

Supplement 1

Summary statistics for baseline variables

Peter Kamerman and Tory Madden

Last updated: 07 May 2019

Contents

Import data	1
Clean data	2
Demographic data	2
Brief Pain Inventory (BPI)	3
Beck's Depression Inventory (BDI)	3
EQ-5D (3L)	3
Self-efficacy Questionnaire 6 (SE6)	4
Analysis	5
Demographic data	5
Continuous variables	5
Factor variables	6
Brief Pain Inventory	7
Continuous variables	7
Beck's Depression Index	8
Continuous variables	8
EQ5D (3L)	9
Continuous variables	9
Self-efficacy Questionnaire 6	10
Continuous variables	10
Session information	11

Tabular summary statistics of baseline variables (time = 0 weeks) for the whole cohort, and stratified by sex and study site.

Import data

```
# Demographic data
demo <- read_rds('data-cleaned/demographics.rds')

# BDI
bdi <- read_rds('data-cleaned/bdi.rds')

# BPI
bpi <- read_rds('data-cleaned/bpi.rds')

# EQ5D
eq5d <- read_rds('data-cleaned/eq5d.rds')

# SE6
se6 <- read_rds('data-cleaned/se6.rds')
```

Clean data

Demographic data

```
demo %<>%  
  # Convert sex to factor  
  mutate(Sex = factor(Sex)) %>%  
  # Rename Years_on_ART  
  rename(Years_on_HAART = Years_on_ART) %>%  
  # Convert SOS_mnemonic to factor  
  mutate(SOS_mnemonic = factor(SOS_mnemonic)) %>%  
  # Transfer CD4_nadir data to CD4_recent if missing CD4_recent data  
  # (i.e. get the most updated CD4 count available)  
  mutate(CD4_most_recent =  
    ifelse(is.na(CD4_recent),  
           yes = CD4_nadir,  
           no = CD4_recent)) %>%  
  # Categorise years of schooling into 7 years or less, 8-12 years,  
  # and more than 12 years of education, factorize, and then order  
  mutate(Education = case_when(  
    Years_education <= 7 ~ '0-7 years',  
    Years_education > 7 & Years_education <= 12 ~ '8-12 years',  
    Years_education > 12 ~ 'More than 12 years'),  
    Education = factor(Education,  
                       levels = c('0-7 years',  
                                   '8-12 years',  
                                   'More than 12 years'),  
                       ordered = TRUE)) %>%  
  # Recode HAART and order  
  mutate(HAART = case_when(  
    HAART == 'first-line' ~ 'first-line HAART',  
    HAART == 'second-line' ~ 'second-line HAART',  
    HAART == 'monitoring' ~ 'no HAART'),  
    HAART = factor(HAART,  
                   levels = c('no HAART',  
                               'first-line HAART',  
                               'second-line HAART'),  
                   ordered = TRUE)) %>%  
  # Recode and order occupation categories  
  mutate(Employment = str_replace_all(Occupation,  
                                       pattern = '^unemployed - .+',  
                                       replacement = 'unemployed'),  
    Employment = factor(Employment,  
                        levels = c('employed', 'unemployed',  
                                    'student/volunteer',  
                                    'unable to work - disability grant')) %>%  
  # Select required columns  
  select(ID, Study_site, Sex, Age_years, Years_on_HAART,  
         CD4_most_recent, HAART, Education, Employment,  
         SOS_mnemonic)  
  
# Make a site/sex filter
```

```
sorter <- demo %>%
  select(ID, Study_site, Sex)
```

Brief Pain Inventory (BPI)

```
bpi %<>%
  # Capitalize IDs
  mutate(ID = stringr::str_to_upper(ID)) %>%
  # Select baseline values
  select(ID,
    ends_with('BL')) %>%
  # Select columns
  select(ID, 3:6, 9:15) %>%
  # Calculate Pain Severity Score (PSS) at baseline
  mutate(PSS = rowMeans(.[2:5], na.rm = TRUE),
    # Treat PSS as a discrete scale
    PSS = round(PSS)) %>%
  # Calculate Pain Interference Index (PIS) at baseline
  mutate(PIS = rowMeans(.[6:12], na.rm = TRUE),
    # Treat PIS as a discrete scale
    PIS = round(PIS)) %>%
  #remove unwanted columns
  select(ID, PSS, PIS) %>%
  left_join(sorter)
```

Beck's Depression Inventory (BDI)

```
bdi %<>%
  # Capitalize IDs
  mutate(ID = stringr::str_to_upper(ID)) %>%
  # Make a total score column
  mutate_at(2:ncol(bdi),
    as.numeric) %>%
  mutate(BDI = rowSums(.[2:ncol(bdi)], na.rm = TRUE),
    # Treat BDI as a discrete scale
    BDI = round(BDI)) %>%
  select(ID, BDI) %>%
  left_join(sorter)
```

EQ-5D (3L)

```
eq5d %<>%
  # Capitalize IDs
  mutate(ID = stringr::str_to_upper(ID))

# Calculate eq5d index score
## Create basic term = 1 for all cases in new column
eq5d$index_core <- 1

## Sum all rows for total index score
eq5d %<>%
  mutate(index_sum = rowSums(.[2:6], na.rm = TRUE))

# Create constant term to subtract for domain scores > 1 (i.e. sum > 5)
```

```

eq5d %<>%
  mutate(index_constant = ifelse(index_sum > 5,
                                yes = 0.081,
                                no = 0))

## Create variable for subtraction for each domain
eq5d %<>%
  mutate(Mobility_index = ifelse(Mobility.BL == 2,
                                yes = 0.069,
                                no = ifelse(Mobility.BL == 3,
                                              yes = 0.314,
                                              no = 0))) %>%
  mutate(Self_care_index = ifelse(Self_care.BL == 2,
                                yes = 0.104,
                                no = ifelse(Self_care.BL == 3,
                                              yes = 0.214,
                                              no = 0))) %>%
  mutate(Usual_activities_index = ifelse(Usual_activities.BL == 2,
                                         yes = 0.036,
                                         no = ifelse(Usual_activities.BL == 3,
                                                       yes = 0.094,
                                                       no = 0))) %>%
  mutate(Pain_index = ifelse(Pain.BL == 2,
                             yes = 0.123,
                             no = ifelse(Pain.BL == 3,
                                           yes = 0.386,
                                           no = 0))) %>%
  mutate(Anxiety_depression_index = ifelse(Anxiety_and_depression.BL == 2,
                                           yes = 0.071,
                                           no = ifelse(Anxiety_and_depression.BL == 3,
                                                         yes = 0.236,
                                                         no = 0)))

## Compute the index score using:
## index = index_core - constant_index - Mobility_index...
eq5d %<>%
  mutate(EQ5D_index = index_core - index_constant - Mobility_index
         - Self_care_index - Usual_activities_index - Pain_index
         - Anxiety_depression_index) %>%
  # Convert State_of_health VAS to double
  mutate(EQ5D_VAS = as.numeric(State_of_health.BL))

# Select columns
eq5d %<>% select(ID,
               EQ5D_index,
               EQ5D_VAS) %>%
  left_join(sorter)

```

Self-efficacy Questionnaire 6 (SE6)

```

se6 %<>%
  # Capitalize IDs
  mutate(ID = stringr::str_to_upper(ID)) %>%
  # Calculate SE6 at baseline
  mutate(SE6 = rowMeans(.[2:7], na.rm = TRUE),
         # Treat SE6 as a discrete scale

```

```
SE6 = round(SE6)) %>%
#remove unwanted columns
select(ID,
       SE6) %>%
left_join(sorter)
```

Analysis

Demographic data

Continuous variables

```
demo %>%
  select_if(is.numeric) %>%
  skim_to_wide(.) %>%
  kable(., caption = 'Whole cohort',
        align = 'llrrrrrrrrrr')
```

Table 1: Whole cohort

type	variable	missing	complete	n	mean	sd	median	q25	q75	min	max
numeric	Age_years	0	160	160	35.23	5.65	35	32	38	18	58
numeric	CD4_most_recent	8	152	160	406.45	249.51	376	224.75	547	3	1189
numeric	Years_on_HAART	78	82	160	3.56	2.83	3	1	5.06	0.25	13

```
demo %>%
  select(Study_site, Age_years, Years_on_HAART, CD4_most_recent) %>%
  group_by(Study_site) %>%
  skim_to_wide(.) %>%
  kable(., caption = 'By study site',
        align = 'lllrrrrrrrrrr')
```

Table 2: By study site

type	Study_site	variable	missing	complete	n	mean	sd	median	q25	q75	min	max
numeric	R1	Age_years	0	47	47	35.28	2.99	36	33	38	28	40
numeric	R1	CD4_most_recent	0	47	47	415.43	195.84	397	277.5	538	114	1180
numeric	R1	Years_on_HAART	29	18	47	3.92	2.05	4.17	2.31	5.46	0.67	8.3
numeric	R2	Age_years	0	49	49	32.9	4.63	35	30	36	18	40
numeric	R2	CD4_most_recent	2	47	49	450.53	241.69	407	268.5	562.5	36	1120
numeric	R2	Years_on_HAART	2	47	49	3.97	3.25	3	1	6	0.25	13
numeric	U1	Age_years	0	47	47	39.34	6.27	38	35	43.5	27	58
numeric	U1	CD4_most_recent	6	41	47	302.73	284.63	206	113	368	3	1189
numeric	U1	Years_on_HAART	47	0	47	NaN	NA	NA	NA	NA	Inf	-Inf
numeric	U2	Age_years	0	17	17	30.41	4.77	30	26	34	23	37
numeric	U2	CD4_most_recent	0	17	17	509.94	248.8	471	414	648	119	1097
numeric	U2	Years_on_HAART	0	17	17	2.07	1.69	1	0.67	3.42	0.33	5.25

```
demo %>%
  select(Sex, Age_years, Years_on_HAART, CD4_most_recent) %>%
  group_by(Sex) %>%
  skim_to_wide(.) %>%
```

```
kable(., caption = 'By sex',
      align = 'lllrrrrrrrrr')
```

Table 3: By sex

type	Sex	variable	missing	complete	n	mean	sd	median	q25	q75	min	max
numeric	female	Age_years	0	97	97	34.23	5.97	35	30	37	18	58
numeric	female	CD4_most_recent	5	92	97	433.73	273.51	410.5	244.75	570.5	3	1189
numeric	female	Years_on_HAART	33	64	97	3.46	3.03	3	0.96	5	0.25	13
numeric	male	Age_years	0	63	63	36.76	4.76	36	33.5	39	27	50
numeric	male	CD4_most_recent	3	60	63	364.63	202.5	335	210.5	491.75	30	1180
numeric	male	Years_on_HAART	45	18	63	3.92	2.05	4.17	2.31	5.46	0.67	8.3

Factor variables

```
demo %>%
  select_if(is.factor) %>%
  skim_to_wide(.) %>%
  kable(., caption = 'Whole cohort',
        align = 'llrrrrr')
```

Table 4: Whole cohort

type	variable	missing	complete	n	n_unique	top_counts
factor	Education	2	158	160	3	8-1: 112, 0-7: 44, Mor: 2
factor	Employment	3	157	160	4	une: 96, emp: 51, una: 8, stu: 2
factor	HAART	4	156	160	3	fir: 115, sec: 36, no : 5
factor	Sex	0	160	160	2	fem: 97, mal: 63
factor	SOS_mnemonic	47	113	160	2	low: 78, hea: 35

```
demo %>%
  select(Study_site, Education, Employment, HAART, Sex, SOS_mnemonic) %>%
  group_by(Study_site) %>%
  skim_to_wide(.) %>%
  kable(., caption = 'By study site',
        align = 'lllrrrrrr')
```

Table 5: By study site

type	Study_site	variable	missing	complete	n	n_unique	top_counts
factor	R1	Education	0	47	47	2	8-1: 30, 0-7: 17, Mor: 0
factor	R1	Employment	2	45	47	3	emp: 21, une: 20, una: 4, stu: 0
factor	R1	HAART	0	47	47	2	fir: 43, sec: 4, no : 0
factor	R1	Sex	0	47	47	1	mal: 47, fem: 0
factor	R1	SOS_mnemonic	0	47	47	2	hea: 28, low: 19
factor	R2	Education	0	49	49	2	8-1: 30, 0-7: 19, Mor: 0
factor	R2	Employment	0	49	49	3	une: 40, emp: 8, una: 1, stu: 0
factor	R2	HAART	0	49	49	3	fir: 40, sec: 5, no : 4
factor	R2	Sex	0	49	49	1	fem: 49, mal: 0
factor	R2	SOS_mnemonic	0	49	49	1	low: 49, hea: 0
factor	U1	Education	2	45	47	3	8-1: 41, 0-7: 2, Mor: 2
factor	U1	Employment	1	46	47	4	une: 25, emp: 19, stu: 1, una: 1
factor	U1	HAART	4	43	47	3	sec: 24, fir: 18, no : 1
factor	U1	Sex	0	47	47	2	fem: 31, mal: 16

type	Study_site	variable	missing	complete	n	n_unique	top_counts
factor	U1	SOS_mnemonic	47	0	47	0	hea: 0, low: 0
factor	U2	Education	0	17	17	2	8-1: 11, 0-7: 6, Mor: 0
factor	U2	Employment	0	17	17	4	une: 11, emp: 3, una: 2, stu: 1
factor	U2	HAART	0	17	17	2	fir: 14, sec: 3, no : 0
factor	U2	Sex	0	17	17	1	fem: 17, mal: 0
factor	U2	SOS_mnemonic	0	17	17	2	low: 10, hea: 7

```
demo %>%
  select(Sex, Education, Employment, HAART, Sex, SOS_mnemonic) %>%
  group_by(Sex) %>%
  skim_to_wide(.) %>%
  kable(., caption = 'By sex',
        align = 'lllrrrrr')
```

Table 6: By sex

type	Sex	variable	missing	complete	n	n_unique	top_counts
factor	female	Education	1	96	97	3	8-1: 68, 0-7: 26, Mor: 2
factor	female	Employment	1	96	97	4	une: 65, emp: 26, una: 3, stu: 2
factor	female	HAART	3	94	97	3	fir: 65, sec: 24, no : 5
factor	female	SOS_mnemonic	31	66	97	2	low: 59, hea: 7
factor	male	Education	1	62	63	2	8-1: 44, 0-7: 18, Mor: 0
factor	male	Employment	2	61	63	3	une: 31, emp: 25, una: 5, stu: 0
factor	male	HAART	1	62	63	2	fir: 50, sec: 12, no : 0
factor	male	SOS_mnemonic	16	47	63	2	hea: 28, low: 19

Brief Pain Inventory

- PSS: Pain severity
- PIS: Pain interference

Continuous variables

```
bpi %>%
  select_if(is.numeric) %>%
  skim_to_wide(.) %>%
  kable(., caption = 'Whole cohort',
        align = 'llrrrrrrrrrr')
```

Table 7: Whole cohort

type	variable	missing	complete	n	mean	sd	median	q25	q75	min	max
numeric	PIS	17	143	160	5.12	2.58	5	3	7	0	10
numeric	PSS	16	144	160	5.03	2.14	5	4	6	0	10

```
bpi %>%
  select(Study_site, PIS, PSS) %>%
  group_by(Study_site) %>%
  skim_to_wide(.) %>%
```

```
kable(., caption = 'By study site',
      align = 'l1lrrrrrrrrrr')
```

Table 8: By study site

type	Study_site	variable	missing	complete	n	mean	sd	median	q25	q75	min	max
numeric	R1	PIS	12	35	47	4.69	3.19	5	1.5	7	0	9
numeric	R1	PSS	12	35	47	5	3.01	5	3	8	0	10
numeric	R2	PIS	1	48	49	4.79	2.25	5	3	6.25	0	10
numeric	R2	PSS	0	49	49	4.61	1.74	5	4	6	2	10
numeric	U1	PIS	4	43	47	5.12	2.31	5	3.5	7	0	9
numeric	U1	PSS	4	43	47	5	1.6	5	4	6	0	8
numeric	U2	PIS	0	17	17	6.94	2.05	7	6	9	3	10
numeric	U2	PSS	0	17	17	6.41	1.84	6	6	7	3	10

```
bpi %>%
  select(Sex, PIS, PSS) %>%
  group_by(Sex) %>%
  skim_to_wide(.) %>%
  kable(., caption = 'By sex',
        align = 'l1lrrrrrrrrrr')
```

Table 9: By sex

type	Sex	variable	missing	complete	n	mean	sd	median	q25	q75	min	max
numeric	female	PIS	5	92	97	5.32	2.39	5	4	7	0	10
numeric	female	PSS	4	93	97	5.05	1.89	5	4	6	0	10
numeric	male	PIS	12	51	63	4.76	2.88	5	2	7	0	9
numeric	male	PSS	12	51	63	5	2.56	5	4	6	0	10

Beck's Depression Index

Continuous variables

```
bdi %>%
  select_if(is.numeric) %>%
  skim_to_wide(.) %>%
  kable(., caption = 'Whole cohort',
        align = 'llrrrrrrrrrr')
```

Table 10: Whole cohort

type	variable	missing	complete	n	mean	sd	median	q25	q75	min	max
numeric	BDI	0	160	160	20.07	13.22	18.5	10	29	0	55

```
bdi %>%
  select(Study_site, BDI) %>%
  group_by(Study_site) %>%
  skim_to_wide(.) %>%
  kable(., caption = 'By study site',
        align = 'l1lrrrrrrrrrr')
```


Table 11: By study site

type	Study_site	variable	missing	complete	n	mean	sd	median	q25	q75	min	max
numeric	R1	BDI	0	47	47	13.47	12.1	13	0	21.5	0	45
numeric	R2	BDI	0	49	49	25.73	11.12	27	17	33	5	49
numeric	U1	BDI	0	47	47	17.51	12.39	16	8	25.5	0	46
numeric	U2	BDI	0	17	17	29.06	13.15	31	22	35	0	55

```

bdi %>%
  select(Sex, BDI) %>%
  group_by(Sex) %>%
  skim_to_wide(.) %>%
  kable(., caption = 'By sex',
        align = 'lllrrrrrrrrrr')

```

Table 12: By sex

type	Sex	variable	missing	complete	n	mean	sd	median	q25	q75	min	max
numeric	female	BDI	0	97	97	24.26	12.65	25	16	33	0	55
numeric	male	BDI	0	63	63	13.62	11.44	13	4	20	0	45

EQ5D (3L)

Continuous variables

```

eq5d %>%
  select_if(is.numeric) %>%
  skim_to_wide(.) %>%
  kable(., caption = 'Whole cohort',
        align = 'llrrrrrrrrrr')

```

Table 13: Whole cohort

type	variable	missing	complete	n	mean	sd	median	q25	q75	min	max
numeric	EQ5D_index	18	142	160	0.62	0.19	0.69	0.49	0.76	-0.21	1
numeric	EQ5D_VAS	16	144	160	59.52	21.17	60	50	76.25	0	100

```

eq5d %>%
  select(Study_site, EQ5D_index, EQ5D_VAS) %>%
  group_by(Study_site) %>%
  skim_to_wide(.) %>%
  kable(., caption = 'By study site',
        align = 'lllrrrrrrrrrr')

```

Table 14: By study site

type	Study_site	variable	missing	complete	n	mean	sd	median	q25	q75	min	max
numeric	R1	EQ5D_index	12	35	47	0.67	0.21	0.73	0.66	0.78	-0.21	1
numeric	R1	EQ5D_VAS	12	35	47	62.46	23.1	69	50	80	10	100
numeric	R2	EQ5D_index	0	49	49	0.66	0.18	0.73	0.62	0.8	-0.05	0.8
numeric	R2	EQ5D_VAS	0	49	49	59.9	16.6	60	50	70	20	90

type	Study_site	variable	missing	complete	n	mean	sd	median	q25	q75	min	max
numeric	U1	EQ5D_index	6	41	47	0.56	0.17	0.5	0.43	0.73	0.36	1
numeric	U1	EQ5D_VAS	4	43	47	59.77	20.64	60	55	75	0	90
numeric	U2	EQ5D_index	0	17	17	0.53	0.22	0.52	0.46	0.66	0.008	0.85
numeric	U2	EQ5D_VAS	0	17	17	51.76	29.04	60	40	75	0	80

```
eq5d %>%
  select(Sex, EQ5D_index, EQ5D_VAS) %>%
  group_by(Sex) %>%
  skim_to_wide(.) %>%
  kable(., caption = 'By sex',
        align = 'lllrrrrrrrrrr')
```

Table 15: By sex

type	Sex	variable	missing	complete	n	mean	sd	median	q25	q75	min	max
numeric	female	EQ5D_index	5	92	97	0.61	0.19	0.66	0.47	0.76	-0.05	1
numeric	female	EQ5D_VAS	4	93	97	58.12	21.02	60	50	75	0	90
numeric	male	EQ5D_index	13	50	63	0.64	0.2	0.73	0.51	0.76	-0.21	1
numeric	male	EQ5D_VAS	12	51	63	62.08	21.42	70	50	80	10	100

Self-efficacy Questionnaire 6

Continuous variables

```
se6 %>%
  select_if(is.numeric) %>%
  skim_to_wide(.) %>%
  kable(., caption = 'Whole cohort',
        align = 'llrrrrrrrrrr')
```

Table 16: Whole cohort

type	variable	missing	complete	n	mean	sd	median	q25	q75	min	max
numeric	SE6	18	142	160	6.93	2.25	7.5	5	9	1	10

```
se6 %>%
  select(Study_site, SE6) %>%
  group_by(Study_site) %>%
  skim_to_wide(.) %>%
  kable(., caption = 'By study site',
        align = 'lllrrrrrrrrrr')
```

Table 17: By study site

type	Study_site	variable	missing	complete	n	mean	sd	median	q25	q75	min	max
numeric	R1	SE6	12	35	47	6.4	2.24	7	5	8	1	10
numeric	R2	SE6	0	49	49	6.16	2.32	6	5	8	1	10
numeric	U1	SE6	6	41	47	8.54	1.43	9	8	10	4	10
numeric	U2	SE6	0	17	17	6.35	1.8	6	5	8	2	9

```
se6 %>%
  select(Sex, SE6) %>%
  group_by(Sex) %>%
  skim_to_wide(.) %>%
  kable(., caption = 'By sex',
        align = 'lllrrrrrrrrrr')
```

Table 18: By sex

type	Sex	variable	missing	complete	n	mean	sd	median	q25	q75	min	max
numeric	female	SE6	5	92	97	6.9	2.31	8	5	9	1	10
numeric	male	SE6	13	50	63	6.98	2.16	7	6	8.75	1	10

Session information

```
## R version 3.6.0 (2019-04-26)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS Mojave 10.14.4
##
## 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] knitr_1.22      skimr_1.0.5     magrittr_1.5    forcats_0.4.0
## [5] stringr_1.4.0   dplyr_0.8.0.1   purrr_0.3.2     readr_1.3.1
## [9] tidyr_0.8.3     tibble_2.1.1    ggplot2_3.1.1   tidyverse_1.2.1
##
## loaded via a namespace (and not attached):
## [1] Rcpp_1.0.1      highr_0.8        cellranger_1.1.0 pillar_1.3.1
## [5] compiler_3.6.0  plyr_1.8.4       tools_3.6.0     digest_0.6.18
## [9] lubridate_1.7.4 jsonlite_1.6     evaluate_0.13   nlme_3.1-139
## [13] gtable_0.3.0    lattice_0.20-38  pkgconfig_2.0.2 rlang_0.3.4
## [17] cli_1.1.0       rstudioapi_0.10  yaml_2.2.0      haven_2.1.0
## [21] xfun_0.6        withr_2.1.2.9000 xml2_1.2.0      httr_1.4.0
## [25] hms_0.4.2       generics_0.0.2   grid_3.6.0      tidyselect_0.2.5
## [29] glue_1.3.1      R6_2.4.0         readxl_1.3.1    rmarkdown_1.12
## [33] modelr_0.1.4    backports_1.1.4  scales_1.0.0    htmltools_0.3.6
## [37] rvest_0.3.3     assertthat_0.2.1 colorspace_1.4-1 stringi_1.4.3
## [41] lazyeval_0.2.2  munsell_0.5.0    broom_0.5.2     crayon_1.3.4
```