

Script 2c

Descriptive stats by pain status

Peter Kamerman

22 January 2020

Contents

Analysis notes	1
Import data	1
First look	2
Basic clean	2
Quick tabulation	2
Analysis data set for the period 0 to 48 weeks	2
Analysis	2
Prepare summary data for tabulation and plotting	2
Tabulate	3
Bivarite plots	4
Session information	6

Analysis notes

Pain was defined as *pain in the last week* being ‘Yes’, and *pain at its worst* being > 0 .

These two measurements were then the “gatekeeper” measurements, such that the two measurements both had to be positive (‘Yes’ and ‘ > 0 ’, respectively) in order for there to be any entries for *site of pain* and *site of worst pain*. Were the data were incongruent (e.g., when there was no *pain in the last week* and *pain at its worst* = 0, but there were entries for *site of pain* and *site of worst pain*), then the *site of pain* and *site of worst pain* entries were marked as **incongruent**.

Data also were considered **incongruent** when *pain in the last week* = ‘Yes’, but *site of worst pain* = ‘None’.

Lastly, data were considered **incongruent** when *site of worst pain* was not listed as one of the pain locations for a given measurement week.

For analysis purposes, missing data in the *site of pain* columns were changed to ‘**No**’ (pain not present in the site). This approach was conservative, but we believed that the approach would have the least effect on the outcome, while still retaining as many participants as possible.

Import data

```
df <- read_rds('data-cleaned/data-ADVANCE.rds') %>%
  select(ranid, interval_name, pain_in_the_last_week,
         cd4_cells.ul, viral_load_cp.ml,
         any_missing, interval_numeric)
```

First look

```
head(df)
```

```
## # A tibble: 6 x 7
##   ranid interval_name pain_in_the_las... cd4_cells.ul viral_load_cp.ml
##   <chr> <ord>          <chr>          <dbl>          <dbl>
## 1 01-0... 0 weeks      No              642            641
## 2 01-0... 12 weeks     No              NA              50
## 3 01-0... 24 weeks     No             525            50
## 4 01-0... 36 weeks     No              NA              50
## 5 01-0... 48 weeks     No             668            50
## 6 01-0... 0 weeks      No             241           3851
## # ... with 2 more variables: any_missing <chr>, interval_numeric <dbl>
```

```
glimpse(df)
```

```
## Observations: 5,265
## Variables: 7
## $ ranid          <chr> "01-0001", "01-0001", "01-0001", "01-0001"...
## $ interval_name  <ord> 0 weeks, 12 weeks, 24 weeks, 36 weeks, 48 ...
## $ pain_in_the_last_week <chr> "No", "No", "No", "No", "No", "No", "Yes",...
## $ cd4_cells.ul   <dbl> 642, NA, 525, NA, 668, 241, NA, 364, NA, 4...
## $ viral_load_cp.ml <dbl> 641, 50, 50, 50, 50, 3851, 50, 50, 50, 50,...
## $ any_missing    <chr> "No", "No", "No", "No", "No", "No", "No", ...
## $ interval_numeric <dbl> 0, 12, 24, 36, 48, 0, 12, 24, 36, 48, 0, 1...
```

Basic clean

```
# Clean and process data
df %<>%
  filter(any_missing == 'No') %>%
  select(-any_missing)
```

Quick tabulation

Analysis data set for the period 0 to 48 weeks

```
# Tabulate data
xtabs(~interval_name, data = df)
```

```
## interval_name
## 0 weeks 12 weeks 24 weeks 36 weeks 48 weeks
##      787      787      787      787      787
```

Analysis

Prepare summary data for tabulation and plotting

```
# Pain in the last week
df_pain <- as.data.frame(xtabs(~interval_name + pain_in_the_last_week,
```

```

                                data = df)) %>%
  rename(count = Freq) %>%
  group_by(interval_name) %>%
  mutate(total = sum(count)) %>%
  mutate(proportion = round(count / total, 3)) %>%
  filter(pain_in_the_last_week == 'Yes')

# Viral load
df_vl <- df %>%
  group_by(interval_name) %>%
  mutate(viral_load_cp.ml = ifelse(viral_load_cp.ml < 50,
                                   yes = 50,
                                   no = viral_load_cp.ml)) %>%
  summarise(missing = sum(is.na(viral_load_cp.ml)),
            median = median(viral_load_cp.ml, na.rm = TRUE),
            Q25 = quantile(viral_load_cp.ml, probs = 0.25, na.rm = TRUE),
            Q75 = quantile(viral_load_cp.ml, probs = 0.75, na.rm = TRUE),
            min = min(viral_load_cp.ml, na.rm = TRUE),
            max = max(viral_load_cp.ml, na.rm = TRUE))

# CD4
df_cd4 <- df %>%
  group_by(interval_name) %>%
  summarise(missing = sum(is.na(cd4_cells.ul)),
            median = median(cd4_cells.ul, na.rm = TRUE),
            Q25 = quantile(cd4_cells.ul, probs = 0.25, na.rm = TRUE),
            Q75 = quantile(cd4_cells.ul, probs = 0.75, na.rm = TRUE),
            min = min(cd4_cells.ul, na.rm = TRUE),
            max = max(cd4_cells.ul, na.rm = TRUE))

df_cd4 <- df_cd4[c(1, 3, 5), ]

```

Tabulate

```

# Pain
df_pain %>%
  knitr::kable(caption = 'Proportion with pain in the last week')

```

Table 1: Proportion with pain in the last week

interval_name	pain_in_the_last_week	count	total	proportion
0 weeks	Yes	151	787	0.192
12 weeks	Yes	143	787	0.182
24 weeks	Yes	127	787	0.161
36 weeks	Yes	114	787	0.145
48 weeks	Yes	96	787	0.122

```

# CD4
df_cd4 %>%
  knitr::kable(caption = '5-number summary of CD4 T-cell count (cells/ul)')

```

Table 2: 5-number summary of CD4 T-cell count (cells/ul)

interval_name	missing	median	Q25	Q75	min	max
0 weeks	0	290	173.50	441.50	1	1721

interval_name	missing	median	Q25	Q75	min	max
24 weeks	11	411	277.75	604.25	5	1615
48 weeks	16	450	313.50	629.00	75	1626

```
# Viral load
df_vl %>%
  knitr::kable(caption = '5-number summary of viral load (copies/ml)')
```

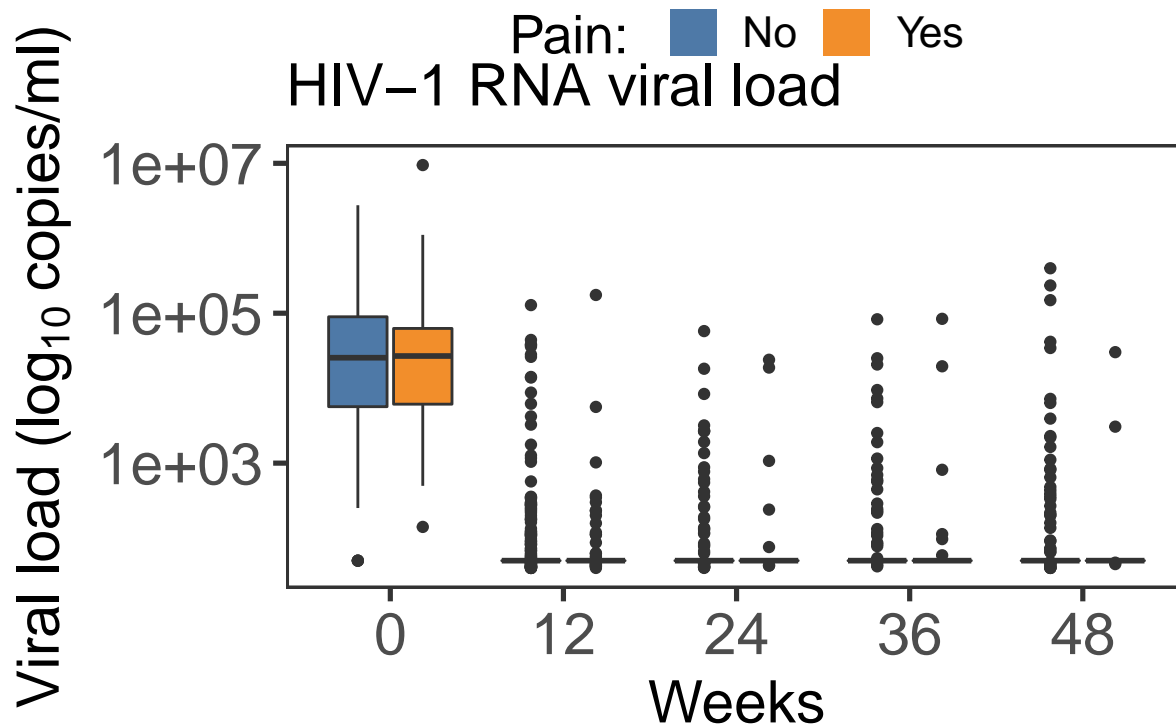
Table 3: 5-number summary of viral load (copies/ml)

interval_name	missing	median	Q25	Q75	min	max
0 weeks	0	25853	5704.5	85574	50	9475772
12 weeks	3	50	50.0	50	50	175168
24 weeks	2	50	50.0	50	50	57754
36 weeks	9	50	50.0	50	50	84167
48 weeks	9	50	50.0	50	50	397926

Bivarite plots

```
# Viral load
p_vl <- df %>%
  select(interval_name, pain_in_the_last_week, viral_load_cp.ml) %>%
  ggplot(data = .) +
  aes(y = viral_load_cp.ml,
      x = interval_name,
      fill = pain_in_the_last_week) +
  geom_boxplot(key_glyph = "rect") +
  labs(title = 'HIV-1 RNA viral load ',
       y = expression('Viral load (log'[10]~'copies/ml)'),
       x = 'Weeks') +
  scale_x_discrete(labels = c('0', '12', '24', '36', '48')) +
  scale_y_log10() +
  scale_fill_tableau(name = 'Pain: ') +
  theme(plot.margin = margin(t = 4, l = 1, r = 1, b = 1, "lines"),
        legend.direction = 'horizontal',
        legend.position = c(0.5, 1.25),
        axis.title = element_text(size = 22),
        plot.title = element_text(size = 22),
        axis.text = element_text(size = 22),
        panel.grid = element_blank()); p_vl
```

```
## Warning: Removed 23 rows containing non-finite values (stat_boxplot).
```

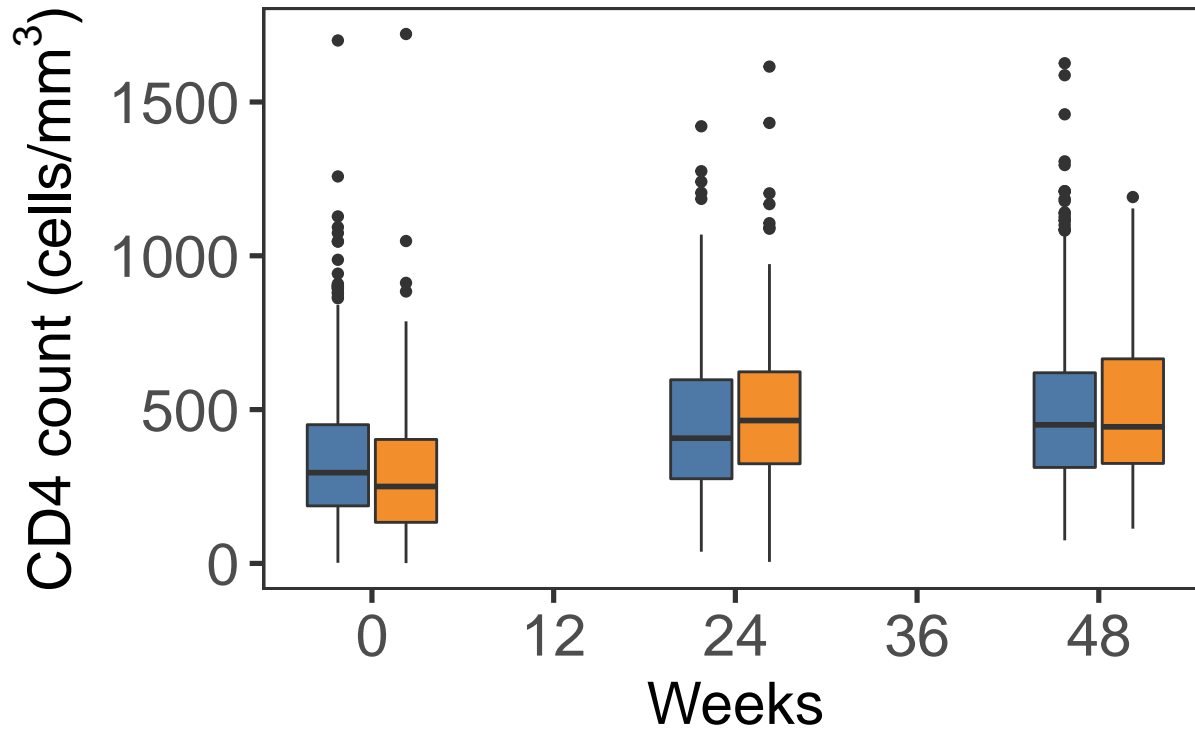


```
# CD4
p_cd4 <- df %>%
  select(interval_name, interval_numeric,
         pain_in_the_last_week, cd4_cells.ul) %>%
  mutate(cd4_cells.ul = ifelse(interval_numeric == 12 | interval_numeric == 36,
                              yes = NA,
                              no = cd4_cells.ul)) %>%

  ggplot(data = .) +
  aes(y = cd4_cells.ul,
      x = interval_name,
      fill = pain_in_the_last_week) +
  geom_boxplot() +
  labs(title = 'CD4 T-cell count',
       y = expression('CD4 count (cells/mm'^3*')),
       x = 'Weeks') +
  scale_x_discrete(labels = c('0', '12', '24', '36', '48')) +
  scale_fill_tableau(name = 'Pain in the\nlast week') +
  theme(legend.position = 'none',
        axis.title = element_text(size = 22),
        plot.title = element_text(size = 22),
        axis.text = element_text(size = 22),
        panel.grid = element_blank()); p_cd4
```

```
## Warning: Removed 1601 rows containing non-finite values (stat_boxplot).
```

CD4 T-cell count



Session information

```
sessionInfo()
```

```
## R version 3.6.1 (2019-07-05)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS Mojave 10.14.6
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
## other attached packages:
## [1] patchwork_0.0.1 ggthemes_4.2.0 magrittr_1.5   forcats_0.4.0
## [5] stringr_1.4.0   dplyr_0.8.3   purrr_0.3.3   readr_1.3.1
## [9] tidyr_1.0.0     tibble_2.1.3  ggplot2_3.2.1 tidyverse_1.2.1
##
## loaded via a namespace (and not attached):
## [1] tidyselect_0.2.5 xfun_0.10      haven_2.1.1    lattice_0.20-38
## [5] colorspace_1.4-1 vctrs_0.2.0    generics_0.0.2 htmltools_0.4.0
## [9] yaml_2.2.0       utf8_1.1.4     rlang_0.4.2    pillar_1.4.2
## [13] glue_1.3.1       withr_2.1.2    modelr_0.1.5   readxl_1.3.1
## [17] lifecycle_0.1.0  munsell_0.5.0  gtable_0.3.0   cellranger_1.1.0
```

## [21]	rvest_0.3.4	evaluate_0.14	labeling_0.3	knitr_1.25
## [25]	fansi_0.4.0	highr_0.8	broom_0.5.2	Rcpp_1.0.3
## [29]	scales_1.0.0	backports_1.1.5	jsonlite_1.6	hms_0.5.1
## [33]	digest_0.6.23	stringi_1.4.3	grid_3.6.1	cli_2.0.0
## [37]	tools_3.6.1	lazyeval_0.2.2	crayon_1.3.4	pkgconfig_2.0.3
## [41]	zeallot_0.1.0	xml2_1.2.2	lubridate_1.7.4	assertthat_0.2.1
## [45]	rmarkdown_1.16	httr_1.4.1	rstudioapi_0.10	R6_2.4.1
## [49]	nlme_3.1-141	compiler_3.6.1		