Descriptive Analyses for SANE Training Program Data

Steven J. Pierce[1]

[1]Michigan State University, Center for Statistical Training and Consulting, https://cstat.msu.edu

Table of contents

1 Purpose

This file is part of a research compendium (Pierce, 2026) associated with a study about a sexual assault nurse examiner training program (Dontje & Campbell, 07/01/2021–06/30/2025). It contains results from descriptive analyses.

2 Setup

This section documents some setup tasks that are useful to the statistician on the team. Most readers of this document will probably want skip directly to Section 3.

2.1 Define Global Options

Global R chunk options are defined in the YAML header but local chunk options will over-ride global options. We can temporarily disable an individual chunk by inserting `#| eval: false` on a line at the top of the chunk. The method for creating a `cfsize` option that controls font size in code chunks and their text output is based on an answer to a question posted on stackoverflow.com.

```
```{r}
#| label: global-options

# Create a custom chunk hook/option for controlling font size in chunk & output.
def.chunk.hook  <- knitr::knit_hooks$get("chunk")
knitr::knit_hooks$set(chunk = function(x, options) {
  x <- def.chunk.hook(x, options)
  ifelse(options$cfsize != "normalsize",
         paste0("\n \\", options$cfsize,"\n\n", x, "\n\n \\normalsize"),
         x)
  })
```
```

2.2 Load Packages

R packages usually add new functions to the base R software, allowing you to do more things. Here, we load the specific R packages required for this script to work.

```
```{r}
#| label: load-packages
library(devtools)       # for session_info()
```
```

```
Loading required package: usethis
```

```
```{r}
#| label: load-packages
library(here)           # for here(), i_am(), makes code more portable.
```
```

```
here() starts at P:/Consulting/Cases_1600-1799/C1788/SANETPA/scripts
```

```
```{r}
#| label: load-packages
library(rmarkdown)      # for pandoc_version()
library(knitr)          # for kable()
library(dplyr)          # for %>%, filter(), group_by(), mutate(), rename(), etc.
```
```

```
Attaching package: 'dplyr'


The following objects are masked from 'package:stats':

    filter, lag


The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

```{r}
#| label: load-packages
library(effectsize)     # for cramers_v()
library(tidyverse)      # for map_dfr(), map_chr(), rowid_to_column(), etc.
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v forcats   1.0.1     v readr     2.1.6
v ggplot2   4.0.2     v stringr   1.6.0
v lubridate 1.9.5     v tibble    3.3.1
v purrr     1.2.1     v tidyr     1.3.2


-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```{r}
#| label: load-packages
library(haven)          # for as_factor()
library(labelled)       # for var_label()
library(janitor)        # for adorn_totals()
```

```
Attaching package: 'janitor'

The following objects are masked from 'package:stats':

    chisq.test, fisher.test
```

```{r}
#| label: load-packages
options(kableExtra.latex.load_packages = FALSE)
library(kableExtra)     # for kable_styling(), add_header_above(),
```

```
Attaching package: 'kableExtra'

The following object is masked from 'package:dplyr':

    group_rows
```

```{r}
#| label: load-packages
                        # column_spec(), row_spec() etc.
library(piercer)        # for file_details(), git_report(), which_latex()
library(PropCIs)        # for scoreci()
library(psych)          # for alpha()
```

```
Attaching package: 'psych'

The following objects are masked from 'package:ggplot2':

    %+%, alpha

The following object is masked from 'package:effectsize':

    phi
```

```{r}
#| label: load-packages
library(quarto)         # for quarto_version()
library(SANETPA)        # for version info
```

2.3 Declare Path

This next chunk declares the path to this script relative to the project-level root directory. If the file is not in the right location under the project root you'll get a warning message. This helps ensure relative paths are all working as expected. The chunk below uses the `SourceDir` and `SourceFile` parameters set in the YAML header.

```
```{r}
#| label: declare-path

# Declare path to this script relative to the project root directory.
here::i_am(path = paste0(params$SourceDir, params$SourceFile))
```
```

```
here() starts at P:/Consulting/Cases_1600-1799/C1788/SANETPA
```

3 Read Data

We start be reading in the datafile, which contains the following datasets:

- `Applicants`. This is a person-level dataset containing one row for every applicant, regardless of eligibility status.
- `Eligible_Applicants`. This is a person-level file containing one row for every person who both applied and was eligible for the program. It omits those who were ineligible.
- `Enrolled_Applicants`. This is a person-level file containing one row for every person who was eligible for the program and actually enrolled. It omits those who were not enrolled.
- `Enrolled_Applicants_CD`. This is a person-level file containing one row for every person who was eligible for the program, actually enrolled, and had complete data on the variables to be used in the stopping-ratio models for RQ2 and RQ3. It omits those who had incomplete data on those variables.
- `StartedDT_Applicants`. This is a person-level file containing one row for every person who was eligible for the program, enrolled, and actually started the didactic training.
- `Thresholds`. This is a person-threshold level file with one row per threshold encountered by each eligible applicant.

```
```{r}
#| label: load-data
#| eval: true

# Store path to data file.
DataFile <- here("data/Imported_SANETP_Data.RData")

load(file = DataFile)
```
```

Table 1 shows meta-data about the data file we just loaded and Table 2 shows the sizes of the datasets it contains.

```
```{r}
#| label: tbl-imported-data-file
#| tbl-cap: "Meta-Data About the Data File Loaded"

file_details(DataFile) %>%
  kable(, format = "latex", booktabs = TRUE,
        col.names = c("File Name", "Size", "Last Modified")) %>%
  kable_styling()
```
```

Table 1: Meta-Data About the Data File Loaded

File Name	Size	Last Modified
Imported_SANETP_Data.RData	181K	2026-02-14 12:56:34

Next I expand the `Enrolled_Applicants_CD` dataset by converting a couple variables to factors for later convenience.

```{r}
#| label: update-Enrolled-Appplicants-CD

Enrolled_Applicants_CD <- Enrolled_Applicants_CD %>%
  # Convert categorical predictors to factors with a "." variable name suffix
  mutate(Setting. = as_factor(Setting),
         Motivation_NeedSANE. = as_factor(Motivation_NeedSANE))
```

```{r}
#| label: tbl-datasets
#| tbl-cap: "Sizes of the Datasets"

data.frame(Dataset = c("Applicants", "Eligible_Applicants",
                       "Enrolled_Applicants", "Enrolled_Applicants_CD",
                       "StartedDT_Applicants", "Thresholds"),
           N_Rows = c(nrow(Applicants), nrow(Eligible_Applicants),
                      nrow(Enrolled_Applicants), nrow(Enrolled_Applicants_CD),
                      nrow(StartedDT_Applicants), nrow(Thresholds)),
           N_Cols = c(ncol(Applicants), ncol(Eligible_Applicants),
                      ncol(Enrolled_Applicants), ncol(Enrolled_Applicants_CD),
                      ncol(StartedDT_Applicants), ncol(Thresholds))) %>%
  kable(, format = "latex", booktabs = TRUE,
        col.names = c("Dataset", "N Rows", "N Columns")) %>%
  kable_styling()
```

Table 2: Sizes of the Datasets

Dataset	N Rows	N Columns
Applicants	497	154
Eligible_Applicants	327	154
Enrolled_Applicants	254	154
Enrolled_Applicants_CD	252	161
StartedDT_Applicants	235	159
Thresholds	661	21

4 Summaries of Applicants Data

There are 497 rows in the `Applicants` dataset. The first thing we need to examine is the distribution for a few key variables. Table 3 shows the crosstabulation of the raw stage variable by the eligibility status.

```{r}
#| label: tbl-Stage-Raw-Eligible
#| tbl-cap: Stage_Raw by Eligible Contingency Table

# Footnote text.
FN <- "CSW, clinical skills workshop; DT, didactic training."

Applicants %>%
  xtabs(~as_factor(Stage_Raw) + Eligible, addNA = TRUE, data = .) %>%
  addmargins() %>%
```

```
  kable(format = "latex", booktabs = TRUE,
        col.names = c("Stage_Raw", "No", "Yes", "Sum")) %>%
  kable_styling() %>%
  add_header_above(header = c(" ", "Eligible" = 2, " ")) %>%
  column_spec(column = 4, italic = TRUE) %>%
  row_spec(row = 8, italic = TRUE) %>%
  footnote(kable_input = ., general = FN, footnote_as_chunk = TRUE,
           threeparttable = TRUE)
```

Table 3: Stage_Raw by Eligible Contingency Table

	Eligible		
Stage_Raw	No	Yes	Sum
Applied	170	0	*170*
Eligible	0	73	*73*
Enrolled	0	17	*17*
Started DT	0	61	*61*
Finished DT	0	16	*16*
Started CSW	0	1	*1*
Finished CSW	0	159	*159*
Sum	*170*	*327*	497

Note: CSW, clinical skills workshop; DT, didactic training.

> **Tip**
>
> - People who exited at the Applied stage should all be coded as ineligible.
> - People in the remaining stages should all be coded as eligible.
> - Cell count for Applied/Yes should be 0. Positive values there may be a result of inaccurate data entry (we previously detected and fixed 4 cases with this check).

Table 4 and Table 5 respectively show the joint distributions of the `Barrier_FO` and `Barrier_WR` scores with the `Eligible` indicator. This allows us to demonstrate that all eligible applicants have observed scores in the expected range and many ineligible applicants have missing values.

```
```{r}
#| label: tbl-Barrier-FO-Eligible
#| tbl-cap: Barrier_FO by Eligible Contingency Table

# Footnote text.
FN <- paste("The range of possible values for Barrier_FO is [1, 5].",
            "NA, not available (missing data).")

Applicants %>%
  xtabs(~Barrier_FO + Eligible, addNA = TRUE, data = .) %>%
  addmargins() %>%
  kable(format = "latex", booktabs = TRUE,
        col.names = c("Barrier_FO", "No", "Yes", "Sum")) %>%
  kable_styling() %>%
  add_header_above(header = c(" ", "Eligible" = 2, " ")) %>%
  column_spec(column = 4, italic = TRUE) %>%
  row_spec(row = 8, italic = TRUE) %>%
  footnote(kable_input = ., general = FN, footnote_as_chunk = TRUE,
           threeparttable = TRUE)
```
```

```
```{r}
#| label: tbl-Barrier-WR-Eligible
#| tbl-cap: Barrier_WR by Eligible Contingency Table
```
```

Table 4: Barrier_FO by Eligible Contingency Table

	Eligible		
Barrier_FO	No	Yes	Sum
1	15	70	*85*
2	32	101	*133*
3	33	114	*147*
4	12	37	*49*
5	3	5	*8*
NA	75	0	*75*
Sum	170	327	*497*

Note: The range of possible values for Barrier_FO is [1, 5]. NA, not available (missing data).

```
# Footnote text.
FN <- paste("The range of possible values for Barrier_WR is [1, 5].",
            "NA, not available (missing data).")

Applicants %>%
  xtabs(~Barrier_WR + Eligible, addNA = TRUE, data = .) %>%
  addmargins() %>%
  kable(format = "latex", booktabs = TRUE,
        col.names = c("Barrier_WR", "No", "Yes", "Sum")) %>%
  kable_styling() %>%
  add_header_above(header = c(" ", "Eligible" = 2, " ")) %>%
  column_spec(column = 4, italic = TRUE) %>%
  row_spec(row = 8, italic = TRUE) %>%
  footnote(kable_input = ., general = FN, footnote_as_chunk = TRUE,
           threeparttable = TRUE)
```

Table 5: Barrier_WR by Eligible Contingency Table

	Eligible		
Barrier_WR	No	Yes	Sum
1	2	16	*18*
2	9	44	*53*
3	32	136	*168*
4	38	95	*133*
5	14	36	*50*
NA	75	0	*75*
Sum	170	327	*497*

Note:
The range of possible values for Barrier_WR is [1, 5]. NA, not available (missing data).

5 Summaries of Enrolled Applicants Data

`Enrolled_Applicants` is the dataset before listwise deletion of applicants with missing data on any variable to be used in the models for RQ2 and RQ3. It has 254 rows. `Enrolled_Applicants_CD` is the dataset of enrolled applicants with complete data (remaining after listwise deletion). It has 252 rows.

5.1 Demographics

Here we provide descriptive statistics on demographic variables for the set of enrolled applicants with complete data (remaining after listwise deletion). While our RQ1 analyses use a broader sample, the SR models used for RQ2 and RQ3 are the main study results, so we wanted to provide descriptives for the sample used in those analyses. Table 6 shows the age distribution, Table 7 shows the sex distribution, and Table 8 shows the race distribution.

```
```{r}
#| label: tbl-Age
#| tbl-cap: Descriptive Statistics for Age Variables Among Enrolled Applicants
#|          (After Listwise Deletion)

VNames <- c("Age")
VLabels <- c("Age (years)")
CNames <- c("Variable", "Construct", "N", "Mean", "SD", "Min", "Max")

Enrolled_Applicants %>%
  select(ID, Age) %>%
  pivot_longer(cols = c(Age), names_to = "VarName", values_to = "Value") %>%
  mutate(VarName = factor(VarName, levels = VNames, labels = VNames),
         Construct = factor(VarName, levels = VNames, labels = VLabels)) %>%
  relocate(ID, VarName, Construct) %>%
  group_by(VarName, Construct) %>%
  summarise(N = n(),
            Mean = mean(Value, na.rm = TRUE),
            SD = sd(Value, na.rm = TRUE),
            Min = min(Value, na.rm = TRUE),
            Max = max(Value, na.rm = TRUE)) %>%
  kable(., format = "latex", booktabs = TRUE, digits = 1,
        col.names = CNames, row.names = FALSE) %>%
  kable_styling()
```
```

```
`summarise()` has regrouped the output.
i Summaries were computed grouped by VarName and Construct.
i Output is grouped by VarName.
i Use `summarise(.groups = "drop_last")` to silence this message.
i Use `summarise(.by = c(VarName, Construct))` for per-operation grouping
  (`?dplyr::dplyr_by`) instead.
```

Table 6: Descriptive Statistics for Age Variables Among Enrolled Applicants (After Listwise Deletion)

Variable	Construct	N	Mean	SD	Min	Max
Age	Age (years)	254	38.5	10.4	22	72

```
```{r}
#| label: tbl-Sex
#| tbl-cap: Sex Frequency Distribution Among Enrolled Applicants
#|          (After Listwise Deletion)

# Footnote text.
FN <- paste("Denominator for percents is the frequency sum.")

Enrolled_Applicants_CD %>%
  xtabs(~as_factor(Sex), addNA = TRUE, data = .) %>%
  addmargins() %>%
  as.data.frame() %>%
```
```

```
  rename(Sex = as_factor.Sex.) %>%
  mutate(Percent = 100*Freq/nrow(Enrolled_Applicants_CD),
         Cumulative = if_else(Sex == "Sum",
                              true = 100,
                              false = cumsum(Percent))) %>%
  kable(format = "latex", booktabs = TRUE, digits = 1,
        col.names = c("Sex", "Freq", "%", "Cumulative %")) %>%
  kable_styling() %>%
  row_spec(row = 3, italic = TRUE) %>%
  footnote(kable_input = ., general = FN, footnote_as_chunk = TRUE,
           threeparttable = TRUE)
```

Table 7: Sex Frequency Distribution Among Enrolled Applicants (After Listwise Deletion)

Sex	Freq	%	Cumulative %
Male	4	1.6	1.6
Female	248	98.4	100.0
Sum	*252*	*100.0*	*100.0*

Note: Denominator for percents is the frequency sum.

```
```{r}
#| label: tbl-Race-Ethnicity
#| tbl-cap: Race/Ethnicity Frequency Distribution Among Enrolled Applicants
#|          (After Listwise Deletion)

# Footnote text.
FN <- paste("Denominator for percents is the frequency sum.")

Enrolled_Applicants_CD %>%
  xtabs(~as_factor(Race_Ethnicity), addNA = TRUE, data = .) %>%
  addmargins() %>%
  as.data.frame() %>%
  rename(Race = as_factor.Race_Ethnicity.) %>%
  mutate(Percent = 100*Freq/nrow(Enrolled_Applicants_CD),
         Cumulative = if_else(Race == "Sum",
                              true = 100,
                              false = cumsum(Percent))) %>%
  kable(format = "latex", booktabs = TRUE, digits = 1,
        col.names = c("Race/Ethnicity", "Freq", "%", "Cumulative %")) %>%
  kable_styling() %>%
  row_spec(row = 8, italic = TRUE) %>%
  footnote(kable_input = ., general = FN, footnote_as_chunk = TRUE,
           threeparttable = TRUE)
```
```

Table 8: Race/Ethnicity Frequency Distribution Among Enrolled Applicants (After Listwise Deletion)

Race/Ethnicity	Freq	%	Cumulative %
Caucasian/White (non-hipanic)	205	81.3	81.3
African American/Black	23	9.1	90.5
Hispanic/Latinx	8	3.2	93.7
Native American/Alaska Native	10	4.0	97.6
Enter your own	1	0.4	98.0
Asian American/Asian	2	0.8	98.8
Native Hawaiian/Pacific Islander	3	1.2	100.0
Sum	*252*	*100.0*	*100.0*

Note: Denominator for percents is the frequency sum.

5.2 Outcome Distribution

`Stage_Reached` is the stage variable we will use for modeling purposes. Table 9 shows the frequency distribution for the stage reached among all eligible applicants (before listwise deletion), while @ tbl-Stage-Reached-cd shows it after removing those with incomplete data on the variables to be included in the model (after listwise deletion).

```
```{r}
#| label: tbl-Stage-Reached-all
#| tbl-cap: Stage_Reached Frequency Distribution Among Enrolled Applicants
#|          (Before Listwise Deletion)

# Footnote text.
FN <- paste("Denominator for percents is the frequency sum.",
      "CSW, clinical skills workshop; DT, didactic training.")

Enrolled_Applicants %>%
  xtabs(~as_factor(Stage_Reached), addNA = TRUE, data = .) %>%
  addmargins() %>%
  as.data.frame() %>%
  rename(Stage_Reached = as_factor.Stage_Reached.) %>%
  mutate(Percent = 100*Freq/nrow(Enrolled_Applicants),
         Cumulative = if_else(Stage_Reached == "Sum",
                              true = 100,
                              false = cumsum(Percent))) %>%
  kable(format = "latex", booktabs = TRUE, digits = 1,
        col.names = c("Stage_Reached", "Freq", "%", "Cumulative %")) %>%
  kable_styling() %>%
  row_spec(row = 6, italic = TRUE) %>%
  footnote(kable_input = ., general = FN, footnote_as_chunk = TRUE,
           threeparttable = TRUE)
```
```

Table 9: Stage_Reached Frequency Distribution Among Enrolled Applicants (Before Listwise Deletion)

Stage_Reached	Freq	%	Cumulative %
Enrolled	17	6.7	6.7
Started DT	61	24.0	30.7
Finished DT	17	6.7	37.4
Finished CSW	159	62.6	100.0
Sum	254	100.0	100.0

Note: Denominator for percents is the frequency sum. CSW, clinical skills workshop; DT, didactic training.

```
```{r}
#| label: tbl-Stage-Reached-cd
#| tbl-cap: Stage_Reached Frequency Distribution Among Enrolled Applicants
#|          (After Listwise Deletion)

# Footnote text.
FN <- paste("Denominator for percents is the frequency sum.",
      "CSW, clinical skills workshop; DT, didactic training.")

Enrolled_Applicants_CD %>%
  xtabs(~as_factor(Stage_Reached), addNA = TRUE, data = .) %>%
  addmargins() %>%
  as.data.frame() %>%
  rename(Stage_Reached = as_factor.Stage_Reached.) %>%
  mutate(Percent = 100*Freq/nrow(Enrolled_Applicants_CD),
         Cumulative = if_else(Stage_Reached == "Sum",
                              true = 100,
                              false = cumsum(Percent))) %>%
  kable(format = "latex", booktabs = TRUE, digits = 1,
        col.names = c("Stage_Reached", "Freq", "%", "Cumulative %")) %>%
  kable_styling() %>%
  row_spec(row = 6, italic = TRUE) %>%
  footnote(kable_input = ., general = FN, footnote_as_chunk = TRUE,
           threeparttable = TRUE)
```
```

Table 10: Stage_Reached Frequency Distribution Among Enrolled Applicants (After Listwise Deletion)

Stage_Reached	Freq	%	Cumulative %
Enrolled	17	6.7	6.7
Started DT	61	24.2	31.0
Finished DT	17	6.7	37.7
Finished CSW	157	62.3	100.0
Sum	252	100.0	100.0

Note: Denominator for percents is the frequency sum. CSW, clinical skills workshop; DT, didactic training.

5.3 Relationship Between Setting and Motivation_NeedSANE

Previous research found nurses in rural settings were more likely to complete a SANE training program than nurses from other settings (Patterson & Resko, 2015). Our analysis did not find that, but instead found that Motivation_NeedSANE was associated with lower attrition (see RQ2/RQ3 output). Patterson & Resko (2015) had also measured similar motivation variable but it did not make it into their logistic regression models (they used a bivariate association screening process). To help us think about the difference in findings, I tested whether Setting and Motivation_NeedSANE are independent using a chi-square test, then computed a bias-corrected estimate of Cramer's V (which measures association between nominal variables and can range from 0 to 1) (Ben-Shachar et al., 2020, 2023). We use the data for enrolled applicants with complete data on the variables used in the SR models for this.

```
```{r}
#| label: tbl-Setting-Motivation_NeedSANE
#| tbl-cap: Contingency Table for Setting by Motivation_NeedSANE Among Enrolled
#|          Applicants with Complete Data
#| message: false

CT <- Enrolled_Applicants_CD %>%
  xtabs(~ Setting. + Motivation_NeedSANE., addNA = FALSE, data = .)

CT.CV <- cramers_v(x = CT, adjust = TRUE, ci = 0.95, alternative = "greater")

FN <- paste0("Chi-square test for independence, X^2(",
             chisq.test(x = CT)$parameter, ") = ",
             round(chisq.test(x = CT)$statistic, digits = 2),
             ", p = ", display_num(chisq.test(x = CT)$p.value), "; ",
             "Cramer's V (adjusted) = ",
             format(round(CT.CV$Cramers_v_adjusted, digits = 2), nsmall = 2),
             ", 1-tailed 95% CI = [",
             round(CT.CV$CI_low, digits = 2), ", ",
             format(CT.CV$CI_high, nsmall = 2), "].")

bind_rows(dc_summary(data = Enrolled_Applicants_CD,
          vars = c("Motivation_NeedSANE")),
          dc_summary(data = Enrolled_Applicants_CD %>% filter(Setting == 1),
          vars = c("Motivation_NeedSANE")),
          dc_summary(data = Enrolled_Applicants_CD %>% filter(Setting == 2),
          vars = c("Motivation_NeedSANE")),
          dc_summary(data = Enrolled_Applicants_CD %>% filter(Setting == 3),
          vars = c("Motivation_NeedSANE"))) %>%
  bind_cols(Setting = c("Overall", "Urban", "Rural/Tribal", "Suburban")) %>%
  mutate(No = as.character(No),
         Yes = as.character(Yes),
         Total = as.character(Total),
         Setting = factor(Setting, levels = c("Overall", "Urban",
                                              "Rural/Tribal", "Suburban"))) %>%
  relocate(Variable, Setting) %>%
  arrange(Variable, Setting) %>%
  kable(format = "latex", booktabs = TRUE, digits = 2, align = "llrrrrrr",
        format.args = list(nsmall = 2)) %>%
  kable_styling() %>%
  add_header_above(header = c(" " = 2,"Motivation (N)" = 3, " ",
```
```

```
                          "95% CI" = 2)) %>%
  collapse_rows(columns = 1:2, valign = "top", latex_hline = "major",
                row_group_label_position = "first") %>%
  footnote(kable_input = ., general = FN, footnote_as_chunk = TRUE,
           threeparttable = TRUE)
```

Table 11: Contingency Table for Setting by Motivation_NeedSANE Among Enrolled Applicants with Complete Data

		Motivation (N)				95% CI	
Variable	Setting	No	Yes	Total	Rate	LL	UL
Motivation_NeedSANE	Overall	59	193	252	0.77	0.71	0.81
	Urban	23	65	88	0.74	0.64	0.82
	Rural/Tribal	8	71	79	0.90	0.81	0.95
	Suburban	28	57	85	0.67	0.57	0.76

Note: Chi-square test for independence, X^2(2) = 12.44, p = 0.002; Cramer's V (adjusted) = 0.20, 1-tailed 95% CI = [0.06, 1.00].

Tip

Table 11 shows that Setting and Motivation_NeedSANE are not independent. Nurses from rural/tribal settings report the highest rate of endorsing being motivated to participate in the training by a need for SANE service in their communities. However, the association has a small effect size.

5.4 Reliability: Barrier Time Demands

We considered creating a measure of the extent to which competing personal and professional demands on participants' time are a barrier to completing the training program. The program staff believe that this construct is likely to be a major driver of attrition from the program. Similarly, external stakeholders familiar with forensic nursing emphasized competing demands on time/bandwidth as a likely factor in attrition. Therefore, it is a high priority to have some measure of barriers in our models. The challenge is operationalizing it.

There are three items from the training application form that are good candidates for inclusion in such a scale.

- `Train_Goal_Fac_Bar_5R`: I have time in my schedule to focus on this course.
- `Train_Goal_Fac_Bar_8`: I have a lot of work responsibilities.
- `Train_Goal_Fac_Bar_10`: I have a lot of family obligations right now.

These ordinal Likert-response items are all coded such that low values indicate few barrier time demands and high values mean more barrier time demands.

> **Tip**
>
> We use the data for all eligible applicants (before listwise deletion) in this section. It does not make any meaningful difference if we run the same analyses on the subset of eligible applicants with complete data on the variables to be used in modeling.

5.4.1 Item Distributions

The frequency distributions for each of those items are shown in Table 12, Table 13, and Table 14.

```
```{r}
#| label: tbl-Train-Goal-Fac-Bar-5R
#| tbl-cap: Train_Goal_Fac_Bar_5R Frequency Distribution Among Enrolled Applicants

# Footnote text.
FN <- paste("Item text: 'I have time in my schedule to focus on this course.'",
            "Denominator for percents is total number of enrolled applicants.")

Enrolled_Applicants %>%
xtabs(~as_factor(Train_Goal_Fac_Bar_5R), addNA = TRUE, data = .) %>%
  addmargins() %>%
  as.data.frame() %>%
  rename(Train_Goal_Fac_Bar_5R = as_factor.Train_Goal_Fac_Bar_5R.) %>%
  mutate(Percent = 100*Freq/nrow(Enrolled_Applicants),
         Cumulative = if_else(Train_Goal_Fac_Bar_5R == "Sum",
                              true = 100,
                              false = cumsum(Percent))) %>%
  kable(format = "latex", booktabs = TRUE, digits = 1,
        col.names = c("Train_Goal_Fac_Bar_5R", "Freq", "%", "Cumulative %")) %>%
  kable_styling() %>%
  row_spec(row = 6, italic = TRUE) %>%
  footnote(kable_input = ., general = FN, footnote_as_chunk = TRUE,
           threeparttable = TRUE)
```
```

Table 12: Train_Goal_Fac_Bar_5R Frequency Distribution Among Enrolled Applicants

Train_Goal_Fac_Bar_5R	Freq	%	Cumulative %
strongly agree	219	86.2	86.2
somewhat agree	34	13.4	99.6
neither agree nor disagree	1	0.4	100.0
somewhat disagree	0	0.0	100.0
strongly disagree	0	0.0	100.0
Sum	*254*	*100.0*	*100.0*

Note: Item text: 'I have time in my schedule to focus on this course.' Denominator for percents is total number of enrolled applicants.

```
```{r}
#| label: tbl-Train-Goal-Fac-Bar-8
#| tbl-cap: Train_Goal_Fac_Bar_8 Frequency Distribution Among Enrolled Applicants

# Footnote text.
FN <- paste("Item text: 'I have a lot of work responsibilities.'",
            "Denominator for percents is total number of enrolled applicants.")

Enrolled_Applicants %>%
xtabs(~as_factor(Train_Goal_Fac_Bar_8), addNA = TRUE, data = .) %>%
  addmargins() %>%
  as.data.frame() %>%
  rename(Train_Goal_Fac_Bar_8 = as_factor.Train_Goal_Fac_Bar_8.) %>%
  mutate(Percent = 100*Freq/nrow(Enrolled_Applicants),
```
```

```
        Cumulative = if_else(Train_Goal_Fac_Bar_8 == "Sum",
                             true = 100,
                             false = cumsum(Percent))) %>%
  kable(format = "latex", booktabs = TRUE, digits = 1,
        col.names = c("Train_Goal_Fac_Bar_8", "Freq", "%", "Cumulative %")) %>%
  kable_styling() %>%
  row_spec(row = 6, italic = TRUE) %>%
  footnote(kable_input = ., general = FN, footnote_as_chunk = TRUE,
           threeparttable = TRUE)
```

Table 13: Train_Goal_Fac_Bar_8 Frequency Distribution Among Enrolled Applicants

Train_Goal_Fac_Bar_8	Freq	%	Cumulative %
strongly disagree	12	4.7	4.7
somewhat disagree	38	15.0	19.7
neither agree or disagree	109	42.9	62.6
somewhat agree	69	27.2	89.8
strongly agree	26	10.2	100.0
Sum	*254*	*100.0*	*100.0*

Note: Item text: 'I have a lot of work responsibilities.' Denominator for percents is total number of enrolled applicants.

```
```{r}
#| label: tbl-Train-Goal-Fac-Bar-10
#| tbl-cap: Train_Goal_Fac_Bar_10 Frequency Distribution Among Enrolled Applicants

# Footnote text.
FN <- paste("Item text: 'I have a lot of family obligations right now.'",
            "Denominator for percents is total number of enrolled applicants.")

Enrolled_Applicants %>%
xtabs(~as_factor(Train_Goal_Fac_Bar_10), addNA = TRUE, data = .) %>%
  addmargins() %>%
  as.data.frame() %>%
  rename(Train_Goal_Fac_Bar_10 = as_factor.Train_Goal_Fac_Bar_10.) %>%
  mutate(Percent = 100*Freq/nrow(Enrolled_Applicants),
         Cumulative = if_else(Train_Goal_Fac_Bar_10 == "Sum",
                              true = 100,
                              false = cumsum(Percent))) %>%
  kable(format = "latex", booktabs = TRUE, digits = 1,
        col.names = c("Train_Goal_Fac_Bar_10", "Freq", "%", "Cumulative %")) %>%
  kable_styling() %>%
  row_spec(row = 6, italic = TRUE) %>%
  footnote(kable_input = ., general = FN, footnote_as_chunk = TRUE,
           threeparttable = TRUE)
```
```

5.4.2 3-Item Scale Reliability

Our first attempt at building this scale used all three items.

```
```{r}
#| label: reliability-Barrier-TD-3item

Enrolled_Applicants %>%
  select(Train_Goal_Fac_Bar_5R, Train_Goal_Fac_Bar_8, Train_Goal_Fac_Bar_10) %>%
  as.data.frame() %>%
  alpha(x = ., cumulative = TRUE, discrete = TRUE)
```
```

```
Warning in response.frequencies(x, max = max): response.frequency has been
deprecated and replaced with responseFrequecy.  Please fix your call
```

Table 14: Train_Goal_Fac_Bar_10 Frequency Distribution Among Enrolled Applicants

Train_Goal_Fac_Bar_10	Freq	%	Cumulative %
strongly disagree	57	22.4	22.4
somewhat disagree	81	31.9	54.3
neither agree or disagree	90	35.4	89.8
somewhat agree	24	9.4	99.2
strongly agree	2	0.8	100.0
Sum	*254*	*100.0*	*100.0*

Note: Item text: 'I have a lot of family obligations right now.' Denominator for percents is total number of enrolled applicants.

```
Reliability analysis
Call: alpha(x = ., cumulative = TRUE, discrete = TRUE)

  raw_alpha std.alpha G6(smc) average_r S/N  ase mean  sd median_r
      0.54      0.58    0.49      0.32 1.4 0.04  6.7 1.8     0.28

    95% confidence boundaries
         lower alpha upper
Feldt     0.44  0.54  0.63
Duhachek  0.47  0.54  0.62

 Reliability if an item is dropped:
                     raw_alpha std.alpha G6(smc) average_r  S/N alpha se var.r
Train_Goal_Fac_Bar_5R      0.58      0.58    0.41      0.41 1.37    0.053    NA
Train_Goal_Fac_Bar_8       0.31      0.44    0.28      0.28 0.79    0.055    NA
Train_Goal_Fac_Bar_10      0.29      0.42    0.26      0.26 0.71    0.056    NA
                     med.r
Train_Goal_Fac_Bar_5R  0.41
Train_Goal_Fac_Bar_8   0.28
Train_Goal_Fac_Bar_10  0.26

 Item statistics
                       n raw.r std.r r.cor r.drop mean   sd
Train_Goal_Fac_Bar_5R 254  0.50  0.70  0.42   0.33  1.1 0.36
Train_Goal_Fac_Bar_8  254  0.83  0.75  0.55   0.43  3.2 0.98
Train_Goal_Fac_Bar_10 254  0.82  0.76  0.57   0.44  2.3 0.96

Non missing response frequency for each item
                        1    2    3    4    5 miss
Train_Goal_Fac_Bar_5R 0.86 0.13 0.00 0.00 0.00    0
Train_Goal_Fac_Bar_8  0.05 0.15 0.43 0.27 0.10    0
Train_Goal_Fac_Bar_10 0.22 0.32 0.35 0.09 0.01    0
```

The results above show poor to questionable reliability and that omitting `Train_Goal_Fac_Bar_5R` may increase the reliability by a very small amount. Table 12 provides some clues about why that may be happening. The item has a very skewed distribution: none of the enrolled applicants disagreed with the idea that they had enough time in their schedule to focus on this course. One concern is that applicants may have perceived this item to be a screening question where *expressing disagreement might prevent them from getting into the program.* It may have elicited less honest and accurate answers than the other two questions, which have more variability, less skew, and wording less likely to be seen as a pure screening tool.

> **Note**
>
> We decided to omit the problematic item `Train_Goal_Fac_Bar_5R` and consider just using the other two items.

5.4.3 2-Item Scale Reliability

We know from basic psychometric theory that reliability tends to increase with the number of items used to measure the construct. Scales with two items are unlikely to have high reliability. The point estimate of Cronbach's alpha for this reduced version of the scale is slightly better than the one for the three-item version but still in the questionable range. The upper bound on the 95% CI is now just below the 0.70 threshold often used as a value considered acceptable for research purposes.

```
```{r}
#| label: reliability-Barrier-TD-2item

Enrolled_Applicants %>%
  select(Train_Goal_Fac_Bar_8, Train_Goal_Fac_Bar_10) %>%
  as.data.frame() %>%
  alpha(x = ., cumulative = TRUE, discrete = TRUE)
```
```

```
Warning in response.frequencies(x, max = max): response.frequency has been
deprecated and replaced with responseFrequecy.  Please fix your call


Reliability analysis
Call: alpha(x = ., cumulative = TRUE, discrete = TRUE)

  raw_alpha std.alpha G6(smc) average_r S/N   ase mean  sd median_r
      0.58      0.58    0.41      0.41 1.4 0.053  5.6 1.6     0.41

    95% confidence boundaries
         lower alpha upper
Feldt     0.46  0.58  0.67
Duhachek  0.47  0.58  0.68

 Reliability if an item is dropped:
                    raw_alpha std.alpha G6(smc) average_r  S/N alpha se var.r
Train_Goal_Fac_Bar_8       0.42      0.41    0.17      0.41 0.68       NA     0
Train_Goal_Fac_Bar_10      0.39      0.41    0.17      0.41 0.68       NA     0
                    med.r
Train_Goal_Fac_Bar_8   0.41
Train_Goal_Fac_Bar_10  0.41

 Item statistics
                      n raw.r std.r r.cor r.drop mean   sd
Train_Goal_Fac_Bar_8  254  0.84  0.84  0.53   0.41  3.2 0.98
Train_Goal_Fac_Bar_10 254  0.83  0.84  0.53   0.41  2.3 0.96

Non missing response frequency for each item
                       1    2    3    4    5 miss
Train_Goal_Fac_Bar_8  0.05 0.15 0.43 0.27 0.10    0
Train_Goal_Fac_Bar_10 0.22 0.32 0.35 0.09 0.01    0
```

> **Note**
>
> Rather than use a two-item scale with questionable reliability for measuring barriers, the PIs decided it was better to use those two items as separate predictors. Single item measures are not optimal, but these items have strong face validity. They are only modestly correlated, so it may be that treating them as separate predictors for family obligations and work responsibilities will be useful for better understanding attrition. We will recommend pursuing better measurement of these constructs in future studies.
> One consequence of low reliability in a predictor should be a weaker regression coefficient relating it to an outcome. I suspect that we are more likely to underestimate the effects of family obligations and work responsibilities on attrition than to overestimate them.
> We created new versions of these two items for use in modeling because shorter variable names make code and output more readable.
>
> - `Train_Goal_Fac_Bar_8` became `Barrier_WR`.
> - `Train_Goal_Fac_Bar_10` became `Barrier_FO`.

5.5 Predictor Distributions

5.5.1 Continuous Predictors

We plan to use the ProQual compassion (`ProQOL_CS`), burnout (`ProQOL_BO`), and secondary traumatic stress (`ProQOL_STS`) scale scores (Stamm, 2010) as predictors, along with two single-item measures of barriers to participation due to family obligations (`Barrier_FO`) and work responsibilities (`Barrier_WR`) as competing demands on applicants' time.

Table 15 shows descriptive statistics for each of these continuous predictors among all enrolled applicants before application of listwise deletion, while Table 16 shows descriptive statistics for each of these continuous predictors among enrolled applicants with complete data on the variables to be used in modeling, along with corresponding mean-centered versions of each one. Histograms of the uncentered versions of the variables after listwise deletion are shown in Figure 1, Figure 2, Figure 3, Figure 4, and Figure 5.

```
```{r}
#| label: tbl-continuous-vars-all
#| tbl-cap: Descriptive Statistics for Continuous Predictor Variables Among
#|          Enrolled Applicants (Before Listwise Deletion)

# Footnote text.
FN <- paste("N_m, number of missing values; N_o, number of observed values.")

VNames <- c("ProQOL_BO", "ProQOL_CS", "ProQOL_STS", "Barrier_FO", "Barrier_WR",
            "CProQOL_BO", "CProQOL_CS", "CProQOL_STS", "CBarrier_FO",
            "CBarrier_WR")
VLabels <- rep(c("Burnout", "Compassion Satisfaction",
                 "Secondary Traumatic Stress", "Family Obligations",
                 "Work Responsibilities"), times = 2)
CNames <- c("Variable", "Construct", "N", "N_m", "N_o", "Mean", "SD",
            "Min", "Max")

Enrolled_Applicants %>%
  select(ID, ProQOL_BO, ProQOL_CS, ProQOL_STS, Barrier_FO, Barrier_WR) %>%
  pivot_longer(cols = c(ProQOL_BO, ProQOL_CS, ProQOL_STS, Barrier_FO,
                        Barrier_WR),
               names_to = "VarName",
               values_to = "Value") %>%
  mutate(VarName = factor(VarName, levels = VNames, labels = VNames),
         Construct = factor(VarName, levels = VNames, labels = VLabels)) %>%
  relocate(ID, VarName, Construct) %>%
  group_by(VarName, Construct) %>%
  summarise(N = n(),
            N_missing = sum(is.na(Value)),
            N_observed = sum(!is.na(Value)),
            Mean = mean(Value, na.rm = TRUE),
            SD = sd(Value, na.rm = TRUE),
            Min = min(Value, na.rm = TRUE),
            Max = max(Value, na.rm = TRUE)) %>%
  kable(., format = "latex", booktabs = TRUE, digits = 1,
        col.names = CNames, row.names = FALSE) %>%
  kable_styling() %>%
  footnote(kable_input = ., general = FN, footnote_as_chunk = TRUE)
```
```

```
`summarise()` has regrouped the output.
i Summaries were computed grouped by VarName and Construct.
i Output is grouped by VarName.
i Use `summarise(.groups = "drop_last")` to silence this message.
i Use `summarise(.by = c(VarName, Construct))` for per-operation grouping
  (`?dplyr::dplyr_by`) instead.
```

```
```{r}
#| label: tbl-continuous-vars-cd
#| tbl-cap: Descriptive Statistics for Continuous Predictor Variables Among
#|          Enrolled Applicants (After Listwise Deletion)

# Footnote text.
VNames <- c("ProQOL_BO", "ProQOL_CS", "ProQOL_STS", "Barrier_FO", "Barrier_WR",
            "CProQOL_BO", "CProQOL_CS", "CProQOL_STS", "CBarrier_FO",
            "CBarrier_WR")
VLabels <- rep(c("Burnout", "Compassion Satisfaction",
                 "Secondary Traumatic Stress", "Family Obligations",
```
```

Table 15: Descriptive Statistics for Continuous Predictor Variables Among Enrolled Applicants (Before Listwise Deletion)

Variable	Construct	N	N_m	N_o	Mean	SD	Min	Max
ProQOL_BO	Burnout	254	2	252	17.1	3.8	10	30
ProQOL_CS	Compassion Satisfaction	254	2	252	45.4	3.7	36	50
ProQOL_STS	Secondary Traumatic Stress	254	2	252	17.4	3.9	10	30
Barrier_FO	Family Obligations	254	0	254	2.3	1.0	1	5
Barrier_WR	Work Responsibilities	254	0	254	3.2	1.0	1	5

Note: N_m, number of missing values; N_o, number of observed values.

```
               "Work Responsibilities"), times = 2)
CNames <- c("Variable", "Construct", "N", "Mean", "SD", "Min", "Max")

Enrolled_Applicants_CD %>%
  select(ID, ProQOL_BO, ProQOL_CS, ProQOL_STS, Barrier_FO, Barrier_WR,
         CProQOL_BO, CProQOL_CS, CProQOL_STS, CBarrier_FO, CBarrier_WR) %>%
  pivot_longer(cols = c(ProQOL_BO, ProQOL_CS, ProQOL_STS, Barrier_FO,
                        Barrier_WR, CProQOL_BO, CProQOL_CS, CProQOL_STS,
                        CBarrier_FO, CBarrier_WR),
               names_to = "VarName",
               values_to = "Value") %>%
  mutate(VarName = factor(VarName, levels = VNames, labels = VNames),
         Construct = factor(VarName, levels = VNames, labels = VLabels)) %>%
  relocate(ID, VarName, Construct) %>%
  group_by(VarName, Construct) %>%
  summarise(N = n(),
            Mean = mean(Value, na.rm = TRUE),
            SD = sd(Value, na.rm = TRUE),
            Min = round(min(Value, na.rm = TRUE), digits = 1),
            Max = round(max(Value, na.rm = TRUE), digits = 1)) %>%
  kable(., format = "latex", booktabs = TRUE, digits = 1,
        col.names = CNames, row.names = FALSE) %>%
  kable_styling() %>%
  group_rows(., group_label = "Uncentered Scale Scores", start_row = 1,
             end_row = 5, italic = TRUE) %>%
  group_rows(., group_label = "Mean-Centered Scale Scores", start_row = 6,
             end_row = 10, italic = TRUE)
```

```
`summarise()` has regrouped the output.
i Summaries were computed grouped by VarName and Construct.
i Output is grouped by VarName.
i Use `summarise(.groups = "drop_last")` to silence this message.
i Use `summarise(.by = c(VarName, Construct))` for per-operation grouping
  (`?dplyr::dplyr_by`) instead.
```

```
```{r}
#| label: fig-ProQOL-BO
#| fig-cap: Histogram of ProQOL Burnout Among Enrolled Applicants (After
#|          Listwise Deletion)
#| fig-width: 5
#| fig-height: 2.5
#| warning: false

ggplot(data = Enrolled_Applicants_CD, aes(x = ProQOL_BO)) +
  geom_histogram(aes(y = after_stat(density)), binwidth = 1, color = "black",
                 fill = "grey") +
  geom_density(alpha = .6, fill = "#FF6666") +
  geom_vline(aes(xintercept = mean(ProQOL_BO, na.rm = TRUE)), color = "blue",
             linetype = "solid", linewidth = 1) +
  geom_vline(xintercept = c(22.5, 41.5), color = "black", linetype = "dashed",
             linewidth = 1) +
  coord_cartesian(xlim = c(10, 50), ylim = c(0, 0.16)) +
  scale_x_continuous(breaks = seq(from = 10, to = 50, by = 5)) +
  scale_y_continuous(breaks = seq(from = 0, to = 0.16, by = 0.02)) +
  labs(subtitle = "Low / Moderate / High Cutoffs (Dashed Lines)",
       x = "ProQOL Burnout",
       y = "Density")
```
```

Table 16: Descriptive Statistics for Continuous Predictor Variables Among Enrolled Applicants (After Listwise Deletion)

Variable	Construct	N	Mean	SD	Min	Max
Uncentered Scale Scores						
ProQOL_BO	Burnout	252	17.1	3.8	10.0	30.0
ProQOL_CS	Compassion Satisfaction	252	45.4	3.7	36.0	50.0
ProQOL_STS	Secondary Traumatic Stress	252	17.4	3.9	10.0	30.0
Barrier_FO	Family Obligations	252	2.3	1.0	1.0	5.0
Barrier_WR	Work Responsibilities	252	3.2	1.0	1.0	5.0
Mean-Centered Scale Scores						
CProQOL_BO	Burnout	252	0.0	3.8	-7.1	12.9
CProQOL_CS	Compassion Satisfaction	252	0.0	3.7	-9.4	4.6
CProQOL_STS	Secondary Traumatic Stress	252	0.0	3.9	-7.4	12.6
CBarrier_FO	Family Obligations	252	0.0	1.0	-1.3	2.7
CBarrier_WR	Work Responsibilities	252	0.0	1.0	-2.2	1.8

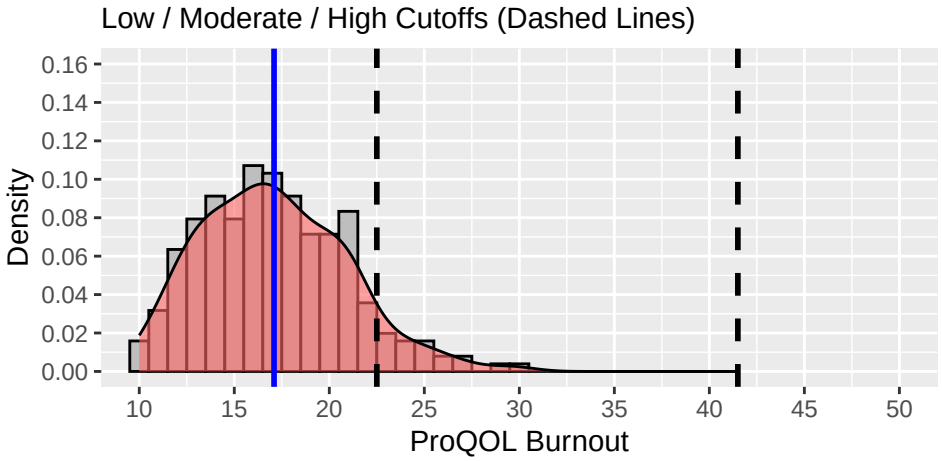

Figure 1: Histogram of ProQOL Burnout Among Enrolled Applicants (After Listwise Deletion)

```
```{r}
#| label: fig-ProQOL-CS
#| fig-cap: Histogram of ProQOL Compassion Satisfaction Among Enrolled
#|          Applicants (After Listwise Deletion)
#| fig-width: 5
#| fig-height: 2.5
#| warning: false

ggplot(data = Enrolled_Applicants_CD, aes(x = ProQOL_CS)) +
  geom_histogram(aes(y = after_stat(density)), binwidth = 1, color = "black",
                 fill = "grey") +
  geom_density(alpha = .6, fill = "#FF6666") +
  geom_vline(aes(xintercept = mean(ProQOL_CS, na.rm = TRUE)), color = "blue",
             linetype = "solid", linewidth = 1) +
  geom_vline(xintercept = c(22.5, 41.5), color = "black", linetype = "dashed",
             linewidth = 1) +
  coord_cartesian(xlim = c(10, 50), ylim = c(0, 0.16)) +
  scale_x_continuous(breaks = seq(from = 10, to = 50, by = 5)) +
  scale_y_continuous(breaks = seq(from = 0, to = 0.16, by = 0.02)) +
  labs(subtitle = "Low / Moderate / High Cutoffs (Dashed Lines)",
       x = "ProQOL Compassion Satisfaction",
       y = "Density")
```
```

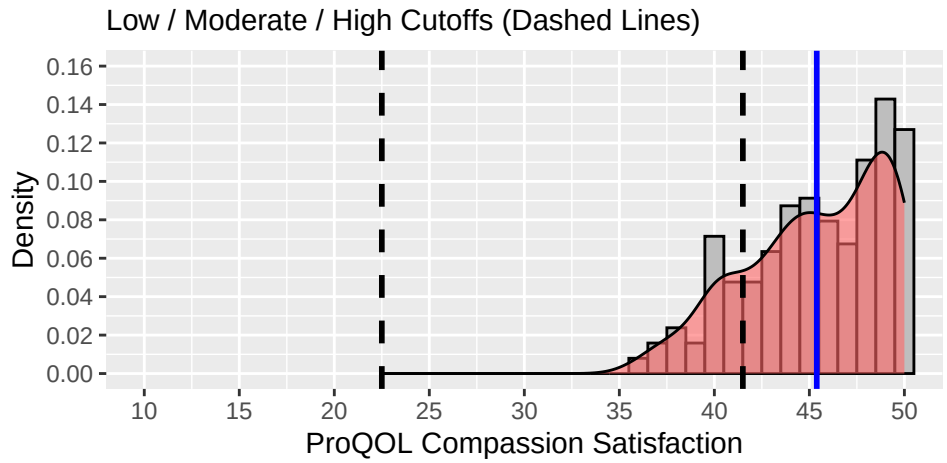

Figure 2: Histogram of ProQOL Compassion Satisfaction Among Enrolled Applicants (After Listwise Deletion)

```
```{r}
#| label: fig-ProQOL-STS
#| fig-cap: Histogram of ProQOL Secondary Traumatic Stress Among Enrolled
#|          Applicants (After Listwise Deletion)
#| fig-width: 5
#| fig-height: 2.5
#| warning: false

ggplot(data = Enrolled_Applicants_CD, aes(x = ProQOL_STS)) +
  geom_histogram(aes(y = after_stat(density)), binwidth = 1, color = "black",
                 fill = "grey") +
  geom_density(alpha = .6, fill = "#FF6666") +
  geom_vline(aes(xintercept = mean(ProQOL_STS, na.rm = TRUE)), color = "blue",
             linetype = "solid", linewidth = 1) +
  geom_vline(xintercept = c(22.5, 41.5), color = "black", linetype = "dashed",
             linewidth = 1) +
  coord_cartesian(xlim = c(10, 50), ylim = c(0, 0.16)) +
  scale_x_continuous(breaks = seq(from = 10, to = 50, by = 5)) +
  scale_y_continuous(breaks = seq(from = 0, to = 0.16, by = 0.02)) +
  labs(subtitle = "Low / Moderate / High Cutoffs (Dashed Lines)",
       x = "ProQOL Secondary Traumatic Stress",
       y = "Density")
```
```

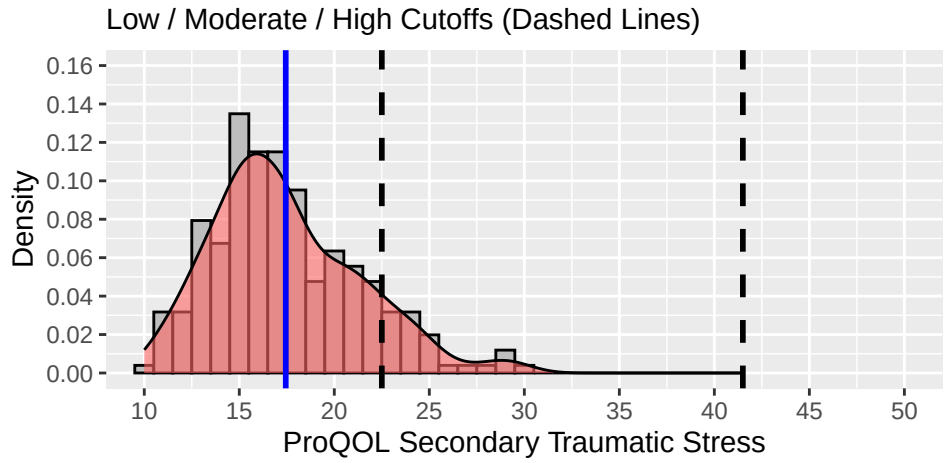

Figure 3: Histogram of ProQOL Secondary Traumatic Stress Among Enrolled Applicants (After Listwise Deletion)

```{r}
#| label: fig-Barrier-FO
#| fig-cap: Histogram of Family Obligations Among Enrolled Applicants (After
#|          Listwise Deletion)
#| fig-width: 5
#| fig-height: 2.5

ggplot(data = Enrolled_Applicants_CD, aes(x = Barrier_FO)) +
  geom_histogram(aes(y = after_stat(density)), binwidth = 1, color = "black",
                 fill = "grey") +
  geom_density(alpha = .6, fill = "#FF6666") +
  geom_vline(aes(xintercept = mean(Barrier_FO, na.rm = TRUE)), color = "blue",
             linetype = "solid", linewidth = 1) +
  coord_cartesian(xlim = c(1, 5), ylim = c(0, 0.70)) +
  scale_x_continuous(breaks = seq(from = 1, to = 5, by = 1)) +
  scale_y_continuous(breaks = seq(from = 0, to = 0.70, by = 0.10)) +
  labs(subtitle = "",
       x = "Family Obligations",
       y = "Density")
```

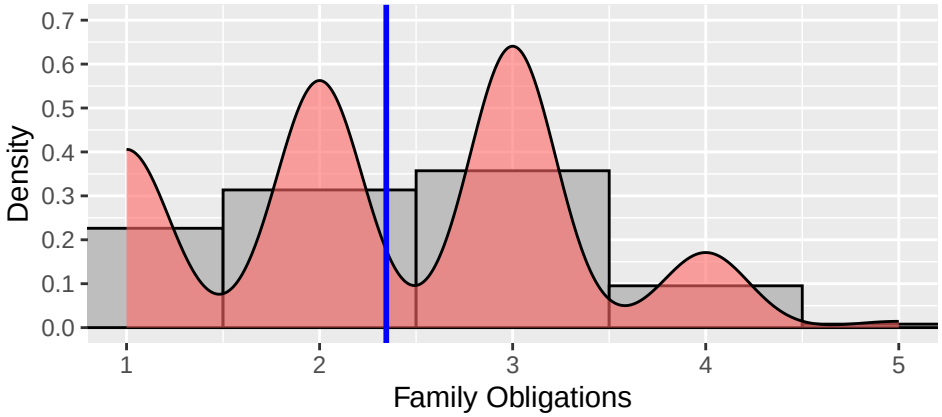

Figure 4: Histogram of Family Obligations Among Enrolled Applicants (After Listwise Deletion)

```{r}
#| label: fig-Barrier-WR
#| fig-cap: Histogram of Family Obligations Among Enrolled Applicants (After
#|          Listwise Deletion)
#| fig-width: 5
#| fig-height: 2.5

ggplot(data = Enrolled_Applicants_CD, aes(x = Barrier_WR)) +
  geom_histogram(aes(y = after_stat(density)), binwidth = 1, color = "black",
                 fill = "grey") +
  geom_density(alpha = .6, fill = "#FF6666") +
  geom_vline(aes(xintercept = mean(Barrier_WR, na.rm = TRUE)), color = "blue",
             linetype = "solid", linewidth = 1) +
  coord_cartesian(xlim = c(1, 5), ylim = c(0, 0.80)) +
  scale_x_continuous(breaks = seq(from = 1, to = 5, by = 1)) +
  scale_y_continuous(breaks = seq(from = 0, to = 0.80, by = 0.10)) +
  labs(subtitle = "",
       x = "Work Responsibilities",
       y = "Density")
```

Figure 5: Histogram of Family Obligations Among Enrolled Applicants (After Listwise Deletion)

Table 17 shows the Pearson correlations between the continuous predictors we will use in our stopping-ratio models.

```
```{r}
#| label: tbl-continuous-corr
#| tbl-cap: Pearson Correlation Matrix for Continuous Predictor Variables Among
#|          Enrolled Applicants (After Listwise Deletion)

N_used <- nrow(Enrolled_Applicants_CD)
N_omit <- nrow(Enrolled_Applicants) - N_used

FN <- paste0("N = ", N_used, " after listwise deletion of missing data (",
             N_omit, " observations).")

Enrolled_Applicants_CD %>%
  select(ProQOL_BO, ProQOL_CS, ProQOL_STS, Barrier_FO, Barrier_WR) %>%
  cor(x = ., use = "complete.obs", method = "pearson") %>%
  kable(format = "latex", booktabs = TRUE, digits = 2) %>%
  kable_styling() %>%
  footnote(kable_input = ., general = FN, footnote_as_chunk = TRUE)
```
```

Table 17: Pearson Correlation Matrix for Continuous Predictor Variables Among Enrolled Applicants (After Listwise Deletion)

	ProQOL_BO	ProQOL_CS	ProQOL_STS	Barrier_FO	Barrier_WR
ProQOL_BO	1.00	-0.71	0.63	0.15	0.11
ProQOL_CS	-0.71	1.00	-0.37	-0.06	0.02
ProQOL_STS	0.63	-0.37	1.00	0.20	0.22
Barrier_FO	0.15	-0.06	0.20	1.00	0.41
Barrier_WR	0.11	0.02	0.22	0.41	1.00

Note: N = 252 after listwise deletion of missing data (2 observations).

5.5.2 Categorical Predictors

Table 18 and Table 19 respectively show the frequency distributions for `Motivation_NeedSANE` and `Motivation_PersonalConn`, which are both binary variables.

```{r}
#| label: tbl-Motivation-NeedSANE
#| tbl-cap: Motivation_NeedSANE Frequency Distribution Among Enrolled Applicants
#|          (After Listwise Deletion)

# Footnote text.
FN <- paste("Denominator for percents is the frequency sum.")

Enrolled_Applicants_CD %>%
xtabs(~as_factor(Motivation_NeedSANE), addNA = TRUE, data = .) %>%
  addmargins() %>%
  as.data.frame() %>%
  rename(Motivation_NeedSANE = as_factor.Motivation_NeedSANE.) %>%
  mutate(Percent = 100*Freq/nrow(Enrolled_Applicants_CD),
         Cumulative = if_else(Motivation_NeedSANE == "Sum",
                              true = 100,
                              false = cumsum(Percent))) %>%
  kable(format = "latex", booktabs = TRUE, digits = 1,
        col.names = c("Motivation_NeedSANE", "Freq", "%", "Cumulative %")) %>%
  kable_styling() %>%
  row_spec(row = 3, italic = TRUE) %>%
  footnote(kable_input = ., general = FN, footnote_as_chunk = TRUE,
           threeparttable = TRUE)
```

Table 18: Motivation_NeedSANE Frequency Distribution Among Enrolled Applicants (After Listwise Deletion)

Motivation_NeedSANE	Freq	%	Cumulative %
No	59	23.4	23.4
Yes	193	76.6	100.0
Sum	*252*	*100.0*	*100.0*

Note: Denominator for percents is the frequency sum.

```{r}
#| label: tbl-Motivation-PersonalConn
#| tbl-cap: Motivation_PersonalConn Frequency Distribution Among Enrolled
#|          Applicants (After Listwise Deletion)

# Footnote text.
FN <- paste("Denominator for percents is the frequency sum.")

Enrolled_Applicants_CD %>%
  xtabs(~as_factor(Motivation_PersonalConn), addNA = TRUE, data = .) %>%
  addmargins() %>%
  as.data.frame() %>%
  rename(Motivation_PersonalConn = as_factor.Motivation_PersonalConn.) %>%
  mutate(Percent = 100*Freq/nrow(Enrolled_Applicants_CD),
         Cumulative = if_else(Motivation_PersonalConn == "Sum",
                              true = 100,
                              false = cumsum(Percent))) %>%
  kable(format = "latex", booktabs = TRUE, digits = 1,
        col.names = c("Motivation_PersonalConn", "Freq", "%", "Cumulative %")) %>%
  kable_styling() %>%
  row_spec(row = 3, italic = TRUE) %>%
  footnote(kable_input = ., general = FN, footnote_as_chunk = TRUE,
           threeparttable = TRUE)
```

Table 19: Motivation_PersonalConn Frequency Distribution Among Enrolled Applicants (After Listwise Deletion)

Motivation_PersonalConn	Freq	%	Cumulative %
No	141	56	56
Yes	111	44	100
Sum	*252*	*100*	*100*

Note: Denominator for percents is the frequency sum.

Table 20 shows the frequency distribution for the setting where enrolled applicants practice in its

original coding. It shows that the Tribal category is too small to model as a separate level, so we need to combine it with another category.

```{r}
#| label: tbl-Setting-original
#| tbl-cap: Original Practice Setting Frequency Distribution Among Enrolled
#|          Applicants (After Listwise Deletion)

# Footnote text.
FN <- paste("Denominator for percents is the frequency sum.")

Enrolled_Applicants_CD %>%
  xtabs(~as_factor(practicset_dem_num), addNA = TRUE, data = .) %>%
  addmargins() %>%
  as.data.frame() %>%
  rename(practicset_dem_num = as_factor.practicset_dem_num.) %>%
  mutate(Percent = 100*Freq/nrow(Enrolled_Applicants_CD),
         Cumulative = if_else(practicset_dem_num == "Sum",
                              true = 100,
                              false = cumsum(Percent))) %>%
  kable(format = "latex", booktabs = TRUE, digits = 1,
        col.names = c("practicset_dem_num", "Freq", "%", "Cumulative %")) %>%
  kable_styling() %>%
  row_spec(row = 5, italic = TRUE) %>%
  footnote(kable_input = ., general = FN, footnote_as_chunk = TRUE,
           threeparttable = TRUE)
```

Table 20: Original Practice Setting Frequency Distribution Among Enrolled Applicants (After Listwise Deletion)

practicset_dem_num	Freq	%	Cumulative %
Urban	88	34.9	34.9
Rural	77	30.6	65.5
Suburban	85	33.7	99.2
Tribal	2	0.8	100.0
Sum	*252*	*100.0*	*100.0*

Note: Denominator for percents is the frequency sum.

Table 21 shows the updated frequency distribution for setting after collapsing the Rural and Tribal categories together. This glosses over important cultural differences between those settings, but these two settings are similar in size.

```{r}
#| label: tbl-Setting-recoded
#| tbl-cap: Recoded Practice Setting Frequency Distribution Among Enrolled
#|          Applicants (After Listwise Deletion)

# Footnote text.
FN <- paste("Denominator for percents is the frequency sum.")

Enrolled_Applicants_CD %>%
  xtabs(~as_factor(Setting), addNA = TRUE, data = .) %>%
  addmargins() %>%
  as.data.frame() %>%
  rename(Setting = as_factor.Setting.) %>%
  mutate(Percent = 100*Freq/nrow(Enrolled_Applicants_CD),
         Cumulative = if_else(Setting == "Sum",
                              true = 100,
                              false = cumsum(Percent))) %>%
  kable(format = "latex", booktabs = TRUE, digits = 1,
        col.names = c("Setting (Recoded)", "Freq", "%", "Cumulative %")) %>%
  kable_styling() %>%
  row_spec(row = 4, italic = TRUE) %>%
  footnote(kable_input = ., general = FN, footnote_as_chunk = TRUE,
           threeparttable = TRUE)
```

Table 21: Recoded Practice Setting Frequency Distribution Among Enrolled Applicants (After Listwise Deletion)

Setting (Recoded)	Freq	%	Cumulative %
Urban	88	34.9	34.9
Rural/Tribal	79	31.3	66.3
Suburban	85	33.7	100.0
Sum	*252*	*100.0*	*100.0*

Note:
Denominator for percents is the frequency sum.

6 Summaries of Thresholds Data

Figure 6 shows the final set of stages and thresholds between them at which attrition from the training program could occur. The arrows associated with each threshold are labeled according to how the outcome variable is coded on the corresponding person-threshold record, depending on whether the person stopped participating at the current stage or moved on to the next stage.

```
```{dot}
//| label: fig-Stages
//| fig-cap: Stages, Thresholds (T1-T3), and Stopping Ratios (SR1-SR3) in the
//|          SANE Training Program. CSW, clinical skills workshop; DT, didactic
//|          training.
//| fig-width: 5
//| fig-height: 2.5

digraph StagesModeled {

graph [rankdir="LR"];

node [shape = "box", style= "filled", fillcolor = "Gray90", fontsize = "7"];
A1 [label = "Attrited\nBefore DT\nSR1"]
A2 [label = "Attrited\nDuring DT\nSR2"]
A3 [label = "Attrited\nBefore/During CSW\nSR3"]

S1 [label = "Stage 1\nEnrolled"]
S2 [label = "Stage 2\nStarted DT"]
S3 [label = "Stage 3\nFinished DT"]
S4 [label = "Stage 4\nFinished CSW"]

edge [fontsize = "7", arrowsize = 0.5];

S1 -> A1 [label = "T1\nAttrit = 1"]
S2 -> A2 [label = "T2\nAttrit = 1"]
S3 -> A3 [label = "T3\nAttrit = 1"]

S1 -> S2 [label = "Attrit = 0"]
S2 -> S3 [label = "Attrit = 0"]
S3 -> S4 [label = "Attrit = 0"]

}
```
```

> **Tip**
>
> The `Thresholds` dataset contains only data from eligible applicants remaining after listwise deletion of applicants with missing data on the variables used for modeling.

```
```{r}
#| label: tbl-Threshold-Attrit
#| tbl-cap: Threshold by Attrit Contingency Table (After Listwise Deletion)

# Footnote text.
FN <- paste("Denominator for percents is row N.",
            "CSW, clinical skills workshop; DT, didactic training; N, number",
            "of participants who reached the threshold.")
```
```

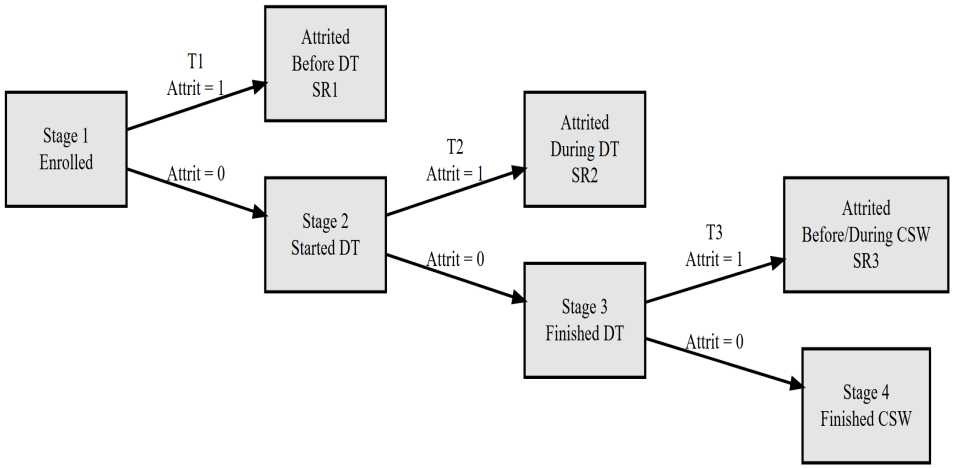

Figure 6: Stages, Thresholds (T1-T3), and Stopping Ratios (SR1-SR3) in the SANE Training Program. CSW, clinical skills workshop; DT, didactic training.

```
Thresholds %>%
  mutate(Threshold = as_factor(Threshold),
         Attrit = as_factor(Attrit)) %>%
  group_by(Threshold, Attrit) %>%
  summarise(N = n()) %>%
  pivot_wider(names_from = Attrit, values_from = N) %>%
  mutate(N = No + Yes,
         Percent = 100*Yes/N) %>%
  select(Threshold, N, No, Yes, Percent) %>%
  kable(format = "latex", booktabs = TRUE, digits = 1,
        col.names= c("Threshold", "N", "No (0)", "Yes (1)", "%")) %>%
  kable_styling() %>%
  add_header_above(header = c(" " = 2, "Attrit" = 3)) %>%
  footnote(kable_input = ., general = FN, footnote_as_chunk = TRUE,
           threeparttable = TRUE)
```

```
`summarise()` has regrouped the output.
i Summaries were computed grouped by Threshold and Attrit.
i Output is grouped by Threshold.
i Use `summarise(.groups = "drop_last")` to silence this message.
i Use `summarise(.by = c(Threshold, Attrit))` for per-operation grouping
  (`?dplyr::dplyr_by`) instead.
```

Table 22: Threshold by Attrit Contingency Table (After Listwise Deletion)

		Attrit		
Threshold	N	No (0)	Yes (1)	%
T1 Before DT	252	235	17	6.7
T2 During DT	235	174	61	26.0
T3 Before/During CSW	174	157	17	9.8

Note: Denominator for percents is row N. CSW, clinical skills workshop; DT, didactic training; N, number of participants who reached the threshold.

7 References

Ben-Shachar, M. S., Lüdecke, D., & Makowski, D. (2020). Effectsize: Estimation of effect size indices and standardized parameters. *Journal of Open Source Software*, *5*(56), 2815. https://doi.org/10.21105/joss.02815

Ben-Shachar, M. S., Patil, I., Thériault, R., Wiernik, B. M., & Lüdecke, D. (2023). Phi, fei, fo, fum: Effect sizes for categorical data that use the chi-squared statistic. *Mathematics*, *11*(9), 1982. https://doi.org/10.3390/math11091982

Dontje, K., & Campbell, R. (07/01/2021–06/30/2025). *Increasing access, recruitment, and retention of sexual assault nurse examiners in rural michigan* (Grant No. T96HP42059). Health Resources and Services Administration.

Patterson, D., & Resko, S. (2015). Predictors of attrition for a sexual assault forensic examiner (SAFE) blended learning training program. *Journal of Continuing Education in the Health Professions*, *35*(2), 99–108. https://doi.org/10.1002/chp.21277

Pierce, S. J. (2026). *SANETPA: Research compendium for a study of sexual assault nurse examiner training program attrition* (Version 1.0.1) [Reproducible Research Materials and Computer Program, R Package, Public Repository]. GitHub. https://github.com/sjpierce/SANETPA. Zenodo. https://doi.org/10.5281/zenodo.18643254

Stamm, B. H. (2010). *The concise ProQOL manual*. The ProQOL.org. https://ProQOL.org

8 Software Information

This section documents information that is important for reproducibility. Most users will not need to read it. It is primarily here for use by the statistician on the team if we need to troubleshoot reproducibility issues because someone else is unable to get the same results from the same code. Start by checking for differences in package versions.

We used R as our main computing environment and Quarto scripts to enhance reproducibility. We used RStudio as the editor to interface with R and Quarto.

- Software chain: **qmd file > RStudio > Quarto > R > knitr > md file > Pandoc > tex file > TinyTeX > PDF file**.
- Source file: `Descriptive_Analyses.qmd`
- Output file: `Descriptive_Analyses_2026-02-14.pdf`
- Quarto 1.8.27 runs `*.qmd` files through R and knitr to produce `*.md` markdown files.
- Pandoc 3.6.3 converts markdown files (`*.md`) to other formats, including LaTeX (`*.tex`) and HTML (`*.html`) among others.
- TinyTeX compiles LaTeX files (`*.tex`) into PDF files. It should be viable to use MiKTeX or another LaTeX distribution instead.

8.1 Versions

This document was generated using the following computational environment and dependencies:

```
# Check and report whether we used TinyTex or other LaTeX software.
which_latex()
```

```
is_tinytex = TRUE. We used TeX Live 2025 (TinyTeX) with tlmgr 2025-11-06.
```

```
tlmgr revision 76773 (2025-11-06 20:43:29 +0100)
tlmgr using installation: C:/Users/pierces1/AppData/Roaming/TinyTeX
TeX Live (https://tug.org/texlive) version 2025
```

```
# Get R and R package version numbers in use.
devtools::session_info()
```

```
Warning in system2("quarto", "-V", stdout = TRUE, env = paste0("TMPDIR=", :
running command '"quarto"
TMPDIR=C:/Users/pierces1/AppData/Local/Temp/RtmpG4LrMQ/file651c7a075c7 -V' had
status 1
```

```
- Session info -------------------------------------------------------------
 setting  value
 version  R version 4.5.2 (2025-10-31 ucrt)
 os       Windows 11 x64 (build 26100)
 system   x86_64, mingw32
 ui       RTerm
 language (EN)
 collate  English_United States.utf8
 ctype    English_United States.utf8
 tz       America/New_York
 date     2026-02-14
 pandoc   3.6.3 @ C:/Program Files/RStudio/resources/app/bin/quarto/bin/tools/ (via rmarkdown)
 quarto   NA @ C:\\PROGRA~1\\Quarto\\bin\\quarto.exe

- Packages -----------------------------------------------------------------
 package      * version  date (UTC) lib source
 assertthat     0.2.1    2019-03-21 [1] CRAN (R 4.5.0)
 backports      1.5.0    2024-05-23 [1] CRAN (R 4.5.0)
 bayestestR     0.17.0   2025-08-29 [1] CRAN (R 4.5.1)
 broom          1.0.12   2026-01-27 [1] CRAN (R 4.5.2)
 cachem         1.1.0    2024-05-16 [1] CRAN (R 4.5.0)
 cli            3.6.5    2025-04-23 [1] CRAN (R 4.5.0)
 coda           0.19-4.1 2024-01-31 [1] CRAN (R 4.5.0)
 codetools      0.2-20   2024-03-31 [1] CRAN (R 4.5.0)
 datawizard     1.3.0    2025-10-11 [1] CRAN (R 4.5.1)
 devtools     * 2.4.6    2025-10-03 [1] CRAN (R 4.5.1)
 digest         0.6.39   2025-11-19 [1] CRAN (R 4.5.2)
 dplyr        * 1.2.0    2026-02-03 [1] CRAN (R 4.5.2)
 effectsize   * 1.0.1    2025-05-27 [1] CRAN (R 4.5.0)
 ellipsis       0.3.2    2021-04-29 [1] CRAN (R 4.5.0)
 emmeans        2.0.1    2025-12-16 [1] CRAN (R 4.5.2)
 estimability   1.5.1    2024-05-12 [1] CRAN (R 4.5.0)
 evaluate       1.0.5    2025-08-27 [1] CRAN (R 4.5.1)
 farver         2.1.2    2024-05-13 [1] CRAN (R 4.5.0)
 fastmap        1.2.0    2024-05-15 [1] CRAN (R 4.5.0)
 forcats      * 1.0.1    2025-09-25 [1] CRAN (R 4.5.1)
 fs             1.6.6    2025-04-12 [1] CRAN (R 4.5.0)
 generics       0.1.4    2025-05-09 [1] CRAN (R 4.5.0)
 ggplot2      * 4.0.2    2026-02-03 [1] CRAN (R 4.5.2)
 git2r          0.36.2   2025-03-29 [1] CRAN (R 4.5.0)
 glue           1.8.0    2024-09-30 [1] CRAN (R 4.5.0)
 gtable         0.3.6    2024-10-25 [1] CRAN (R 4.5.0)
 haven        * 2.5.5    2025-05-30 [1] CRAN (R 4.5.0)
 here         * 1.0.2    2025-09-15 [1] CRAN (R 4.5.1)
 hms            1.1.4    2025-10-17 [1] CRAN (R 4.5.1)
 htmltools      0.5.9    2025-12-04 [1] CRAN (R 4.5.2)
 httr           1.4.8    2026-02-13 [1] CRAN (R 4.5.2)
 insight        1.4.6    2026-02-04 [1] CRAN (R 4.5.2)
 janitor      * 2.2.1    2024-12-22 [1] CRAN (R 4.5.0)
 jsonlite       2.0.0    2025-03-27 [1] CRAN (R 4.5.0)
 kableExtra   * 1.4.0    2024-01-24 [1] CRAN (R 4.5.0)
 knitr        * 1.51     2025-12-20 [1] CRAN (R 4.5.2)
 labelled     * 2.16.0   2025-10-22 [1] CRAN (R 4.5.1)
 later          1.4.5    2026-01-08 [1] CRAN (R 4.5.2)
 lattice        0.22-9   2026-02-09 [1] CRAN (R 4.5.2)
 lifecycle      1.0.5    2026-01-08 [1] CRAN (R 4.5.2)
 lubridate    * 1.9.5    2026-02-04 [1] CRAN (R 4.5.2)
 magrittr       2.0.4    2025-09-12 [1] CRAN (R 4.5.1)
 MASS           7.3-65   2025-02-28 [1] CRAN (R 4.5.0)
 Matrix         1.7-4    2025-08-28 [1] CRAN (R 4.5.1)
 MBESS          4.9.42   2026-01-08 [1] CRAN (R 4.5.2)
 memoise        2.0.1    2021-11-26 [1] CRAN (R 4.5.0)
 mnormt         2.1.2    2026-01-27 [1] CRAN (R 4.5.2)
 multcomp       1.4-29   2025-10-20 [1] CRAN (R 4.5.1)
 mvtnorm        1.3-3    2025-01-10 [1] CRAN (R 4.5.0)
 nlme           3.1-168  2025-03-31 [1] CRAN (R 4.5.0)
 otel           0.2.0    2025-08-29 [1] CRAN (R 4.5.1)
 parameters     0.28.3   2025-11-25 [1] CRAN (R 4.5.2)
 pbivnorm       0.6.0    2015-01-23 [1] CRAN (R 4.5.0)
 piercer      * 0.23.0   2025-09-07 [1] Github (sjpierce/piercer@7e53e10)
 pillar         1.11.1   2025-09-17 [1] CRAN (R 4.5.1)
 pkgbuild       1.4.8    2025-05-26 [1] CRAN (R 4.5.0)
```

```
pkgconfig      2.0.3    2019-09-22 [1] CRAN (R 4.5.0)
pkgload        1.5.0    2026-02-03 [1] CRAN (R 4.5.2)
pROC           1.19.0.1 2025-07-31 [1] CRAN (R 4.5.1)
processx       3.8.6    2025-02-21 [1] CRAN (R 4.5.0)
PropCIs      * 0.3-0    2018-02-23 [1] CRAN (R 4.5.0)
ps             1.9.1    2025-04-12 [1] CRAN (R 4.5.0)
psych        * 2.6.1    2026-02-03 [1] CRAN (R 4.5.2)
purrr        * 1.2.1    2026-01-09 [1] CRAN (R 4.5.2)
quarto       * 1.5.1    2025-09-04 [1] CRAN (R 4.5.1)
R6             2.6.1    2025-02-15 [1] CRAN (R 4.5.0)
RColorBrewer   1.1-3    2022-04-03 [1] CRAN (R 4.5.0)
Rcpp           1.1.1    2026-01-10 [1] CRAN (R 4.5.2)
readr        * 2.1.6    2025-11-14 [1] CRAN (R 4.5.2)
remotes        2.5.0    2024-03-17 [1] CRAN (R 4.5.0)
rlang          1.1.7    2026-01-09 [1] CRAN (R 4.5.2)
rmarkdown    * 2.30     2025-09-28 [1] CRAN (R 4.5.1)
rprojroot      2.1.1    2025-08-26 [1] CRAN (R 4.5.1)
rstudioapi     0.18.0   2026-01-16 [1] CRAN (R 4.5.2)
S7             0.2.1    2025-11-14 [1] CRAN (R 4.5.2)
sandwich       3.1-1    2024-09-15 [1] CRAN (R 4.5.0)
SANETPA      * 1.0.1    2026-02-14 [1] Github (sjpierce/SANETPA@482345e)
scales         1.4.0    2025-04-24 [1] CRAN (R 4.5.0)
sessioninfo    1.2.3    2025-02-05 [1] CRAN (R 4.5.0)
snakecase      0.11.1   2023-08-27 [1] CRAN (R 4.5.0)
stringi        1.8.7    2025-03-27 [1] CRAN (R 4.5.0)
stringr      * 1.6.0    2025-11-04 [1] CRAN (R 4.5.2)
survival       3.8-6    2026-01-16 [1] CRAN (R 4.5.2)
svglite        2.2.2    2025-10-21 [1] CRAN (R 4.5.1)
systemfonts    1.3.1    2025-10-01 [1] CRAN (R 4.5.1)
texreg         1.39.5   2025-12-22 [1] CRAN (R 4.5.2)
textshaping    1.0.4    2025-10-10 [1] CRAN (R 4.5.1)
TH.data        1.1-5    2025-11-17 [1] CRAN (R 4.5.2)
tibble       * 3.3.1    2026-01-11 [1] CRAN (R 4.5.2)
tidyr        * 1.3.2    2025-12-19 [1] CRAN (R 4.5.2)
tidyselect     1.2.1    2024-03-11 [1] CRAN (R 4.5.0)
tidyverse    * 2.0.0    2023-02-22 [1] CRAN (R 4.5.0)
timechange     0.4.0    2026-01-29 [1] CRAN (R 4.5.2)
tinytex        0.58     2025-11-19 [1] CRAN (R 4.5.2)
tzdb           0.5.0    2025-03-15 [1] CRAN (R 4.5.0)
usethis      * 3.2.1    2025-09-06 [1] CRAN (R 4.5.1)
vctrs          0.7.1    2026-01-23 [1] CRAN (R 4.5.2)
viridisLite    0.4.3    2026-02-04 [1] CRAN (R 4.5.2)
withr          3.0.2    2024-10-28 [1] CRAN (R 4.5.0)
xfun           0.56     2026-01-18 [1] CRAN (R 4.5.2)
xml2           1.5.2    2026-01-17 [1] CRAN (R 4.5.2)
xtable         1.8-4    2019-04-21 [1] CRAN (R 4.5.0)
yaml           2.3.12   2025-12-10 [1] CRAN (R 4.5.2)
zoo            1.8-15   2025-12-15 [1] CRAN (R 4.5.2)

[1] C:/Users/pierces1/AppData/Local/R/win-library/4.5
[2] C:/Program Files/R/R-4.5.2/library
* -- Packages attached to the search path.

------------------------------------------------------------------------------
```

8.2 Git Details

The current Git commit details and status are:

```
git_report()
```

```
Local:    main P:/Consulting/Cases_1600-1799/C1788/SANETPA
Remote:   main @ origin (https://github.com/sjpierce/SANETPA.git)
Head:     [482345e] 2026-02-14: Update version number.



Untracked files:
    Untracked:  scripts/Descriptive_Analyses.rmarkdown
    Untracked:  scripts/Descriptive_Analyses_files/
    Untracked:  scripts/Production_Run.rmarkdown
```

This is useful because it tells us exactly which commit in the Git history we would need to be using to make sure we are running the exact same code. Sometimes another person is not using the most current code, or has changed the code in some way since it was last committed.

Tip

- Untracked files are files located in the repository that Git has not been told to entirely ignore, but have also not been committed into the version history.
- Unstaged changes to files indicate that some of the contents have been modified since the last time the file was committed to Git. In production runs, we want the Git output to not show any unstaged changes to key files!