

Supplement 1

Demographics

Peter Kamerman

Last knitted: 16 October 2021

Contents

1	Import and check data	1
2	Basic descriptive statistics	2
2.1	Full cohort	2
2.1.1	Descriptives	2
2.2	By study site	3
2.2.1	Sample size	3
2.2.2	Descriptives	3
3	Session information	5

1 Import and check data

```
# Import
demo <- read_rds('data-cleaned/data-demographics.rds')%>%
  # Fix site names:
  ## DD = Site 1, NM = Site 2, RESI = Site 3, RP = Site 4, STIG = Site 5
  mutate(Site = case_when(
    Site == 'DD' ~ 'Site 1',
    Site == 'NM' ~ 'Site 2',
    Site == 'RESI' ~ 'Site 3',
    Site == 'RP' ~ 'Site 4',
    Site == 'STIG' ~ 'Site 5'
  ))
```

```
# Check
## Demographics
dim(demo)
```

```
## [1] 599 11
```

```
names(demo)
```

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

```
## [10] "Education"          "BDI"
glimpse(demo)

## Rows: 599
## Columns: 11
## $ ID                <chr> "RPB73", "RPB74", "RPB75", "RPB76", "RPB77", "RPB78"~
## $ Site              <chr> "Site 4", "Site 4", "Site 4", "Site 4", "Site 4", "S~
## $ Sex               <chr> "Female", "Female", "Female", "Female", "Female", "F~
## $ Age               <dbl> 36, 27, 39, 36, 31, 32, 28, 37, 31, 25, 31, 24, 35, ~
## $ Employment_status <chr> "Other", "Unemployed", "Other", "Unemployed", "Unemp~
## $ CD4_recent        <dbl> 391, 571, 591, 207, 126, 225, 543, 410, 74, 212, 579~
## $ ART_currently     <chr> "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Ye~
## $ Diabetes          <chr> "No", "No", "No", "No", "No", "No", "No", "No", "No"~
## $ TB               <chr> "No", "No", "No", "No", "Yes", "No", "Yes", "No", "N~
## $ Education         <chr> "Tertiary", "Secondary", "Secondary", "Primary", "Se~
## $ BDI              <dbl> 23, 25, 5, 25, 24, 24, 8, 22, 9, 8, 8, 7, 3, 9, 26, ~
```

2 Basic descriptive statistics

2.1 Full cohort

2.1.1 Descriptives

```
demo %>%
  # Remove ID column
  select(-ID, -Site) %>%
  # Convert character to factor
  mutate_if(is.character, factor) %>%
  my_skim()
```

Table 1: Data summary

Name	Piped data
Number of rows	599
Number of columns	9
Column type frequency:	
factor	6
numeric	3
Group variables	None

Variable type: factor

skim_variable	missing	complete	n_unique	top_counts
Sex	0	599	2	Fem: 484, Mal: 115
Employment_status	49	550	4	Une: 333, Ful: 131, Par: 52, Oth: 34

skim_variable	missing	complete	n_unique	top_counts
ART_currently	5	594	2	Yes: 462, No: 132
Diabetes	241	358	2	No: 342, Yes: 16
TB	228	371	2	No: 261, Yes: 110
Education	37	562	3	Sec: 398, Pri: 99, Ter: 65

Variable type: numeric

skim_variable	missing	complete	mean	sd	p0	p25	p50	p75	p100
Age	8	591	37.23	9.06	19	31	36	42.00	76
CD4_recent	99	500	320.81	238.92	1	155	261	433.25	1232
BDI	348	251	17.91	9.74	1	10	17	25.00	46

2.2 By study site

2.2.1 Sample size

```
demo %>%
  # Select ID
  select(Site) %>%
  # Group by study site
  group_by(Site) %>%
  summarise(Count = n()) %>%
  kable(caption = 'Sample size by study site')
```

Table 4: Sample size by study site

Site	Count
Site 1	60
Site 2	239
Site 3	99
Site 4	151
Site 5	50

2.2.2 Descriptives

```
demo %>%
  # Convert character to factor
  mutate_if(is.character, factor) %>%
  # Remove ID
  select(-ID) %>%
  # Group by site
  group_by(Site) %>%
  my_skim()
```

Table 5: Data summary

Name	Piped data
Number of rows	599
Number of columns	10
Column type frequency:	
factor	6
numeric	3
Group variables	Site

Variable type: factor

skim_variable	Site	missing	complete	n_unique	top_counts
Sex	Site 1	0	60	2	Fem: 39, Mal: 21
Sex	Site 2	0	239	2	Fem: 184, Mal: 55
Sex	Site 3	0	99	2	Fem: 66, Mal: 33
Sex	Site 4	0	151	1	Fem: 151, Mal: 0
Sex	Site 5	0	50	2	Fem: 44, Mal: 6
Employment_status	Site 1	0	60	4	Une: 36, Ful: 17, Par: 6, Oth: 1
Employment_status	Site 2	49	190	3	Une: 126, Ful: 55, Par: 9, Oth: 0
Employment_status	Site 3	0	99	4	Une: 47, Ful: 35, Par: 16, Oth: 1
Employment_status	Site 4	0	151	4	Une: 102, Oth: 31, Par: 14, Ful: 4
Employment_status	Site 5	0	50	4	Une: 22, Ful: 20, Par: 7, Oth: 1
ART_currently	Site 1	0	60	2	Yes: 58, No: 2
ART_currently	Site 2	3	236	2	Yes: 130, No: 106
ART_currently	Site 3	1	98	2	Yes: 95, No: 3
ART_currently	Site 4	0	151	2	Yes: 130, No: 21
ART_currently	Site 5	1	49	1	Yes: 49, No: 0
Diabetes	Site 1	0	60	2	No: 58, Yes: 2
Diabetes	Site 2	239	0	0	No: 0, Yes: 0
Diabetes	Site 3	1	98	2	No: 89, Yes: 9
Diabetes	Site 4	0	151	2	No: 150, Yes: 1
Diabetes	Site 5	1	49	2	No: 45, Yes: 4
TB	Site 1	0	60	2	Yes: 35, No: 25
TB	Site 2	226	13	1	Yes: 13, No: 0
TB	Site 3	1	98	2	No: 56, Yes: 42
TB	Site 4	0	151	2	No: 140, Yes: 11
TB	Site 5	1	49	2	No: 40, Yes: 9
Education	Site 1	2	58	3	Sec: 40, Pri: 9, Ter: 9
Education	Site 2	27	212	3	Sec: 160, Pri: 44, Ter: 8
Education	Site 3	4	95	3	Sec: 61, Pri: 19, Ter: 15
Education	Site 4	2	149	3	Sec: 111, Ter: 20, Pri: 18
Education	Site 5	2	48	3	Sec: 26, Ter: 13, Pri: 9

Variable type: numeric

skim_variable	Site	missing	complete	mean	sd	p0	p25	p50	p75	p100
Age	Site 1	0	60	38.63	5.87	27	34.00	38.5	42.00	58

skim_variable	Site	missing	complete	mean	sd	p0	p25	p50	p75	p100
Age	Site 2	6	233	36.73	8.54	20	31.00	36.0	41.00	76
Age	Site 3	0	99	43.40	9.32	19	35.00	42.0	49.00	66
Age	Site 4	2	149	30.77	4.75	20	27.00	31.0	34.00	40
Age	Site 5	0	50	44.90	9.67	25	37.25	44.5	52.50	65
CD4_recent	Site 1	18	42	428.60	255.93	16	256.00	374.5	577.25	1095
CD4_recent	Site 2	37	202	222.10	180.15	1	103.75	185.5	308.75	963
CD4_recent	Site 3	16	83	452.76	278.07	11	256.50	405.0	651.00	1232
CD4_recent	Site 4	4	147	322.49	206.16	26	174.00	266.0	420.00	1040
CD4_recent	Site 5	24	26	482.88	292.90	50	291.00	434.5	582.00	1138
BDI	Site 1	2	58	18.52	10.79	2	10.00	17.0	25.50	46
BDI	Site 2	239	0	NaN	NA	NA	NA	NA	NA	NA
BDI	Site 3	99	0	NaN	NA	NA	NA	NA	NA	NA
BDI	Site 4	8	143	17.45	9.93	1	9.00	16.0	24.00	42
BDI	Site 5	0	50	18.52	7.87	3	12.00	18.0	25.00	33

3 Session information

```
sessionInfo()
```

```
## R version 4.0.4 (2021-02-15)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Big Sur 10.16
##
## Matrix products: default
## BLAS: /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRblas.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.0/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_1.33      skimr_2.1.3     forcats_0.5.1   stringr_1.4.0
## [5] dplyr_1.0.7     purrr_0.3.4     readr_1.4.0     tidyr_1.1.3
## [9] tibble_3.1.3    ggplot2_3.3.5   tidyverse_1.3.0
##
## loaded via a namespace (and not attached):
## [1] tidyselect_1.1.1 xfun_0.25       repr_1.1.3      haven_2.3.1
## [5] colorspace_2.0-2 vctrs_0.3.8     generics_0.1.0  htmltools_0.5.1.1
## [9] base64enc_0.1-3  yaml_2.2.1      utf8_1.2.2      rlang_0.4.11
## [13] pillar_1.6.2     glue_1.4.2      withr_2.4.2     DBI_1.1.1
## [17] dbplyr_2.1.0     modelr_0.1.8    readxl_1.3.1    lifecycle_1.0.0
## [21] munsell_0.5.0    gtable_0.3.0    cellranger_1.1.0 rvest_1.0.0
## [25] evaluate_0.14    fansi_0.5.0     highr_0.9       broom_0.7.9
## [29] Rcpp_1.0.7       scales_1.1.1    backports_1.2.1 jsonlite_1.7.2
## [33] fs_1.5.0         hms_1.0.0       digest_0.6.27   stringi_1.7.3
```

## [37]	grid_4.0.4	cli_3.0.1	tools_4.0.4	magrittr_2.0.1
## [41]	crayon_1.4.1	pkgconfig_2.0.3	ellipsis_0.3.2	xml2_1.3.2
## [45]	reprex_1.0.0	lubridate_1.7.10	assertthat_0.2.1	rmarkdown_2.10
## [49]	httr_1.4.2	rstudioapi_0.13	R6_2.5.0	compiler_4.0.4