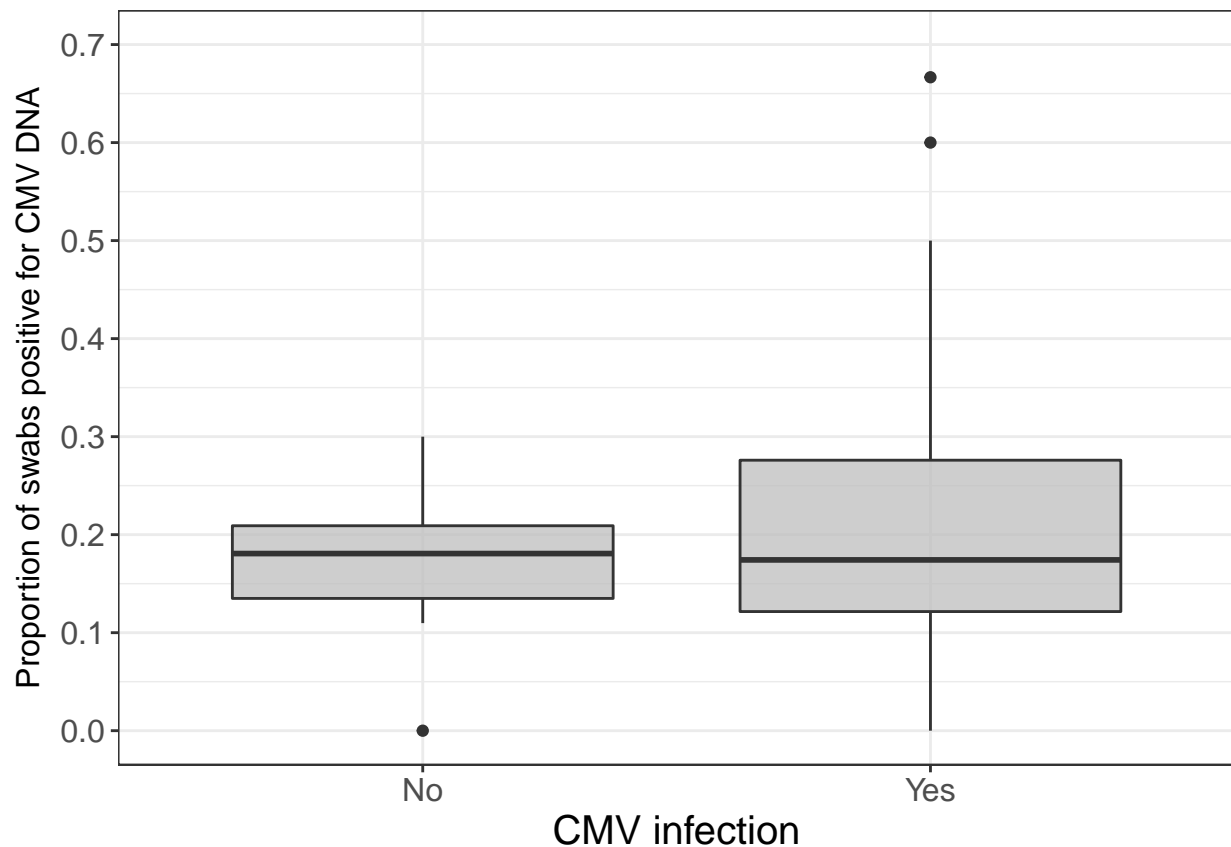
inf_label	total	mean	median	IQR	range
No	10	0.17	0.18	0.135, 0.205	0, 0.3
Yes	20	0.23	0.17	0.1175, 0.275	0, 0.67

3. Breastfeeding data

Note that consecutive positives are removed.

```
ts_covariates = readr::read_csv("data/demographics_data.csv")
```

```
## Parsed with column specification:
## cols(
##   PatientID2 = col_character(),
##   times = col_datetime(format = ""),
##   breastfed = col_integer(),
##   saliva = col_integer(),
##   chewfood = col_integer()
## )
```

```r
#extract negative swabs and then start of transient infections
cmv_blip_period =  subset(all_blips, Virus == "CMV") %>%
    filter(!consecutive_swab | is.na(consecutive_swab)) %>% filter(pos) %>%
  bind_rows(subset(all_blips, Virus == "CMV" & !pos))

bf_pos_swabs = left_join(cmv_blip_period, ts_covariates, by =  c("PatientID2", "times")) %>%
  group_by(PatientID2) %>%
  mutate(
    breastfed_ever = max(breastfed),
    saliva_ever = max(saliva),
    chewfood_ever = max(chewfood)
  )
```

```r
agg_function = function(var){
  subset(bf_pos_swabs, !is.na(get(var))) %>% group_by_(var) %>%
    dplyr::summarize(blip_total = sum(count > 0),
                     total_swabs = n(),
                     blip_prop = blip_total/total_swabs)
}

population_rates = bf_pos_swabs %>%
  group_by(PatientID2, breastfed_ever, saliva_ever, chewfood_ever) %>%
  summarize(
    blip_total = sum(count > 0),
    total_swabs = n(),
    blip_prop = blip_total/total_swabs
  ) %>%
  reshape2::melt(measure.vars = c("breastfed_ever", "saliva_ever", "chewfood_ever")) %>%
  group_by(variable) %>%
  mutate(
    total = n(),
    prop = mean(value)
  ) %>%
    group_by(variable, value, prop) %>%
  summarize(
    median_swabs = median(total_swabs),
    median_blip_prop = median(blip_prop),
    blip_prop_population = sum(blip_total)/sum(total_swabs)
  )

no_breastfeed_data = bf_pos_swabs %>%
  group_by(PatientID2) %>%
  mutate(
    no_bf_end_day = if(unique(breastfed_ever)) min(days[breastfed == 1]) else
      max(days) + 1
  ) %>%
  subset(days < no_bf_end_day)

#any(no_breastfeed_data$breastfed)
#mean(no_breastfeed_data$count > 0)
#mean(subset(no_breastfeed_data, !breastfed_ever)$count > 0)
#no_breastfeed_data %>% group_by(PatientID2) %>% summarize(blip_prop = sum(count > 0)/n()) %>% summariz


covariate_rates = plyr::ldply(c("breastfed", "saliva", "chewfood"), function(x) {
  out = agg_function(x)
  z = which(colnames(out) == x)
  out$value = as.numeric(as.matrix((as.data.frame(out[, z]))))
  out$var = x
  out[, -z]
})


select(covariate_rates, var, value, everything()) %>% ungroup() %>% xtable(caption="Not by patient, by p
```

var	value	blip_total	total_swabs	blip_prop
breastfed	0.00	46	368	0.12
breastfed	1.00	90	492	0.18
saliva	0.00	131	838	0.16
saliva	1.00	5	22	0.23
chewfood	0.00	132	826	0.16
chewfood	1.00	4	34	0.12

Table 1: Not by patient, by positive swab status

```
ungroup(population_rates) %>% xtable(caption="Positive swab proportion by patient. (population = pooled
```

variable	value	prop	median_swabs	median_blip_prop	blip_prop_population
breastfed_ever	0	0.87	38.00	0.16	0.13
breastfed_ever	1	0.87	12.50	0.15	0.16
saliva_ever	0	0.27	12.50	0.16	0.16
saliva_ever	1	0.27	53.00	0.12	0.15
chewfood_ever	0	0.30	11.00	0.14	0.15
chewfood_ever	1	0.30	53.00	0.17	0.16

Table 2: Positive swab proportion by patient. (population = pooled proportion).

```r
bf_pos_swabs_all =  subset(all_blips, Virus == "CMV") %>%
  left_join(ts_covariates, by =  c("PatientID2", "times")) %>%
  group_by(PatientID2) %>%
  mutate(
    breastfed_ever = max(breastfed, na.rm = T), #there is a missing during a transient infection
    saliva_ever = max(saliva),
    chewfood_ever = max(chewfood)
  )

all_swab_rates = bf_pos_swabs_all %>%
  group_by(PatientID2, breastfed_ever, saliva_ever, chewfood_ever) %>%
  summarize(
    blip_total = sum(count > 0),
    total_swabs = n(),
    blip_prop = blip_total/total_swabs
  ) %>%
  reshape2::melt(measure.vars = c("breastfed_ever", "saliva_ever", "chewfood_ever")) %>%
  group_by(variable) %>%
  mutate(
    total = n(),
    prop = mean(value)
  ) %>%
  group_by(variable, value, prop) %>%
  summarize(
    median_swabs = median(total_swabs),
    median_blip_prop = median(blip_prop),
    blip_prop_population = sum(blip_total)/sum(total_swabs)
  )

ungroup(all_swab_rates) %>% xtable(caption="Positive swab proportion by all swabs. (population = pooled
```

variable	value	prop	median_swabs	median_blip_prop	blip_prop_population
breastfed_ever	0	0.87	43.50	0.21	0.19
breastfed_ever	1	0.87	12.50	0.18	0.22
saliva_ever	0	0.30	14.00	0.18	0.23
saliva_ever	1	0.30	55.00	0.16	0.20
chewfood_ever	0	0.33	10.50	0.17	0.20
chewfood_ever	1	0.33	56.00	0.25	0.23

Table 3: Positive swab proportion by all swabs. (population = pooled proportion).