

Supplement 3

Number of pain sites

Peter Kamerman

Last knitted: 03 March 2021

Contents

1	Import and check data	1
1.1	Add acute/chronic pain column to demographic data	4
2	Data analysis	5
2.1	Process data	5
2.2	Summary statistics	5
2.2.1	Total group	5
2.2.2	For each categorical variable	6
2.2.2.1	By sex	6
2.2.2.2	By HAART	6
2.2.2.3	By education	7
2.2.2.4	By employment status	7
2.2.2.5	By pain definition	8
2.3	Exploratory plots	8
2.3.1	Distribution of pain counts	8
2.3.1.1	Overall count frequency	8
2.3.1.2	Count by sex	10
2.3.1.3	Count by HAART	12
2.3.1.4	Count education	14
2.3.1.5	Count by employment status	16
2.3.1.6	Count by pain definition	18
2.4	Regression analysis	20
2.4.1	Process data	20
2.4.2	Mixed model negative binomial regression for count data	20
2.4.3	Negative binomial regression for count data	21
2.5	Publication plot	22
3	Session information	25

1 Import and check data

```
# Import
data <- read_rds('data-cleaned/data-pain-sites.rds')
demo <- read_rds('data-cleaned/data-demographics.rds')

# Check
```

```
## Pain sites
dim(data)

## [1] 596 21

names(data)

## [1] "ID" "Head" "Throat"
## [4] "Shoulder" "Arms" "Elbows"
## [7] "Wrists.Hands" "Chest" "Upper_back"
## [10] "Lower_back" "Abdomen" "Cervical_spine"
## [13] "Thoracic_spine" "Lumbosacral_spine" "Groin"
## [16] "Hips" "Legs" "Knees"
## [19] "Ankles.Feet" "Buttocks" "Site"

glimpse(data)

## Rows: 596
## Columns: 21
## $ ID <chr> "RPB73", "RPB74", "RPB75", "RPB76", "RPB77", "RPB...
## $ Head <chr> "No", "No", "No", "Yes", "Yes", "No", "No", "No", ...
## $ Throat <chr> "No", "No", "No", "No", "No", "No", "No", "No", "..."
## $ Shoulder <chr> "No", "No", "No", "No", "No", "No", "No", "No", "..."
## $ Arms <chr> "No", "No", "No", "No", "No", "No", "No", "No", "..."
## $ Elbows <chr> "No", "No", "No", "No", "No", "No", "No", "No", "..."
## $ Wrists.Hands <chr> "No", "No", "No", "No", "No", "No", "No", "No", "..."
## $ Chest <chr> "No", "No", "No", "Yes", "No", "No", "No", "No", "..."
## $ Upper_back <chr> "No", "No", "No", "No", "No", "No", "No", "No", "..."
## $ Lower_back <chr> "No", "No", "No", "No", "No", "No", "No", "No", "..."
## $ Abdomen <chr> "No", "No", "Yes", "Yes", "No", "No", "Yes", "No", "..."
## $ Cervical_spine <chr> "No", "No", "No", "No", "No", "No", "No", "No", "..."
## $ Thoracic_spine <chr> "No", "No", "No", "No", "No", "No", "No", "No", "..."
## $ Lumbosacral_spine <chr> "No", "No", "No", "No", "No", "No", "No", "No", "..."
## $ Groin <chr> "No", "No", "No", "No", "No", "No", "No", "No", "..."
## $ Hips <chr> "No", "No", "No", "No", "No", "No", "No", "Yes", "..."
## $ Legs <chr> "No", "No", "No", "No", "No", "Yes", "No", "Yes", "..."
## $ Knees <chr> "No", "No", "No", "No", "No", "Yes", "No", "Yes", "..."
## $ Ankles.Feet <chr> "No", "No", "No", "No", "No", "Yes", "No", "Yes", "..."
## $ Buttocks <chr> "No", "No", "No", "No", "No", "No", "No", "Yes", "..."
## $ Site <chr> "RP", "RP", "RP", "RP", "RP", "RP", "RP", "RP", "..."

data %>%
  select(-ID, -Site) %>%
  mutate_if(is.character, factor) %>%
  skim()
```

Table 1: Data summary

Name	Piped data
Number of rows	596
Number of columns	19
Column type frequency:	
factor	19
Group variables	None

Variable type: factor

skim_variable	n_missing	complete_rate	n_unique	top_counts
Head	0	1	2	No: 401, Yes: 195
Throat	0	1	2	No: 577, Yes: 19
Shoulder	0	1	2	No: 549, Yes: 47
Arms	0	1	2	No: 571, Yes: 25
Elbows	0	1	2	No: 574, Yes: 22
Wrists.Hands	0	1	2	No: 561, Yes: 35
Chest	0	1	2	No: 479, Yes: 117
Upper_back	0	1	1	No: 596
Lower_back	0	1	2	No: 554, Yes: 42
Abdomen	0	1	2	No: 437, Yes: 159
Cervical_spine	0	1	2	No: 566, Yes: 30
Thoracic_spine	0	1	2	No: 522, Yes: 74
Lumbosacral_spine	0	1	2	No: 504, Yes: 92
Groin	0	1	2	No: 542, Yes: 54
Hips	0	1	2	No: 556, Yes: 40
Legs	0	1	2	No: 497, Yes: 99
Knees	0	1	2	No: 512, Yes: 84
Ankles.Feet	0	1	2	No: 412, Yes: 184
Buttocks	0	1	2	No: 577, Yes: 19

```
## Demographics
```

```
dim(demo)
```

```
## [1] 596 8
```

```
names(demo)
```

```
## [1] "ID"           "Site"          "Sex"
## [4] "Age"          "Employment_status" "CD4_recent"
## [7] "ART_currently" "Education"
```

```
glimpse(demo)
```

```
## Rows: 596
## Columns: 8
## $ ID      <chr> "RPB73", "RPB74", "RPB75", "RPB76", "RPB77", "RPB...
## $ Site    <chr> "RP", "RP", "RP", "RP", "RP", "RP", "RP", "RP", "...
## $ Sex     <chr> "Female", "Female", "Female", "Female", "Female",...
## $ Age     <dbl> 36, 27, 39, 36, 31, 32, 28, 37, 31, 25, 31, 24, 3...
## $ Employment_status <chr> "Other", "Unemployed", "Other", "Unemployed", "Un...
## $ CD4_recent <dbl> 391, 571, 591, 207, 126, 225, 543, 410, 74, 212, ...
## $ ART_currently <chr> "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", ...
## $ Education <chr> "Tertiary", "Secondary", "Secondary", "Primary", ...
```

```
demo %>%
```

```
  select(-ID, -Site) %>%
```

```
  mutate_if(is.character, factor) %>%
```

```
  skim()
```

Table 3: Data summary

Name	Piped data
------	------------

Table 3: Data summary

Number of rows	596
Number of columns	6
Column type frequency:	
factor	4
numeric	2
Group variables	None

Variable type: factor

skim_variable	n_missing	complete_rate	n_unique	top_counts
Sex	0	1.00	2	Fem: 481, Mal: 115
Employment_status	49	0.92	4	Une: 330, Ful: 131, Par: 52, Oth: 34
ART_currently	5	0.99	2	Yes: 460, No: 131
Education	37	0.94	3	Sec: 395, Pri: 99, Ter: 65

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
Age	8	0.99	37.28	9.06	19	31	36	42	76
CD4_recent	99	0.83	320.71	238.92	1	155	261	432	1232

1.1 Add acute/chronic pain column to demographic data

```

# Based on pain definition used in the study
## Acute assessing pain in the last 1 week or 1 month.
## Chronic assessing persistent pain in the last 3 months
demo <- demo %>%
  mutate(Pain_def = case_when(
    Site == 'RP' |
    Site == 'NM' ~ 'Acute',
    TRUE ~ 'Chronic'
  ))

# Tabulate
demo %>%
  select(Pain_def) %>%
  mutate(Pain_def = factor(Pain_def)) %>%
  skim()

```

Table 6: Data summary

Name	Piped data
Number of rows	596
Number of columns	1
Column type frequency:	

Table 6: Data summary

factor	1
Group variables	None

Variable type: factor

skim_variable	n_missing	complete_rate	n_unique	top_counts
Pain_def	0	1	2	Acu: 387, Chr: 209

2 Data analysis

2.1 Process data

```
# Remove upper_back (only one outcome -- no pain)
data <- data[, !(names(data) %in% c('Upper_back'))]

# Calculate number of sites per individual
pain_count <- data %>%
  # Format as long data
  pivot_longer(cols = -c(ID, Site),
               names_to = 'pain_site',
               values_to = 'pain_present') %>%
  group_by(ID) %>%
  summarise(Count = sum(pain_present == 'Yes', na.rm = TRUE))

# Join datasets
analysis_set <- left_join(pain_count, demo)
```

2.2 Summary statistics

2.2.1 Total group

```
analysis_set %>%
  select(Count) %>%
  skim()
```

Table 8: Data summary

Name	Piped data
Number of rows	596
Number of columns	1
Column type frequency:	
numeric	1
Group variables	None

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
Count	0	1	2.24	1.68	0	1	2	3	12

2.2.2 For each categorical variable

2.2.2.1 By sex

```
analysis_set %>%
  select(Count, Sex) %>%
  group_by(Sex) %>%
  skim()
```

Table 10: Data summary

Name	Piped data
Number of rows	596
Number of columns	2
Column type frequency: numeric	1
Group variables	Sex

Variable type: numeric

skim_variable	Sex	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
Count	Female	0	1	2.27	1.69	0	1	2	3	12
Count	Male	0	1	2.13	1.63	0	1	2	3	9

2.2.2.2 By HAART

```
analysis_set %>%
  select(Count, ART_currently) %>%
  filter(complete.cases(.)) %>%
  group_by(ART_currently) %>%
  skim()
```

Table 12: Data summary

Name	Piped data
Number of rows	591
Number of columns	2
Column type frequency: numeric	1
Group variables	ART_currently

Variable type: numeric

skim_variable	ART_currently	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
Count	No	0	1	2.18	1.64	0	1	2	3	12
Count	Yes	0	1	2.27	1.70	0	1	2	3	9

2.2.2.3 By education

```
analysis_set %>%
  select(Count, Education) %>%
  filter(complete.cases()) %>%
  group_by(Education) %>%
  skim()
```

Table 14: Data summary

Name	Piped data
Number of rows	559
Number of columns	2
Column type frequency: numeric	1
Group variables	Education

Variable type: numeric

skim_variable	Education	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
Count	Primary	0	1	2.38	1.64	0	1	2	3	8
Count	Secondary	0	1	2.24	1.73	0	1	2	3	12
Count	Tertiary	0	1	2.06	1.52	0	1	2	3	6

2.2.2.4 By employment status

```
analysis_set %>%
  select(Count, Employment_status) %>%
  filter(complete.cases()) %>%
  group_by(Employment_status) %>%
  skim()
```

Table 16: Data summary

Name	Piped data
Number of rows	547
Number of columns	2
Column type frequency: numeric	1
Group variables	Employment_status

Variable type: numeric

skim_variable	Employment_status	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
Count	Full-time work	0	1	2.21	1.46	0	1	2	3	6
Count	Other	0	1	2.18	1.64	0	1	2	4	5
Count	Part-time work	0	1	2.50	1.89	0	1	2	3	9
Count	Unemployed	0	1	2.21	1.75	0	1	2	3	12

2.2.2.5 By pain definition

```
analysis_set %>%
  select(Count, Pain_def) %>%
  group_by(Pain_def) %>%
  skim()
```

Table 18: Data summary

Name	Piped data
Number of rows	596
Number of columns	2
Column type frequency: numeric	1
Group variables	Pain_def

Variable type: numeric

skim_variable	Pain_def	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
Count	Acute	0	1	2.19	1.74	0	1	2	3	12
Count	Chronic	0	1	2.33	1.57	0	1	2	3	9

2.3 Exploratory plots

2.3.1 Distribution of pain counts

2.3.1.1 Overall count frequency

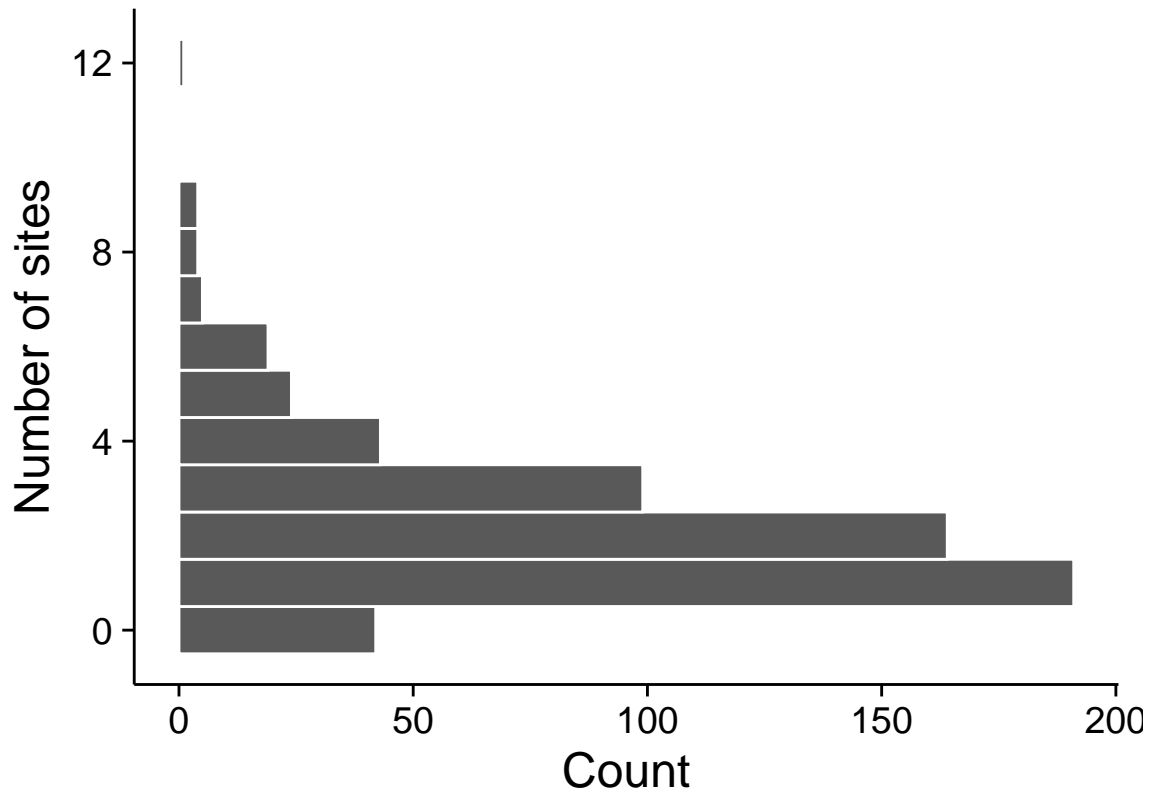
```
all <- analysis_set %>%
  select(Count) %>%
  filter(complete.cases(.)) %>%
  ggplot(data = .) +
  aes(Count) +
  geom_histogram(binwidth = 1,
                 colour = '#FFFFFF') +
  labs(title = 'All participants',
        subtitle = str_glue('(Complete cases = {nrow(analysis_set[!is.na(
          analysis_set$Count), ]})}')),
        x = 'Number of sites',
        y = 'Count') +
  coord_flip() +
  theme_minimal(base_size = 18) +
```



```
theme(plot.title = element_text(size = 18),
      plot.subtitle = element_text(size = 12),
      panel.grid = element_blank(),
      axis.text = element_text(colour = '#000000'),
      axis.line = element_line(size = 0.5),
      axis.ticks = element_line(size = 0.5)); all
```

All participants

(Complete cases = 596)



```
all <- analysis_set %>%
  select(Count) %>%
  filter(complete.cases(.)) %>%
  ggplot(data = .) +
  aes(y = Count,
      x = 'All participants') +
  geom_boxplot(outlier.colour = '#FFFFFF',
              outlier.size = 0,
              width = 0.5) +
  geom_point(size = 2,
            shape = 21,
            position = position_jitter(height = 0.2, width = 0.2),
            fill = '#000000',
            colour = '#000000',
            alpha = 0.5) +
  labs(title = 'All participants',
       subtitle = str_glue('(Complete cases = {nrow(analysis_set[!is.na(
         analysis_set$Count), ]})}'))
```

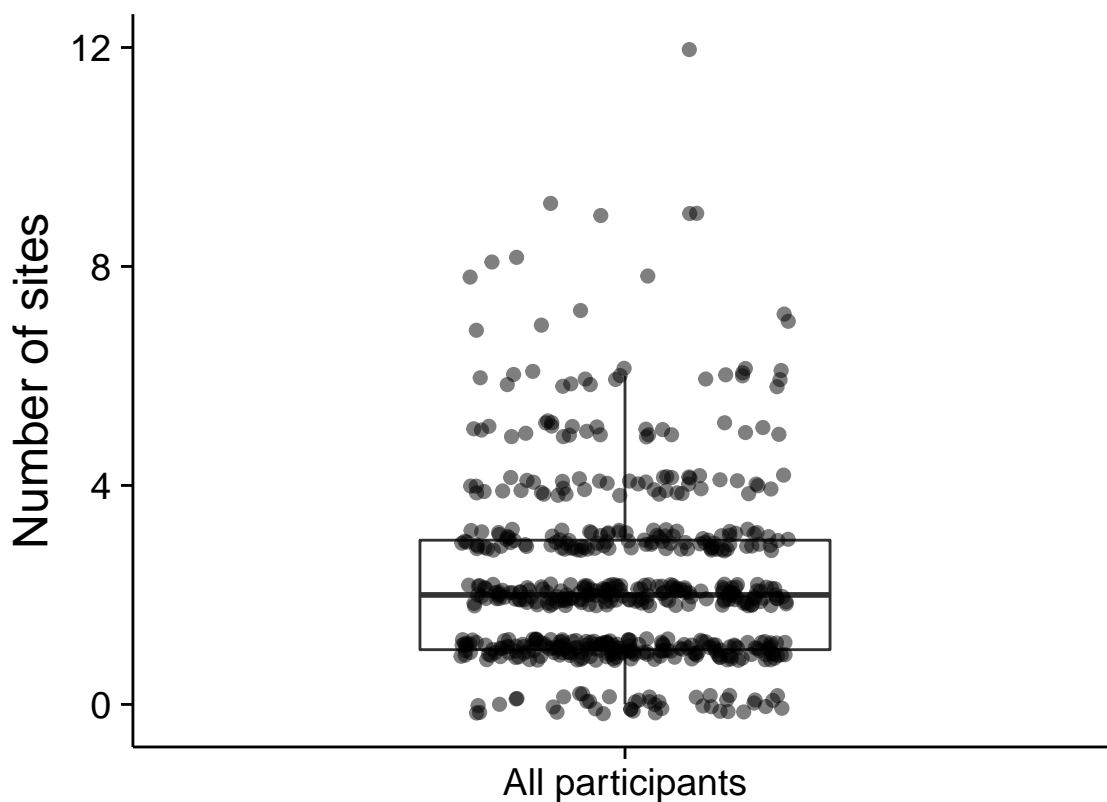
```

y = 'Number of sites') +
theme_minimal(base_size = 18) +
theme(plot.title = element_text(size = 18),
      plot.subtitle = element_text(size = 12),
      plot.caption = element_text(size = 12),
      panel.grid = element_blank(),
      axis.title.x = element_blank(),
      axis.text = element_text(colour = '#000000'),
      axis.line = element_line(size = 0.5),
      axis.ticks = element_line(size = 0.5)); all

```

All participants

(Complete cases = 596)



2.3.1.2 Count by sex

```

analysis_set %>%
  select(Count, Sex) %>%
  filter(complete.cases(.)) %>%
  ggplot(data = .) +
  aes(Count) +
  geom_histogram(binwidth = 1,
                colour = '#FFFFFF') +
  labs(title = 'Sex',
       subtitle = str_glue('(Complete cases = {nrow(analysis_set[!is.na(
         analysis_set$Sex), ]})}')),
       x = 'Number of sites',

```

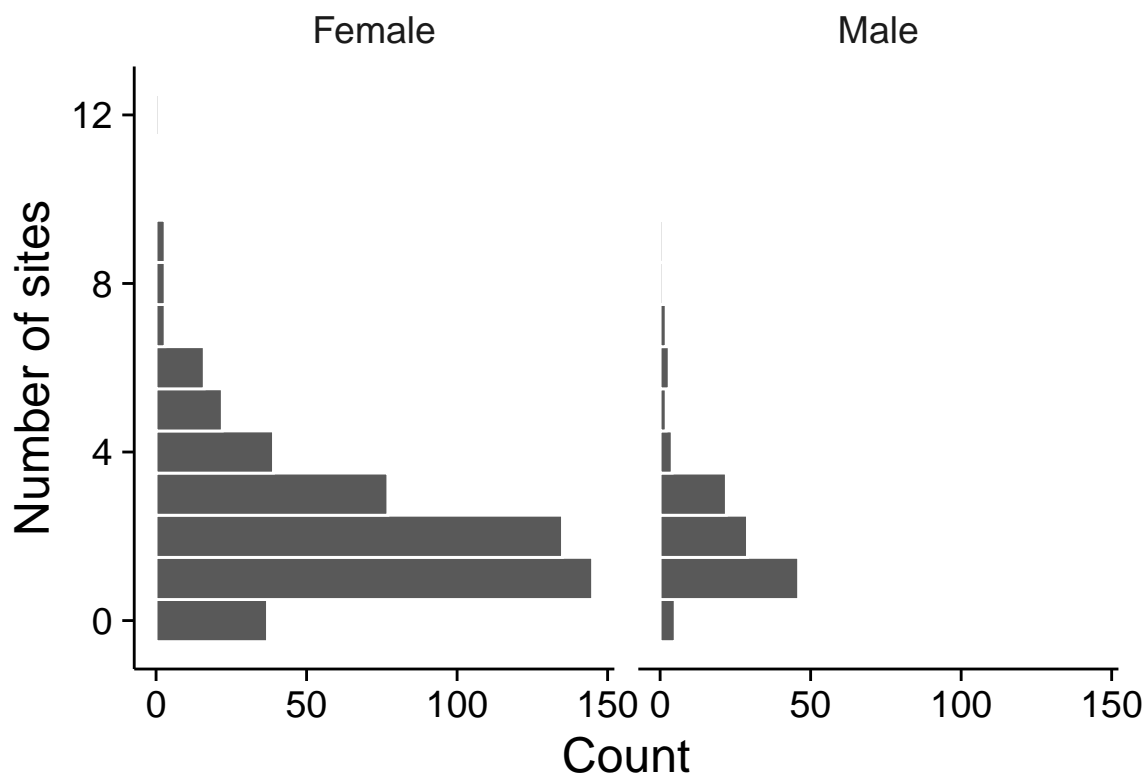
```

y = 'Count') +
facet_wrap(~Sex) +
coord_flip() +
theme_minimal(base_size = 18) +
theme(plot.title = element_text(size = 18),
      plot.subtitle = element_text(size = 12),
      panel.grid = element_blank(),
      axis.text = element_text(colour = '#000000'),
      axis.line = element_line(size = 0.5),
      axis.ticks = element_line(size = 0.5))

```

Sex

(Complete cases = 596)



```

sex <- analysis_set %>%
  select(Count, Sex) %>%
  filter(complete.cases(.)) %>%
  ggplot(data = .) +
  aes(y = Count,
      x = Sex) +
  geom_boxplot(outlier.colour = '#FFFFFF',
              outlier.size = 0) +
  geom_point(size = 2,
            shape = 21,
            position = position_jitter(height = 0.2, width = 0.2),
            fill = '#000000',
            colour = '#000000',
            alpha = 0.5) +

```

```

scale_x_discrete(labels = c('Female', 'Male')) +
labs(title = 'Sex',
      subtitle = str_glue('(Complete cases = {nrow(analysis_set[!is.na(
        analysis_set$Sex), ]})}')),
      y = 'Number of sites') +
theme_minimal(base_size = 18) +
theme(plot.title = element_text(size = 18),
      plot.subtitle = element_text(size = 12),
      plot.caption = element_text(size = 12),
      panel.grid = element_blank(),
      axis.title.x = element_blank(),
      axis.text = element_text(colour = '#000000'),
      axis.line = element_line(size = 0.5),
      axis.ticks = element_line(size = 0.5))

```

2.3.1.3 Count by HAART

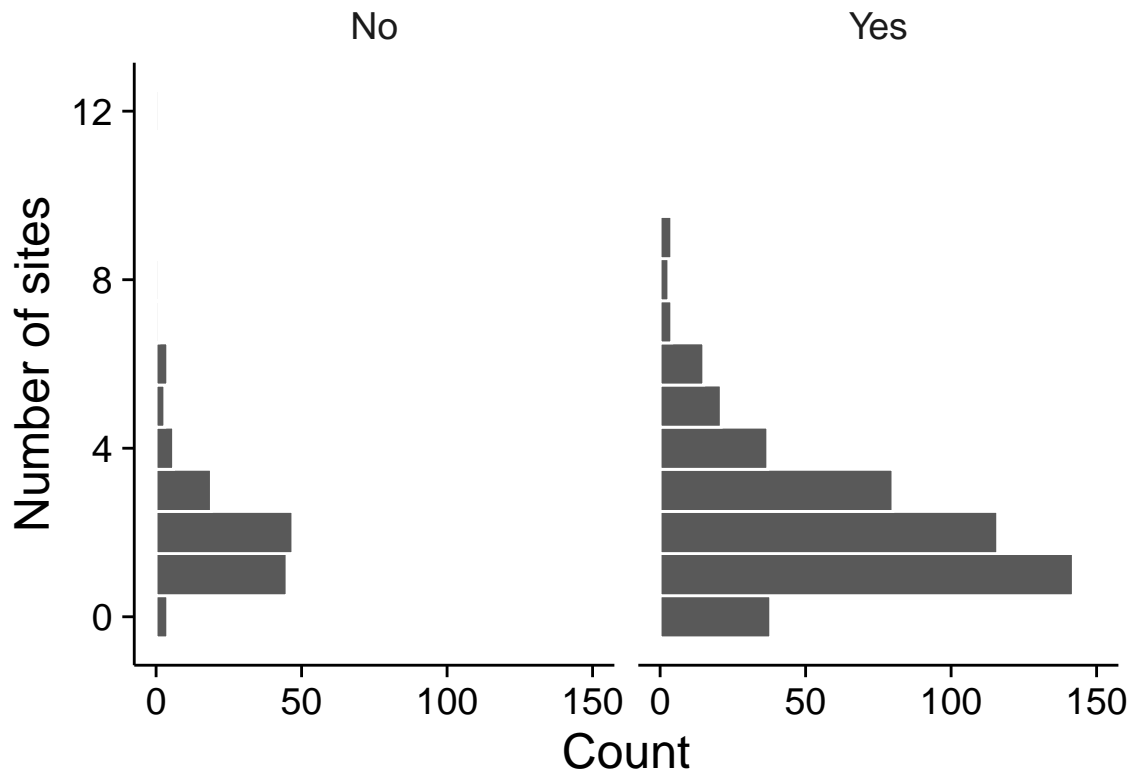
```

analysis_set %>%
  select(Count, ART_currently) %>%
  filter(complete.cases()) %>%
  ggplot(data = .) +
  aes(Count) +
  geom_histogram(binwidth = 1,
                 colour = '#FFFFFF') +
  scale_y_continuous(limits = c(0, 150),
                     breaks = c(0, 50, 100, 150)) +
  labs(title = 'Currently on HAART',
       subtitle = str_glue('(Complete cases = {nrow(analysis_set[!is.na(
         analysis_set$ART_currently), ]})}')),
       x = 'Number of sites',
       y = 'Count') +
  facet_wrap(~ART_currently) +
  coord_flip() +
  theme_minimal(base_size = 18) +
  theme(plot.title = element_text(size = 18),
        plot.subtitle = element_text(size = 12),
        panel.grid = element_blank(),
        axis.text = element_text(colour = '#000000'),
        axis.line = element_line(size = 0.5),
        axis.ticks = element_line(size = 0.5))

```

Currently on HAART

(Complete cases = 591)



```
haart <- analysis_set %>%
  select(Count, ART_currently) %>%
  filter(complete.cases(.)) %>%
  ggplot(data = .) +
  aes(y = Count,
      x = ART_currently) +
  geom_boxplot(outlier.colour = '#FFFFFF',
               outlier.size = 0) +
  geom_point(size = 2,
             shape = 21,
             position = position_jitter(height = 0.2, width = 0.2),
             fill = '#000000',
             colour = '#000000',
             alpha = 0.5) +
  scale_x_discrete(labels = c('No', 'Yes')) +
  labs(title = 'Currently on HAART',
       subtitle = str_glue('(Complete cases = {nrow(analysis_set[!is.na(
         analysis_set$ART_currently), ]})}'),
       y = 'Number of sites') +
  theme_minimal(base_size = 18) +
  theme(plot.title = element_text(size = 18),
        plot.subtitle = element_text(size = 12),
        plot.caption = element_text(size = 12),
        panel.grid = element_blank(),
        axis.title.x = element_blank(),
```

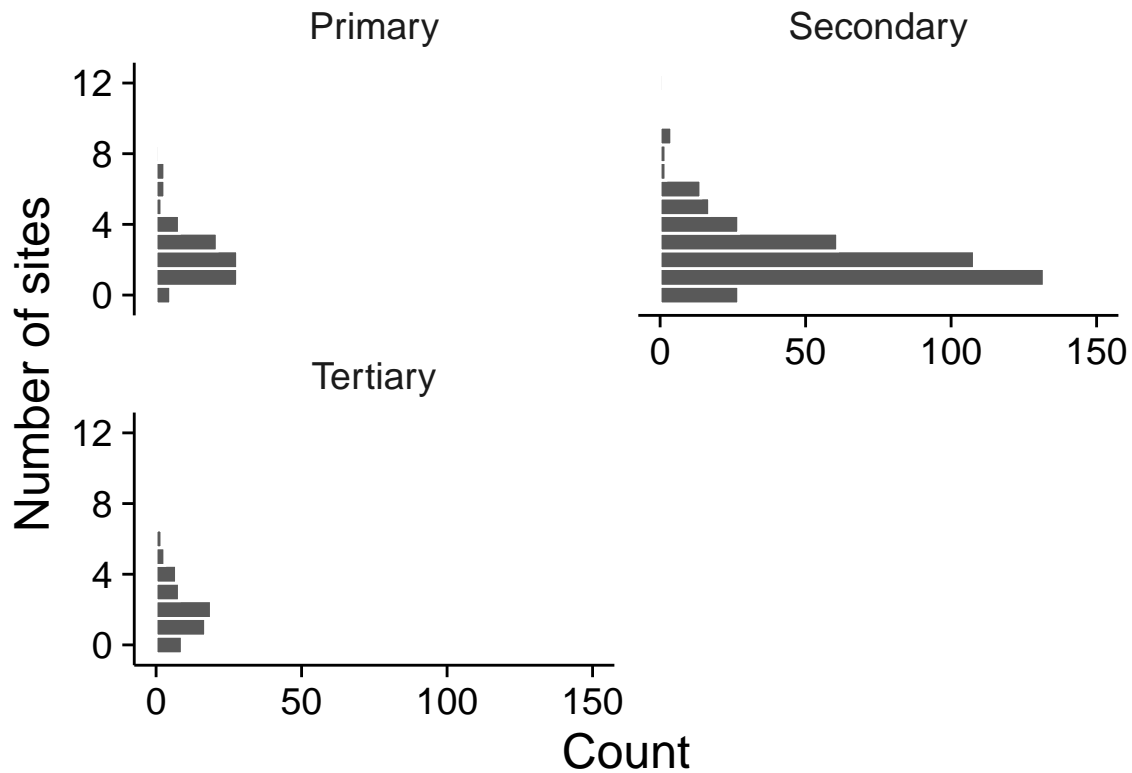
```
axis.text = element_text(colour = '#000000'),
axis.line = element_line(size = 0.5),
axis.ticks = element_line(size = 0.5))
```

2.3.1.4 Count education

```
analysis_set %>%
  select(Count, Education) %>%
  filter(complete.cases(.)) %>%
  ggplot(data = .) +
  aes(Count) +
  geom_histogram(binwidth = 1,
                 colour = '#FFFFFF') +
  scale_y_continuous(limits = c(0, 150),
                     breaks = c(0, 50, 100, 150)) +
  labs(title = 'Level of education',
       subtitle = str_glue('(Complete cases = {nrow(analysis_set[!is.na(
                           analysis_set$Education), ]})}')),
       x = 'Number of sites',
       y = 'Count') +
  facet_wrap(~Education, ncol = 2) +
  coord_flip() +
  theme_minimal(base_size = 18) +
  theme(plot.title = element_text(size = 18),
        plot.subtitle = element_text(size = 12),
        panel.grid = element_blank(),
        axis.text = element_text(colour = '#000000'),
        axis.line = element_line(size = 0.5),
        axis.ticks = element_line(size = 0.5))
```

Level of education

(Complete cases = 559)



```
edu <- analysis_set %>%
  select(Count, Education) %>%
  filter(complete.cases(.)) %>%
  ggplot(data = .) +
  aes(y = Count,
      x = Education) +
  geom_boxplot(outlier.colour = '#FFFFFF',
               outlier.size = 0) +
  geom_point(size = 2,
             shape = 21,
             position = position_jitter(height = 0.2, width = 0.2),
             fill = '#000000',
             colour = '#000000',
             alpha = 0.5) +
  scale_x_discrete(labels = c('Primary', 'Secondary',
                              'Tertiary')) +
  labs(title = 'Level of Education',
       subtitle = str_glue('(Complete cases = {nrow(analysis_set[!is.na(
         analysis_set$Education), ]})}')),
       y = 'Number of sites') +
  theme_minimal(base_size = 18) +
  theme(plot.title = element_text(size = 18),
        plot.subtitle = element_text(size = 12),
        plot.caption = element_text(size = 12),
        panel.grid = element_blank(),
```

```
axis.title.x = element_blank(),
axis.text = element_text(colour = '#000000'),
axis.line = element_line(size = 0.5),
axis.ticks = element_line(size = 0.5))
```

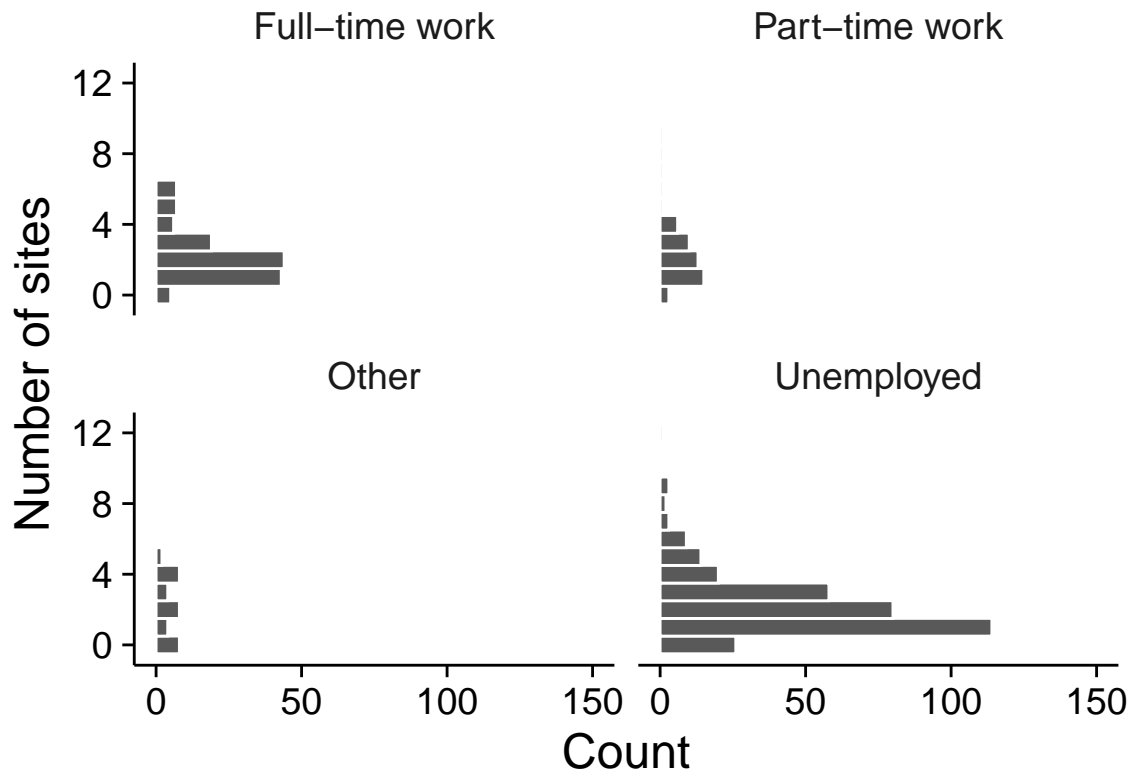
2.3.1.5 Count by employment status

```
analysis_set %>%
  select(Count, Employment_status) %>%
  filter(complete.cases(.)) %>%
  mutate(Employment_status = factor(Employment_status,
                                     levels = c('Full-time work',
                                                  'Part-time work',
                                                  'Other',
                                                  'Unemployed'),
                                     ordered = TRUE)) %>%

  ggplot(data = .) +
  aes(Count) +
  geom_histogram(binwidth = 1,
                 colour = '#FFFFFF') +
  scale_y_continuous(limits = c(0, 150),
                     breaks = c(0, 50, 100, 150)) +
  labs(title = 'Employment status',
       subtitle = str_glue('(Complete cases = {nrow(analysis_set[!is.na(
                           analysis_set$Employment_status), ]})}')),
       x = 'Number of sites',
       y = 'Count') +
  facet_wrap(~Employment_status, ncol = 2) +
  coord_flip() +
  theme_minimal(base_size = 18) +
  theme(plot.title = element_text(size = 18),
        plot.subtitle = element_text(size = 12),
        panel.grid = element_blank(),
        axis.text = element_text(colour = '#000000'),
        axis.line = element_line(size = 0.5),
        axis.ticks = element_line(size = 0.5))
```


Employment status

(Complete cases = 547)



```
employment <- analysis_set %>%
  select(Count, Employment_status) %>%
  filter(complete.cases(.)) %>%
  mutate(Employment_status = factor(Employment_status,
    levels = c('Full-time work',
               'Part-time work',
               'Other',
               'Unemployed'),
    ordered = TRUE)) %>%

  ggplot(data = .) +
  aes(y = Count,
      x = Employment_status) +
  geom_boxplot(outlier.colour = '#FFFFFF',
               outlier.size = 0) +
  geom_point(size = 2,
             shape = 21,
             position = position_jitter(height = 0.2, width = 0.2),
             fill = '#000000',
             colour = '#000000',
             alpha = 0.5) +
  scale_x_discrete(labels = c('Full-time\nwork', 'Part-time\nwork',
                              'Other', 'Unemployed')) +
  labs(title = 'Employment status',
       subtitle = str_glue('(Complete cases = {nrow(analysis_set[!is.na(
         analysis_set$Employment_status), ]})}'))
```

```

caption = 'Other: receiving social grants or were students',
y = 'Number of sites') +
theme_minimal(base_size = 18) +
theme(plot.title = element_text(size = 18),
      plot.subtitle = element_text(size = 12),
      plot.caption = element_text(size = 12),
      panel.grid = element_blank(),
      axis.title.x = element_blank(),
      axis.text = element_text(colour = '#000000'),
      axis.line = element_line(size = 0.5),
      axis.ticks = element_line(size = 0.5))

```

2.3.1.6 Count by pain definition

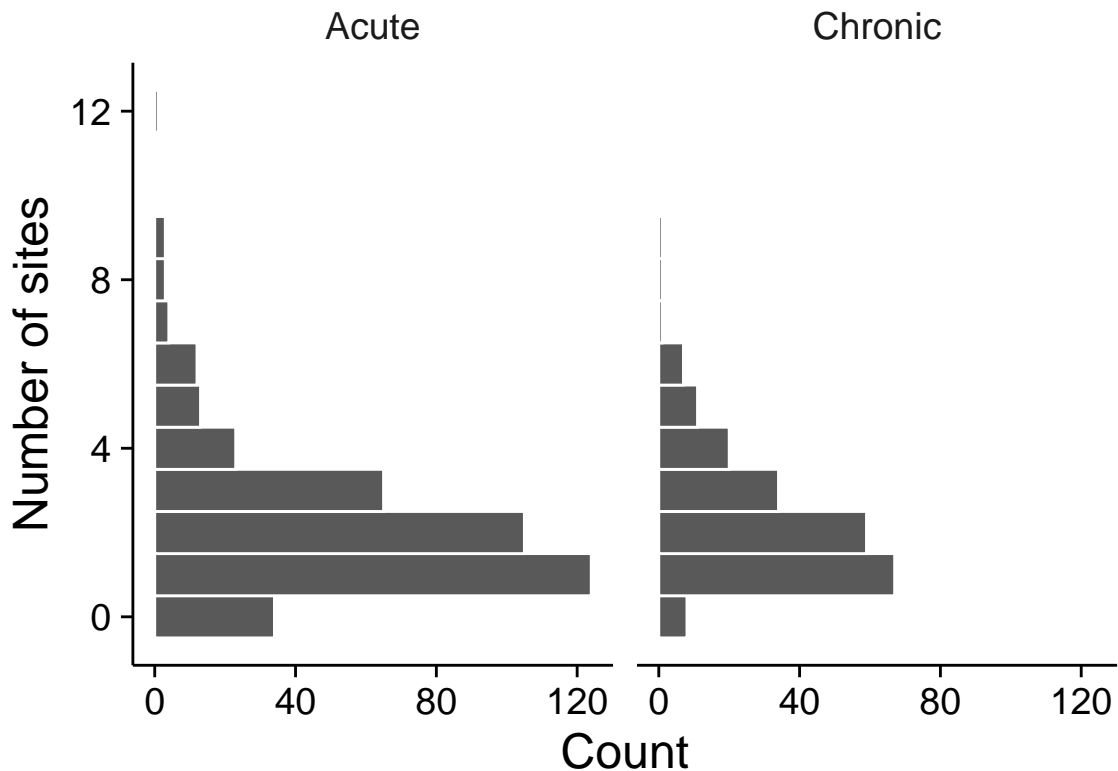
```

analysis_set %>%
  select(Count, Pain_def) %>%
  filter(complete.cases(.)) %>%
  ggplot(data = .) +
  aes(Count) +
  geom_histogram(binwidth = 1,
                 colour = '#FFFFFF') +
  labs(title = 'Pain definition',
       subtitle = str_glue('(Complete cases = {nrow(analysis_set[!is.na(
                           analysis_set$Pain_def), ]})}')),
       x = 'Number of sites',
       y = 'Count') +
  facet_wrap(~Pain_def) +
  coord_flip() +
  theme_minimal(base_size = 18) +
  theme(plot.title = element_text(size = 18),
        plot.subtitle = element_text(size = 12),
        panel.grid = element_blank(),
        axis.text = element_text(colour = '#000000'),
        axis.line = element_line(size = 0.5),
        axis.ticks = element_line(size = 0.5))

```

Pain definition

(Complete cases = 596)



```
pain_def <- analysis_set %>%
  select(Count, Pain_def) %>%
  filter(complete.cases(.)) %>%
  ggplot(data = .) +
  aes(y = Count,
      x = Pain_def) +
  geom_boxplot(outlier.colour = '#FFFFFF',
               outlier.size = 0) +
  geom_point(size = 2,
             shape = 21,
             position = position_jitter(height = 0.2, width = 0.2),
             fill = '#000000',
             colour = '#000000',
             alpha = 0.5) +
  scale_x_discrete(labels = c('Acute', 'Chronic')) +
  labs(title = 'Pain definition',
       subtitle = str_glue('(Complete cases = {nrow(analysis_set[!is.na(
         analysis_set$Pain_def), ]})}')),
       y = 'Number of sites') +
  theme_minimal(base_size = 18) +
  theme(plot.title = element_text(size = 18),
        plot.subtitle = element_text(size = 12),
        plot.caption = element_text(size = 12),
        panel.grid = element_blank(),
        axis.title.x = element_blank(),
```

```
axis.text = element_text(colour = '#000000'),
axis.line = element_line(size = 0.5),
axis.ticks = element_line(size = 0.5))
```

2.4 Regression analysis

2.4.1 Process data

```
analysis_set <- analysis_set %>%
  # Re-factor Employment_status
  mutate(Employment_status = case_when(
    Employment_status == 'Unemployed' ~ ' 1_unemployed',
    Employment_status == 'Part-time work' ~ ' 3_employed PT',
    Employment_status == 'Full-time work' ~ ' 2_employed FT',
    Employment_status == 'Other' ~ ' 4_other'
  )) %>%
  filter(complete.cases(.))
```

2.4.2 Mixed model negative binomial regression for count data

```
# Null model
null <- glmer.nb(Count ~ 1 +
  (1|Site),
  data = analysis_set)

# Full model
mod.mmnbn <- glmer.nb(Count ~ Sex +
  # Center and scale age
  scale(Age) +
  # Center and scale CD4
  scale(CD4_recent) +
  Employment_status +
  Education +
  ART_currently +
  Pain_def +
  (1|Site),
  data = analysis_set)

## Fit is singular, print model to check SD of random effect
summary(mod.mmnbn)

## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: Negative Binomial(9.4933) ( log )
## Formula: Count ~ Sex + scale(Age) + scale(CD4_recent) + Employment_status +
## Education + ART_currently + Pain_def + (1 | Site)
## Data: analysis_set
##
##      AIC      BIC    logLik deviance df.resid
## 1608.4   1661.2   -791.2   1582.4     417
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -1.5293 -0.7298 -0.1483 0.4427 6.0224
##
## Random effects:
## Groups Name Variance Std.Dev.
## Site (Intercept) 4.952e-12 2.225e-06
## Number of obs: 430, groups: Site, 5
##
## Fixed effects:
## Estimate Std. Error z value Pr(>|z|)
## (Intercept) 0.86675 0.11990 7.229 4.88e-13 ***
## SexMale -0.11757 0.10083 -1.166 0.2436
## scale(Age) 0.06919 0.04101 1.687 0.0916 .
## scale(CD4_recent) 0.01573 0.03825 0.411 0.6809
## Employment_status 2_employed FT -0.01663 0.09150 -0.182 0.8558
## Employment_status 3_employed PT 0.18433 0.12035 1.532 0.1256
## Employment_status 4_other 0.02330 0.14169 0.164 0.8694
## EducationSecondary -0.05979 0.09809 -0.610 0.5422
## EducationTertiary -0.20110 0.14198 -1.416 0.1567
## ART_currentlyYes 0.01669 0.09858 0.169 0.8656
## Pain_defChronic 0.08157 0.09758 0.836 0.4032
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
## (Intr) SexMal scl(A) s(CD4_ E_2_FT E_3_PT Emp_4_ EdctnS EdctnT
## SexMale -0.141
## scale(Age) -0.051 -0.149
## scl(CD4_rc) 0.023 0.191 -0.033
## Emplm_2_FT -0.158 -0.009 -0.087 -0.035
## Emplm_3_PT -0.178 0.001 -0.003 -0.018 0.224
## Emplmnt_4_ -0.079 0.038 0.049 -0.047 0.148 0.103
## EdctnScndry -0.646 0.038 0.221 0.047 -0.088 0.050 -0.070
## EductnTrtry -0.399 0.061 0.220 0.016 -0.148 0.032 -0.132 0.586
## ART_crrntly -0.619 0.029 0.019 0.032 0.105 0.036 -0.027 0.013 -0.017
## Pan_dfChrnc 0.029 -0.162 -0.377 -0.334 -0.165 -0.108 0.092 -0.063 -0.153
## ART_cY
## SexMale
## scale(Age)
## scl(CD4_rc)
## Emplm_2_FT
## Emplm_3_PT
## Emplmnt_4_
## EdctnScndry
## EductnTrtry
## ART_crrntly
## Pan_dfChrnc -0.297
## convergence code: 0
## boundary (singular) fit: see ?isSingular
```

2.4.3 Negative binomial regression for count data

```
# Null model
null <- glm.nb(Count ~ 1,
               data = analysis_set)
```

```

# Full model
mod.nb <- glm.nb(Count ~ Sex +
                  # Center and scale age
                  scale(Age) +
                  # Center and scale CD4
                  scale(CD4_recent) +
                  Employment_status +
                  Education +
                  ART_currently +
                  Pain_def,
                  data = analysis_set)

# Compare models
anova(null, mod.nb)

## Likelihood ratio tests of Negative Binomial Models
##
## Response: Count
##
## 1
## 2 Sex + scale(Age) + scale(CD4_recent) + Employment_status + Education + ART_currently + Pain_def
##      theta Resid. df      2 x log-lik.   Test      df LR stat.   Pr(Chi)
## 1 8.192033      429      -1596.185
## 2 9.493334      419      -1582.363 1 vs 2      10 13.82139 0.1812956

# Summary of coefficients
cbind(Estimate = coef(mod.nb)[-1],
      confint(mod.nb)[-1, ],
      summary(mod.nb)$coefficients[-1, 3:4]) %>%
  kable(caption = 'Coefficients and 95% CI',
        digits = 3,
        col.names = c('Estimate', 'Lower 95%CI', 'Upper 95%CI',
                      'z-value', 'P-value'))

```

Table 20: Coefficients and 95% CI

	Estimate	Lower 95%CI	Upper 95%CI	z-value	P-value
SexMale	-0.118	-0.317	0.078	-1.166	0.243
scale(Age)	0.069	-0.012	0.149	1.692	0.091
scale(CD4_recent)	0.016	-0.060	0.090	0.413	0.680
Employment_status 2_employed FT	-0.017	-0.197	0.162	-0.181	0.856
Employment_status 3_employed PT	0.184	-0.055	0.417	1.528	0.127
Employment_status 4_other	0.023	-0.260	0.296	0.164	0.869
EducationSecondary	-0.060	-0.250	0.134	-0.609	0.542
EducationTertiary	-0.201	-0.481	0.076	-1.418	0.156
ART_currentlyYes	0.017	-0.175	0.212	0.169	0.866
Pain_defChronic	0.082	-0.110	0.272	0.842	0.400

2.5 Publication plot

```

# Generate plots for age and CD4
## Count by age

```

```

age <- analysis_set %>%
  select(Count, Age) %>%
  filter(complete.cases()) %>%
  ggplot(data = .) +
  aes(y = Count,
       x = Age) +
  geom_point(size = 2,
             shape = 21,
             position = position_jitter(height = 0.2, width = 0.2),
             fill = '#000000',
             colour = '#000000',
             alpha = 0.5) +
  labs(title = 'Age',
       subtitle = str_glue('(Complete cases = {nrow(analysis_set[!is.na(
           analysis_set$Age), ]})}')),
       x = 'Age (years)',
       y = 'Number of sites') +
  scale_x_continuous(limits = c(18, 80),
                    breaks = seq(20, 80, by = 20)) +
  theme_minimal(base_size = 18) +
  theme(plot.title = element_text(size = 18),
        plot.subtitle = element_text(size = 12),
        plot.caption = element_text(size = 12),
        panel.grid = element_blank(),
        axis.text = element_text(colour = '#000000'),
        axis.line = element_line(size = 0.5),
        axis.ticks = element_line(size = 0.5))

## Count by CD4 T-cell count
cd4 <- analysis_set %>%
  select(Count, CD4_recent) %>%
  filter(complete.cases()) %>%
  ggplot(data = .) +
  aes(y = Count,
       x = CD4_recent) +
  geom_point(size = 2,
             shape = 21,
             position = position_jitter(height = 0.2, width = 0.2),
             fill = '#000000',
             colour = '#000000',
             alpha = 0.5) +
  labs(title = 'Recent CD4 T-cell count',
       subtitle = str_glue('(Complete cases = {nrow(analysis_set[!is.na(
           analysis_set$CD4_recent), ]})}')),
       x = expression('CD4 T-cell count (cells.mm-3)'),
       y = 'Number of sites') +
  scale_x_continuous(limits = c(0, 1400),
                    breaks = seq(0, 1400, by = 400)) +
  scale_y_continuous(breaks = seq(0, 12, 4)) +
  theme_minimal(base_size = 18) +
  theme(plot.title = element_text(size = 18),
        plot.subtitle = element_text(size = 12),
        plot.caption = element_text(size = 12),

```

```

    panel.grid = element_blank(),
    axis.text = element_text(colour = '#000000'),
    axis.line = element_line(size = 0.5),
    axis.ticks = element_line(size = 0.5))

# Fix figures for patchwork plot
all2 <- all

sex2 <- sex +
  theme(axis.title.y = element_blank())

haart2 <- haart

employment2 <- employment

edu2 <- edu +
  theme(axis.title.y = element_blank())

pain_def2 <- pain_def +
  theme(axis.title.y = element_blank())

age2 <- age

cd42 <- cd4 +
  theme(axis.title.y = element_blank())

pubs <- all2 + sex2 + haart2 + edu2 + employment2 + pain_def2 + age2 + cd42 +
  plot_layout(ncol = 2)

ggsave(filename = 'figures/figure_3.png',
        width = 10,
        height = 18)

```


3 Session information

```
sessionInfo()
```

```
## R version 3.6.3 (2020-02-29)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Debian GNU/Linux 10 (buster)
##
## Matrix products: default
## BLAS/LAPACK: /usr/lib/x86_64-linux-gnu/libopenblas-r0.3.5.so
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8      LC_NUMERIC=C
##  [3] LC_TIME=en_US.UTF-8      LC_COLLATE=en_US.UTF-8
##  [5] LC_MONETARY=en_US.UTF-8  LC_MESSAGES=C
##  [7] LC_PAPER=en_US.UTF-8     LC_NAME=C
##  [9] LC_ADDRESS=C             LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods    base
##
## other attached packages:
##  [1] patchwork_1.0.0 knitr_1.28      skimr_2.1.1     forcats_0.5.0
##  [5] stringr_1.4.0   dplyr_0.8.5     purrr_0.3.4     readr_1.3.1
##  [9] tidyr_1.0.2     tibble_3.0.1    ggplot2_3.3.0   tidyverse_1.3.0
## [13] MASS_7.3-51.5   lme4_1.1-23     Matrix_1.2-18
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_1.0.4.6    lubridate_1.7.8 lattice_0.20-38 utf8_1.1.4
##  [5] assertthat_0.2.1 digest_0.6.25   R6_2.4.1        cellranger_1.1.0
##  [9] repr_1.1.0      backports_1.1.6 reprex_0.3.0     evaluate_0.14
## [13] highr_0.8       httr_1.4.1     pillar_1.4.3    rlang_0.4.5
## [17] readxl_1.3.1    minqa_1.2.4    rstudioapi_0.11 nloptr_1.2.2.1
## [21] rmarkdown_2.1   labeling_0.3    splines_3.6.3   statmod_1.4.34
## [25] munsell_0.5.0   broom_0.5.6     compiler_3.6.3  modelr_0.1.6
## [29] xfun_0.13       base64enc_0.1-3 pkgconfig_2.0.3 htmltools_0.4.0
## [33] tidyselect_1.0.0 fansi_0.4.1     crayon_1.3.4    dbplyr_1.4.3
## [37] withr_2.2.0     grid_3.6.3     nlme_3.1-144    jsonlite_1.6.1
## [41] gtable_0.3.0    lifecycle_0.2.0 DBI_1.1.0        magrittr_1.5
## [45] scales_1.1.0    cli_2.0.2       stringi_1.4.6   farver_2.0.3
## [49] fs_1.4.1        xml2_1.3.2      ellipsis_0.3.0  generics_0.0.2
## [53] vctr_0.2.4      boot_1.3-24     tools_3.6.3     glue_1.4.0
## [57] hms_0.5.3       yaml_2.2.1      colorspace_1.4-1 rvest_0.3.5
## [61] haven_2.2.0
```

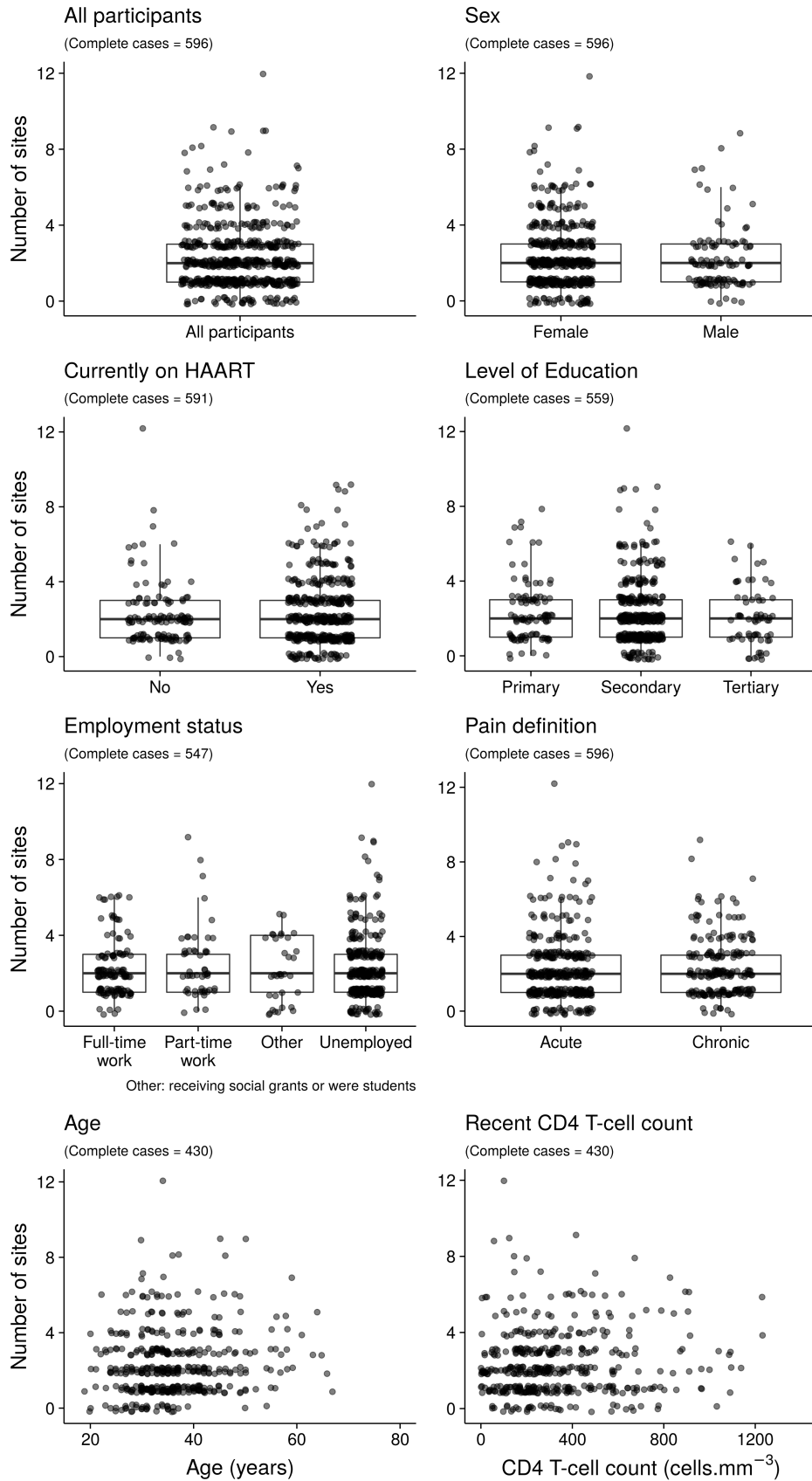


Figure 1: Number of pain sites by predictor variable