Import and Manage 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 imports data from from an SPSS data file provided by the investigators, does some data cleaning and management, then saves out an R data file that will be used by other scripts in this compendium.

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(tidyverse)      # for map_dfr(), map_chr(), rowid_to_column(),
```

```
-- 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
                        # rownames_to_column()
library(haven)          # for read_SPSS()
library(labelled)       # for var_label()
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(kableExtra)     # for kable_styling
library(piercer)        # for cvv_missingness(), file_details(), git_report(),
                        # var_missingness(), which_latex()
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 Data Management

3.1 Read SPSS Data into `Applicants_Raw` Dataset

The raw data provided by the investigators is in an SPSS data file. This section imports it into an R data frame/tibble. The raw data is de-identified to preserve participant privacy and protect confidentiality.

The `ID` variable in the data has no intrinsic meaning and could only be linked to actual participant names via the key retained by the principal investigators. Per email communication that accompanied delivery of the data, it has missing data because "The ids were not generated until the end of application 1- so they would have had to get through all of the questions for part a first" (A. Ashley, personal communication, May 18, 2025).

> **Note**
>
> I received an updated data file called `All_group_1_2025-06-03.sav` that fixed some inaccurate data values detected by a previous iteration of the descriptive analyses. It replaced the previous data file and included all applicants (even those who were ineligible). However, I later edited that file to produce the final raw data file because `All_group_1_2025-06-03.sav` had some data entry errors that created logical discrepancies between binary `Enrolled` and ordinal `stg_reached` variables. The final raw data file is `All_group_1_2025-11-19.sav`. The date in the file name is when I fixed the discrepancies.

Table 1 shows meta-data about the raw data file we will import.

```
```{r}
#| label: tbl-raw-data-file
#| tbl-cap: "Meta-Data About Raw Data File"

# Store path to raw data file.
RawFile <- here("scripts/extdata/All_group_1_2025-11-19.sav")

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

Table 1: Meta-Data About Raw Data File

File Name	Size	Last Modified
All_group_1_2025-11-19.sav	145K	2025-11-19 16:46:23

Next, we actually read that SPSS file into a data frame called `Applicants_Raw`. That will be an intermediate form of the data because we need to do some data management before it is ready for use in descriptive and inferential analyses.

```
```{r}
#| label: import-data

Applicants_Raw <- read_spss(RawFile)
```
```

The imported `Applicants_Raw` data frame has 497 rows and 149 variables. Table 2 shows a list of the variable positions, names, classes, and labels prior to any data management.

```
```{r}
#| label: tbl-Applicants-Raw
#| tbl-cap: "Applicants_Raw: Variable Names and Classes."

ApplicantsVars <- list(Applicants_Raw) %>%
  map_dfr(~ tibble(Name = names(.x),
                   Class = map_chr(.x, all_classes),
                   Label = var_label(.x))) %>%
  rowid_to_column(., "Position")

kable(ApplicantsVars, format = "latex", booktabs = TRUE, longtable = TRUE) %>%
  kable_styling(font_size = 7, latex_options = c("repeat_header")) %>%
  column_spec(column = 3, width = "3cm") %>%
  column_spec(column = 4, width = "6.5cm")
```
```

Table 2: Applicants_Raw: Variable Names and Classes.

Position	Name	Class	Label
1	ID	character	NULL
2	educ_dem_num	haven_labelled, vctrs_vctr, double	What is your highest level of education?
3	license_num	haven_labelled, vctrs_vctr, double	What is your licensure or certification for practice?
4	practicset_dem_num	haven_labelled, vctrs_vctr, double	How would you describe the primary setting you practice nursing in?
5	employment_dem_num	haven_labelled, vctrs_vctr, double	What is your current employment status?
6	employednurse_dem_num	haven_labelled, vctrs_vctr, double	If currently employed, are you working as a nurse?
7	nursexperience_years	numeric	How many years have you been practicing as a nurse (since completing your initial nursing program)?
8	recruitment_num	haven_labelled, vctrs_vctr, double	How did you learn about this opportunity? - Selected Choice
9	prevSANEtrain_num	haven_labelled, vctrs_vctr, double	Previous SANE training (numeric)
10	trained1year_num	numeric	Year training was completed (numeric)
11	trained2year	numeric	What year(s) did you complete your didactic and clinical skills SANE training?
12	Disability_status	haven_labelled, vctrs_vctr, double	Do you have a disability ?
13	sex_dem_num	haven_labelled, vctrs_vctr, double	What was your sex assigned at birth?
14	pronoun_dem_num	haven_labelled, vctrs_vctr, double	What are your pronouns? - Selected Choice
15	age_num	numeric	How old are you?
16	race_ethnicity_num	haven_labelled, vctrs_vctr, double	How would you describe your race/ethnicity? (Select all that apply) - Selected Choice
17	income_dem_num	haven_labelled, vctrs_vctr, double	Income level (numeric)
18	LGBTQ_num	haven_labelled, vctrs_vctr, double	Do you identify as a member of the LGBTQIA2S+ community?
19	npq1	haven_labelled, vctrs_vctr, double	I am happy.
20	npq2	haven_labelled, vctrs_vctr, double	I am preoccupied with more than one person I care for as a nurse
21	npq3	haven_labelled, vctrs_vctr, double	I get satisfaction from being able to care for people.
22	npq4	haven_labelled, vctrs_vctr, double	I feel connected to others.
23	npq5	haven_labelled, vctrs_vctr, double	I jump or am startled by unexpected sounds.
24	npq6	haven_labelled, vctrs_vctr, double	I feel invigorated after working with those I care for.
25	npq7	haven_labelled, vctrs_vctr, double	I find it difficult to separate my personal life from my life as a nurse.
26	npq8	haven_labelled, vctrs_vctr, double	I am not as productive at work because I am losing sleep over traumatic experiences of a person I care for as a nurse.
27	npq9	haven_labelled, vctrs_vctr, double	I think that I might have been affected by the traumatic stress of those I care for as a nurse.
28	npq10	haven_labelled, vctrs_vctr, double	I feel trapped by my job as a nurse.
29	npq11	haven_labelled, vctrs_vctr, double	Because of my nursing, I have felt ”on edge” about various things.
30	npq12	haven_labelled, vctrs_vctr, double	I like my work as a nurse.
31	npq13	haven_labelled, vctrs_vctr, double	I feel depressed because of the traumatic experiences of the people I care for as a nurse.
32	npq14	haven_labelled, vctrs_vctr, double	I feel as though I am experiencing the trauma of someone I have cared for as a nurse.
33	npq15	haven_labelled, vctrs_vctr, double	I have beliefs that sustain me.

(continued)

Position	Name	Class	Label
34	npq16	haven_labelled, vctrs_vctr, double	I am pleased with how I am able to keep up with nursing techniques and protocols.
35	npq17	haven_labelled, vctrs_vctr, double	I am the person I always wanted to be.
36	npq18	haven_labelled, vctrs_vctr, double	My work makes me feel satisfied.
37	npq19	haven_labelled, vctrs_vctr, double	I feel worn out because of my work as a nurse.
38	npq20	haven_labelled, vctrs_vctr, double	I have happy thoughts and feelings about those I care for as a nurse and how I could help them.
39	npq21	haven_labelled, vctrs_vctr, double	I feel overwhelmed because my case work load seems endless.
40	npq22	haven_labelled, vctrs_vctr, double	I believe I can make a difference through my work.
41	npq23	haven_labelled, vctrs_vctr, double	I avoid certain activities or situations because they remind me of frightening experiences of the people I care for as a nurse.
42	npq24	haven_labelled, vctrs_vctr, double	I am proud of what I can do to nurse.
43	npq25	haven_labelled, vctrs_vctr, double	As a result of my nursing, I have intrusive, frightening thoughts.
44	npq26	haven_labelled, vctrs_vctr, double	I feel ”bogged down” by the system.
45	npq27	haven_labelled, vctrs_vctr, double	I have thoughts that I am a ”success” as a nurse.
46	npq28	haven_labelled, vctrs_vctr, double	I can’t recall important parts of my work with trauma victims.
47	npq29	haven_labelled, vctrs_vctr, double	I am a very caring person.
48	npq30	haven_labelled, vctrs_vctr, double	I am happy that I chose to do this work.
49	npq1r	haven_labelled, vctrs_vctr, double	recoded values
50	npq4r	haven_labelled, vctrs_vctr, double	recoded
51	npq15r	haven_labelled, vctrs_vctr, double	recoded
52	npq17r	haven_labelled, vctrs_vctr, double	recoded
53	npq29r	haven_labelled, vctrs_vctr, double	recoded
54	Motivation__ProfGrowth	haven_labelled, vctrs_vctr, double	What are your primary motivations for seeking SANE training (Select all that apply)? - Selected Choice
55	Motivation__PersonalConn	haven_labelled, vctrs_vctr, double	What are your primary motivations for seeking SANE training (Select all that apply)? - Selected Choice
56	Motivation__NeedSANE	haven_labelled, vctrs_vctr, double	What are your primary motivations for seeking SANE training (Select all that apply)? - Selected Choice
57	Motivation__PersonalLearn	haven_labelled, vctrs_vctr, double	What are your primary motivations for seeking SANE training (Select all that apply)? - Selected Choice
58	Motivation__Humanitarian	haven_labelled, vctrs_vctr, double	What are your primary motivations for seeking SANE training (Select all that apply)? - Selected Choice
59	Motivation__EmpRequire	haven_labelled, vctrs_vctr, double	What are your primary motivations for seeking SANE training (Select all that apply)? - Selected Choice
60	Motivation__Other	haven_labelled, vctrs_vctr, double	What are your primary motivations for seeking SANE training (Select all that apply)? - Selected Choice
61	PlansFollowingTraining__Num	haven_labelled, vctrs_vctr, double	Which best describes your plans following the training?
62	train__goal__fac__bar__1__numeric	haven_labelled, vctrs_vctr, double	I will try to learn as much as I can from this course
63	train__goal__fac__bar__2__numeric	haven_labelled, vctrs_vctr, double	I am interested in learning this training material
64	train__goal__fac__bar__3__numeric	haven_labelled, vctrs_vctr, double	My supervisor is supportive of me attending this training
65	train__goal__fac__bar__4__numeric	haven_labelled, vctrs_vctr, double	I am willing to invest effort to improve skills and competencies related to becoming a forensic nurse

(continued)

Position	Name	Class	Label
66	train_goal_fac_bar_5_numeric	haven_labelled, vctrs_vctr, double	I have time in my schedule to focus on this course
67	train_goal_fac_bar_6_numeric	haven_labelled, vctrs_vctr, double	I will get more from this training than most people
68	train_goal_fac_bar_7_numeric	haven_labelled, vctrs_vctr, double	I am motivated to learn the skills needed to be a forensic nurse examiner
69	train_goal_fac_bar_8_numeric	haven_labelled, vctrs_vctr, double	I have a lot of work responsibilities
70	train_goal_fac_bar_9_numeric	haven_labelled, vctrs_vctr, double	I am willing to exert considerable effort in this training in order to improve my skills
71	train_goal_fac_bar_10_numeric	haven_labelled, vctrs_vctr, double	I have a lot of family obligations right now
72	techbasic1_numeric	haven_labelled, vctrs_vctr, double	Have you ever taken an online course before?
73	techbasic2_numeric	haven_labelled, vctrs_vctr, double	Do you have regular access to a computer (e.g., desktop, laptop, tablet) at home?
74	techbasic3_numeric	haven_labelled, vctrs_vctr, double	Do you have reliable access to high-speed internet at home?
75	techfeel1_numeric	haven_labelled, vctrs_vctr, double	I usually get very anxious about using a computer
76	techfeel2_numeric	haven_labelled, vctrs_vctr, double	I'm not the type to do well with computers
77	techfeel3_numeric	haven_labelled, vctrs_vctr, double	I have a lot of self-confidence when it comes to working with computers
78	techfeel4_numeric	haven_labelled, vctrs_vctr, double	I like working with computers
79	techfeel5_numeric	haven_labelled, vctrs_vctr, double	The online course will take less time than a traditional course in a classroom
80	CS	numeric	sum scores compassion from proqol
81	BO	numeric	sum scores burnout from proqol
82	STS	numeric	sum scores sec traumatic stress
83	CSWEvalDate	Date	CSW Eval Date
84	cswsat1_numeric	haven_labelled, vctrs_vctr, double	The pacing of the clinical training was satisfactory.
85	cswsat2_numeric	haven_labelled, vctrs_vctr, double	The clinical concepts were covered_x000D_adequately.
86	cswsat3_numeric	haven_labelled, vctrs_vctr, double	There was enough time to ask_x000D_questions during the clinical training.
87	cswsat4_numeric	haven_labelled, vctrs_vctr, double	There was enough time to practice_x000D_different components of the medical_x000D_forensic exam.
88	cswsat5_numeric	haven_labelled, vctrs_vctr, double	The clinical training met my_x000D_expectations.
89	cswsat6_numeric	haven_labelled, vctrs_vctr, double	I learned a considerable amount during_x000D_the clinical training.
90	cswsat7_numeric	haven_labelled, vctrs_vctr, double	The clinical training helped me feel_x000D_more confident as a forensic examiner.
91	finishedCSW	haven_labelled, vctrs_vctr, double	Enrolled people who completed didactic and CSW
92	finisheddidac	haven_labelled, vctrs_vctr, double	Did the participant complete the didactic?
93	completediddnotcsw	numeric	NULL
94	incompleteCSW	haven_labelled, vctrs_vctr, double	Enrolled people who completed didactic and started CSW
95	didacticsat1_num	haven_labelled, vctrs_vctr, double	There was enough time to learn the material.
96	didacticsat2_num	haven_labelled, vctrs_vctr, double	The topics were covered adequately.
97	didacticsat3_num	haven_labelled, vctrs_vctr, double	The weekly discussion board helped me understand the material better.
98	didacticsat4_num	haven_labelled, vctrs_vctr, double	The online training met my expectations.
99	didacticsat5_num	haven_labelled, vctrs_vctr, double	I learned a considerable amount during the online training.
100	didacticsat6_num	haven_labelled, vctrs_vctr, double	I am satisfied with the amount I have learned in this course.

(continued)

Position	Name	Class	Label
101	didacticsat7_num	haven_labelled, vctrs_vctr, double	Accounting for the materials you learned in this online course, please rate your response to the following question:I have improved knowledge and skills to care for sexual assault patients.
102	DIDEvalDate	Date	Didactic evaluation date.
103	Applied_Date	POSIXct, POSIXt	Date application applied for program.
104	Eligible	haven_labelled, vctrs_vctr, double	Eligible for the program.
105	Enrolled	haven_labelled, vctrs_vctr, double	Whether or not patrticipants were eligible, accepted, and started the program.
106	Enrolled_Date	Date	Date particpant was offered enrollment in the program. All participants in each cohort assisgned same date.
107	Cohort	numeric	Cohort the participant participated in
108	Didactic_Start	Date	Date when participant started the didactic training. All participants in cohort assigned samed date.
109	Didactic_Status	haven_labelled, vctrs_vctr, double	Did the participant complete the didactic training?
110	Didactic_End	Date	Date when participant completed the didactic training.
111	prog_attrited	haven_labelled, vctrs_vctr, double	attrition counts as exiting the training program without completing both didactic and clinical training.
112	exit_date	Date	Participants program exit date.
113	stg_reached	haven_labelled, vctrs_vctr, double	This represents the last stage of training process that the participant actually reached.
114	Mod1_started	haven_labelled, vctrs_vctr, double	Overview of Forensic Nursing and Sexual Violence - did the participant start this module?
115	Mod1_finished	haven_labelled, vctrs_vctr, double	Overview of Forensic Nursing and Sexual Violence - did the participant finish this module?
116	Mod2_started	haven_labelled, vctrs_vctr, double	Victim Responses and Crisis Intervention - did the participant start this module?
117	Mod2_finished	haven_labelled, vctrs_vctr, double	Victim Responses and Crisis Intervention - did the participant finish this module?
118	Mod3_started	haven_labelled, vctrs_vctr, double	Collaborating with Community Agencies - did the participant start this module?
119	Mod3_finished	haven_labelled, vctrs_vctr, double	Collaborating with Community Agencies - did the participant finish this module?
120	Mod4_started	haven_labelled, vctrs_vctr, double	Medical-Forensic History Taking - did the participant start this module?
121	Mod4_finished	haven_labelled, vctrs_vctr, double	Medical-Forensic History Taking - did the participant finish this module?
122	Mod5_started	haven_labelled, vctrs_vctr, double	Observing and Assessing Physical Examination Findings - did the participant start this module?
123	Mod5_finished	haven_labelled, vctrs_vctr, double	Observing and Assessing Physical Examination Findings-did the participant finish this module?
124	Mod6_started	haven_labelled, vctrs_vctr, double	Medical-Forensic Specimen Collection - did the participant start this module?
125	Mod6_finished	haven_labelled, vctrs_vctr, double	Medical-Forensic Specimen Collection - did the participant finish this module?
126	Mod7_started	haven_labelled, vctrs_vctr, double	Medical-Forensic Photography - did the participant start this module?
127	Mod7_finished	haven_labelled, vctrs_vctr, double	Medical-Forensic Photography - did the participant finish this module?
128	Mod8_started	haven_labelled, vctrs_vctr, double	Sexually Transmitted Infection Testing and Prophylaxis - did the participant start this module?
129	Mod8_finished	haven_labelled, vctrs_vctr, double	Sexually Transmitted Infection Testing and Prophylaxis - did the participant finish this module?
130	Mod9_started	haven_labelled, vctrs_vctr, double	Pregnancy Risk Evaluation and Care - did the participant start this module?
131	Mod9_finished	haven_labelled, vctrs_vctr, double	Pregnancy Risk Evaluation and Care- did the participant finish this module?
132	Mod10_started	haven_labelled, vctrs_vctr, double	Medical-Forensic Documentation- did the participant start this module?

(continued)

Position	Name	Class	Label
133	Mod10_finished	haven_labelled, vctrs_vctr, double	Medical-Forensic Documentation- did the participant finish this module?
134	Mod11_started	haven_labelled, vctrs_vctr, double	Discharge and Follow-up Planning - did the participant start this module?
135	Mod11_finished	haven_labelled, vctrs_vctr, double	Discharge and Follow-up Planning- did the participant finish this module?
136	Mod12_started	haven_labelled, vctrs_vctr, double	Legal Considerations and Judicial Proceedings - did the participant start this module?
137	Mod12_finished	haven_labelled, vctrs_vctr, double	Legal Considerations and Judicial Proceedings- did the participant finish this module?
138	Mod1Start	Date	Date when module was scheduled to start for the participants. all participants if started given same date.
139	Mod2Start	Date	Date when module was scheduled to start for the participants. all participants if started given same date.
140	Mod3Start	Date	Date when module was scheduled to start for the participants. all participants if started given same date.
141	Mod4Start	Date	Date when module was scheduled to start for the participants. all participants if started given same date.
142	Mod5Start	Date	Date when module was scheduled to start for the participants. all participants if started given same date.
143	Mod6Start	Date	Date when module was scheduled to start for the participants. all participants if started given same date.
144	Mod7Start	Date	Date when module was scheduled to start for the participants. all participants if started given same date.
145	Mod8Start	Date	Date when module was scheduled to start for the participants. all participants if started given same date.
146	Mod9Start	Date	Date when module was scheduled to start for the participants. all participants if started given same date.
147	Mod10Start	Date	Date when module was scheduled to start for the participants. all participants if started given same date.
148	Mod11Start	Date	Date when module was scheduled to start for the participants. all participants if started given same date.
149	Mod12Start	Date	Date when module was scheduled to start for the participants. all participants if started given same date.

3.2 Check Data for Problems

Tip

Both `finisheddidac` and `Didactic_Status` are binary variables coding whether a person completed the didactic training. The crosstab outputs below show that
the difference between them is that `finisheddidac` has no missing data (`NA`), but a substantial number of applicants have missing `Didactic_Status` paired with `finisheddidac = 0`. All applicants with `finisheddidac = 1` had `Didactic_Status = 1`.

```{r}
#| label: xtab-finisheddidac-Didactic-Status

xtabs(~finisheddidac + Didactic_Status, addNA = TRUE,
      data = Applicants_Raw) %>%
  addmargins()
```

```
             Didactic_Status
finisheddidac   0   1 <NA> Sum
          0    78   0  243 321
          1     0 176    0 176
          Sum  78 176  243 497
```

We can also see that only applicants who were not enrolled have missing values (`NA`) on `Didactic_Status`.

```{r}
#| label: xtab-Enrolled-Didactic-Status

xtabs(~Enrolled + Didactic_Status, addNA = TRUE,
      data = Applicants_Raw) %>%
  addmargins()
```

```
        Didactic_Status
Enrolled   0   1 <NA> Sum
     0     0   0  243 243
     1    78 176    0 254
     Sum  78 176  243 497
```

We will retain both `finisheddidac` and `Didactic_Status` because each may be useful depending on the summary desired.

Table 3 shows the value labels associated with the `prog_attrited` variable. While it is a binary indicator variable (as expected), the coding is reversed compared to what I expected given the variable name. Assigning 0 = completed program and 1 = attrited without completing program would make more sense. We will fix that later after renaming the variable to `Attrited`.

```{r}
#| label: tbl-prog-attrited
#| tbl-cap: Original Value Labels for prog_attrited in Applicants_Raw Data

Applicants_Raw %>%
  select(prog_attrited) %>%
  get_value_labels() %>%
  stack() %>%
  rename(Variable = ind,
         Value = values) %>%
  rownames_to_column(var = "Label") %>%
  relocate(Variable, Value) %>%
  kable(format = "latex", booktabs = TRUE) %>%
  kable_styling()
```

Table 3: Original Value Labels for prog_attrited in Applicants_Raw Data

Variable	Value	Label
prog_attrited	0	didn't complete either or both the didactic and clinical skills workshop
prog_attrited	1	completed both didactic and clinic skills workshop

3.3 Create Applicants Dataset

This section creates the `Applicants` data frame from an updated version of the `Applicants_Raw` data frame. These changes prepare the data for the planned analyses by applying more consistent

and convenient naming conventions, updating value labels, converting categorical variables to factors, and so on.

> **Tip**
>
> The expected uses for the `Applicants` data are:
>
> - Serve as the root source from which other required datasets are derived.
> - Compute eligibility rate and associated confidence interval.

The research team believes that the recruitment materials were sufficiently clear about the program's eligibility criteria, so while it will be useful to report an eligibility rate, they have no substantive interest examining predictors of ineligibility.

The chunk below defines some value labels we will use when creating the `Applicants` dataset from `Applicants_Raw`.

```{r}
#| label: create-value-labels

# Labels for binary variables
Binary_Labels <- c(No = 0, Yes = 1)

# Create vectors of value labels for reverse-coded NPQ items.
reversed_NPQ_value_labels <- c(`Very Often` = 1, Often = 2, Sometimes = 3,
                               Rarely = 4, Never = 5)

# More concise labels for the raw stages.
Stage_Raw_value_labels <- c(`Applied` = 1, `Eligible` = 2, `Enrolled` = 3,
                            `Started DT` = 4, `Finished DT` = 5,
                            `Started CSW` = 6, `Finished CSW` = 7)

# Final stages to be modeled.
Stage_Reached_value_labels <- c(`Enrolled` = 1, `Started DT` = 2,
                                `Finished DT` = 3, `Finished CSW` = 4)

# Final settings to be modeled
Setting_Labels <- c("Urban" = 1, "Rural/Tribal" = 2, "Suburban" = 3)

# Training barriers value labels
Barriers_Labels <-  c(`strongly disagree` = 1, `somewhat disagree` = 2,
                      `neither agree nor disagree` = 3, `somewhat agree` = 4,
                      `strongly agree` = 5)

# Reversed training barriers value labels
Barriers_Reversed_Labels <-  c(`strongly agree` = 1, `somewhat agree` = 2,
                               `neither agree nor disagree` = 3,
                               `somewhat disagree` = 4, `strongly disagree` = 5)
```

In the raw data, `stg_reached` was coded into 7 ordered categories to get the most granular set of categories we thought we might need. Below we replace the imported value labels for those stages with more concise labels. We also rename that variable to `Stage_Raw`, then use it to create the `Stage_Reached` variable by recoding as shown in Table 4.

```{r}
#| label: tbl-stage-recode
#| tbl-cap: Recoding from Stage_Raw to Stage_Reached

tibble(Raw_Value = 1:7,
       Reached_Value = c(NA, NA, 1, 2, 3, 3, 4)) %>%
  set_value_labels(Raw_Value = Stage_Raw_value_labels,
                   Reached_Value = Stage_Reached_value_labels) %>%
  mutate(Raw_Label = as_factor(Raw_Value),
         Reached_Label = as_factor(Reached_Value)) %>%
  select(Raw_Value, Raw_Label, Reached_Value, Reached_Label) %>%
  kable(format = "latex", booktabs = TRUE) %>%
  kable_styling() %>%
  add_header_above(header = c("Stage_Raw" = 2, "Stage_Reached" = 2))
```

Table 4: Recoding from Stage_Raw to Stage_Reached

Stage_Raw		Stage_Reached	
Raw_Value	Raw_Label	Reached_Value	Reached_Label
1	Applied	NA	NA
2	Eligible	NA	NA
3	Enrolled	1	Enrolled
4	Started DT	2	Started DT
5	Finished DT	3	Finished DT
6	Started CSW	3	Finished DT
7	Finished CSW	4	Finished CSW

Only one participant started the clinical skills training without finishing it (`Stage_Raw` = 6). For that stage, the stopping rate would be really close to 0 for the transition between starting and finishing clinical training. That could cause a separation issue (estimation could break down due to insufficient outcome variability at that threshold). Treating that case as having stopped at finishing the didactic training should avoid that problem.

> **!** Important
>
> The raw `practicset_dem_num` variable was coded into 4 categories (Urban, Rural, Suburban, Tribal) but initial descriptive analyses revealed that there are too few observations to use Tribal as a separate category. The study investigators decided that omitting the small number of participants who practice in nursing Tribal settings was inappropriate. Feedback from project stakeholders led them to combine Rural and Tribal categories because those categories represent settings most similar in size. However, we must acknowledge in our paper that this glosses over important cultural differences between Rural and Tribal settings. It is an imperfect way to include data from those participants, but we have no better option.

We want to use measures of the extent to which personal and professional demands on participants' time are a barrier to participation in the training program. We considered using the following three items from the training application form:

- `train_goal_fac_bar_5_numeric`: I have time in my schedule to focus on this course.
- `train_goal_fac_bar_8_numeric`: I have a lot of work responsibilities.
- `train_goal_fac_bar_10_numeric`: I have a lot of family obligations right now.

These ordinal Likert-response items are coded (1, strongly disagree; 2, somewhat disagree; 3, neither agree or disagree; 4, somewhat agree; and 5, strongly agree). We wanted to create an unweighted sum of the items such that large values indicate high external demands on the participant's time, so we reversed the coding of `train_goal_fac_bar_5_numeric` to (1, strongly agree; 2, somewhat agree; 3, neither agree or disagree; 4, somewhat disagree; and 5, strongly disagree) for that to make sense. The `Descriptive_Analyses.qmd` script examined the 3 items and the led us to decide to only use the latter two. See the output from that script for why we excluded the first of those items. This script will create copies of the latter two items with shorter names that reflect potential barriers to participation posed by work responsibilities (`Barrier_WR`) and family obligations (`Barrier_FO`) for use in modeling.

```r
```{r}
#| label: create-Applicants

Applicants <- Applicants_Raw %>%
  rename(Education = educ_dem_num,
         License = license_num,
         Employ_Status = employment_dem_num,
         Employ_Nurse = employednurse_dem_num,
         Nurse_Years = nursexperience_years,
         Recruited_How = recruitment_num,
         Prior_SANE = prevSANEtrain_num,
         Trained1_Year = trained1year_num,
         Trained2_Year = trained2year,
         Disability = Disability_status,
         Sex = sex_dem_num,
         Pronoun = pronoun_dem_num,
         Age = age_num,
         Race_Ethnicity = race_ethnicity_num,
         Income = income_dem_num,
         LGBTQ = LGBTQ_num,
         NPQ_1 = npq1,
         NPQ_2 = npq2,
         NPQ_3 = npq3,
         NPQ_4 = npq4,
         NPQ_5 = npq5,
         NPQ_6 = npq6,
         NPQ_7 = npq7,
         NPQ_8 = npq8,
         NPQ_9 = npq9,
         NPQ_10 = npq10,
         NPQ_11 = npq11,
         NPQ_12 = npq12,
         NPQ_13 = npq13,
         NPQ_14 = npq14,
         NPQ_15 = npq15,
         NPQ_16 = npq16,
         NPQ_17 = npq17,
         NPQ_18 = npq18,
         NPQ_19 = npq19,
         NPQ_20 = npq20,
         NPQ_21 = npq21,
         NPQ_22 = npq22,
         NPQ_23 = npq23,
         NPQ_24 = npq24,
         NPQ_25 = npq25,
         NPQ_26 = npq26,
         NPQ_27 = npq27,
         NPQ_28 = npq28,
         NPQ_29 = npq29,
         NPQ_30 = npq30,
         NPQ_1_r = npq1r,
         NPQ_4_r = npq4r,
         NPQ_15_r = npq15r,
         NPQ_17_r = npq17r,
         NPQ_29_r = npq29r,
         Plan_After_Training = PlansFollowingTraining_Num,
         Train_Goal_Fac_Bar_1 = train_goal_fac_bar_1_numeric,
         Train_Goal_Fac_Bar_2 = train_goal_fac_bar_2_numeric,
         Train_Goal_Fac_Bar_3 = train_goal_fac_bar_3_numeric,
         Train_Goal_Fac_Bar_4 = train_goal_fac_bar_4_numeric,
         Train_Goal_Fac_Bar_5 = train_goal_fac_bar_5_numeric,
         Train_Goal_Fac_Bar_6 = train_goal_fac_bar_6_numeric,
         Train_Goal_Fac_Bar_7 = train_goal_fac_bar_7_numeric,
         Train_Goal_Fac_Bar_8 = train_goal_fac_bar_8_numeric,
         Train_Goal_Fac_Bar_9 = train_goal_fac_bar_9_numeric,
         Train_Goal_Fac_Bar_10 = train_goal_fac_bar_10_numeric,
         Tech_Basic_1 = techbasic1_numeric,
         Tech_Basic_2 = techbasic2_numeric,
         Tech_Basic_3 = techbasic3_numeric,
         Tech_Feel_1 = techfeel1_numeric,
         Tech_Feel_2 = techfeel2_numeric,
         Tech_Feel_3 = techfeel3_numeric,
         Tech_Feel_4 = techfeel4_numeric,
         Tech_Feel_5 = techfeel5_numeric,
         ProQOL_CS = CS,
         ProQOL_BO = BO,
         ProQOL_STS = STS,
         CSW_EvalDate = CSWEvalDate,
         CSW_Sat_1 = cswsat1_numeric,
         CSW_Sat_2 = cswsat2_numeric,
         CSW_Sat_3 = cswsat3_numeric,
         CSW_Sat_4 = cswsat4_numeric,
         CSW_Sat_5 = cswsat5_numeric,
         CSW_Sat_6 = cswsat6_numeric,
         CSW_Sat_7 = cswsat7_numeric,
         Finished_CSW = finishedCSW,
         Finished_Didactic = finisheddidac,
         Completed_Didactic_NotCSW = completediddnotcsw,
         Started_CSW = incompleteCSW,
         Didactic_Sat_1 = didacticsat1_num,
         Didactic_Sat_2 = didacticsat2_num,
```
```

```
        Didactic_Sat_3 = didacticsat3_num,
        Didactic_Sat_4 = didacticsat4_num,
        Didactic_Sat_5 = didacticsat5_num,
        Didactic_Sat_6 = didacticsat6_num,
        Didactic_Sat_7 = didacticsat7_num,
        Didactic_EvalDate = DIDEvalDate,
        Didactic_SDate = Didactic_Start,
        Didactic_EDate = Didactic_End,
        Attrited = prog_attrited,
        Exit_Date = exit_date,
        Stage_Raw = stg_reached,
        Started_Mod_1 = Mod1_started,
        Finished_Mod_1 = Mod1_finished,
        Started_Mod_2 = Mod2_started,
        Finished_Mod_2 = Mod2_finished,
        Started_Mod_3 = Mod3_started,
        Finished_Mod_3 = Mod3_finished,
        Started_Mod_4 = Mod4_started,
        Finished_Mod_4 = Mod4_finished,
        Started_Mod_5 = Mod5_started,
        Finished_Mod_5 = Mod5_finished,
        Started_Mod_6 = Mod6_started,
        Finished_Mod_6 = Mod6_finished,
        Started_Mod_7 = Mod7_started,
        Finished_Mod_7 = Mod7_finished,
        Started_Mod_8 = Mod8_started,
        Finished_Mod_8 = Mod8_finished,
        Started_Mod_9 = Mod9_started,
        Finished_Mod_9 = Mod9_finished,
        Started_Mod_10 = Mod10_started,
        Finished_Mod_10 = Mod10_finished,
        Started_Mod_11 = Mod11_started,
        Finished_Mod_11 = Mod11_finished,
        Started_Mod_12 = Mod12_started,
        Finished_Mod_12 = Mod12_finished,
        Mod_1_SDate = Mod1Start,
        Mod_2_SDate = Mod2Start,
        Mod_3_SDate = Mod3Start,
        Mod_4_SDate = Mod4Start,
        Mod_5_SDate = Mod5Start,
        Mod_6_SDate = Mod6Start,
        Mod_7_SDate = Mod7Start,
        Mod_8_SDate = Mod8Start,
        Mod_9_SDate = Mod9Start,
        Mod_10_SDate = Mod10Start,
        Mod_11_SDate = Mod11Start,
        Mod_12_SDate = Mod12Start) %>%
# Recode Stage_Raw into final set of stages to be modeled.
mutate(Stage_Reached = case_when(Stage_Raw == 2 ~ 1,
                                 Stage_Raw == 3 ~ 1,
                                 Stage_Raw == 4 ~ 2,
                                 Stage_Raw == 5 ~ 3,
                                 Stage_Raw == 6 ~ 3,
                                 Stage_Raw == 7 ~ 4,
                                 .default = NA)) %>%
# Recode Attrited to reverse coding of the binary indicator.
mutate(Attrited = 1 - Attrited) %>%
# Recode practicset_dem_num into Setting, combining Rural & Tribal categories.
mutate(Setting = if_else(practicset_dem_num %in% c(2,4),
                         true = 2,
                         false = practicset_dem_num)) %>%
# Reverse code an item, then compute a sum score for a new scale.
mutate(Train_Goal_Fac_Bar_5R = 6 - Train_Goal_Fac_Bar_5) %>%
rowwise() %>%
mutate(Barrier_WR = Train_Goal_Fac_Bar_8,
       Barrier_FO = Train_Goal_Fac_Bar_10) %>%
ungroup() %>%
# Set variable and value labels.
set_variable_labels(ID = "Applicant ID (may be blank if application was incomplete)",
                    NPQ_1_r = "I am happy. (reverse coded)",
                    NPQ_4_r = "I feel connected to others. (reverse coded)",
                    NPQ_15_r = "I have beliefs that sustain me. (reverse coded)",
                    NPQ_17_r = "I am the person I always wanted to be. (reverse coded)",
                    NPQ_29_r = "I am a very caring person. (reverse coded)",
                    Stage_Raw = "Stage Reached by Participant (Raw Values)",
                    Stage_Reached = "Maximum Stage Reached by Participant (Recoded)",
                    Setting = "How would you describe the primary setting you practice nursing in? (recoded)",
                    Train_Goal_Fac_Bar_5R = "I have time in my schedule to focus on this course",
                    Barrier_WR = "I have a lot of work responsibilities.",
                    Barrier_FO = "I have a lot of family obligations right now.") %>%
set_value_labels(NPQ_1_r = reversed_NPQ_value_labels,
                 NPQ_4_r = reversed_NPQ_value_labels,
                 NPQ_15_r = reversed_NPQ_value_labels,
                 NPQ_17_r = reversed_NPQ_value_labels,
                 NPQ_29_r = reversed_NPQ_value_labels,
                 Stage_Raw = Stage_Raw_value_labels,
                 Stage_Reached = Stage_Reached_value_labels,
                 Eligible = Binary_Labels,
                 Enrolled = Binary_Labels,
                 Attrited = Binary_Labels,
                 Started_Mod_1 = Binary_Labels,
```

```
                  Started_Mod_2 = Binary_Labels,
                  Started_Mod_3 = Binary_Labels,
                  Started_Mod_4 = Binary_Labels,
                  Started_Mod_5 = Binary_Labels,
                  Started_Mod_6 = Binary_Labels,
                  Started_Mod_7 = Binary_Labels,
                  Started_Mod_8 = Binary_Labels,
                  Started_Mod_9 = Binary_Labels,
                  Started_Mod_10 = Binary_Labels,
                  Started_Mod_11 = Binary_Labels,
                  Started_Mod_12 = Binary_Labels,
                  Finished_Mod_1 = Binary_Labels,
                  Finished_Mod_2 = Binary_Labels,
                  Finished_Mod_3 = Binary_Labels,
                  Finished_Mod_4 = Binary_Labels,
                  Finished_Mod_5 = Binary_Labels,
                  Finished_Mod_6 = Binary_Labels,
                  Finished_Mod_7 = Binary_Labels,
                  Finished_Mod_8 = Binary_Labels,
                  Finished_Mod_9 = Binary_Labels,
                  Finished_Mod_10 = Binary_Labels,
                  Finished_Mod_11 = Binary_Labels,
                  Finished_Mod_12 = Binary_Labels,
                  Setting = Setting_Labels,
                  Train_Goal_Fac_Bar_5R = Barriers_Reversed_Labels,
                  Barrier_WR = Barriers_Labels,
                  Barrier_FO = Barriers_Labels)
```

The imported `Applicants` data frame has 497 rows and 154 variables. Table 5 shows a list of the variable positions, names, classes, and labels.

```
```{r}
#| label: tbl-Applicants
#| tbl-cap: "Applicants: Updated Variable Names and Classes."

ApplicantsVars <- list(Applicants) %>%
  map_dfr(~ tibble(Name = names(.x),
                   Class = map_chr(.x, all_classes),
                   Label = var_label(.x))) %>%
  rowid_to_column(., "Position")

kable(ApplicantsVars, format = "latex", booktabs = TRUE, longtable = TRUE) %>%
  kable_styling(font_size = 7, latex_options = c("repeat_header")) %>%
  column_spec(column = 3, width = "3cm") %>%
  column_spec(column = 4, width = "6.5cm")
```
```

Table 5: Applicants: Updated Variable Names and Classes.

Position	Name	Class	Label
1	ID	character	Applicant ID (may be blank if application was incomplete)
2	Education	haven_labelled, vctrs_vctr, double	What is your highest level of education?
3	License	haven_labelled, vctrs_vctr, double	What is your licensure or certification for practice?
4	practicset_dem_num	haven_labelled, vctrs_vctr, double	How would you describe the primary setting you practice nursing in?
5	Employ_Status	haven_labelled, vctrs_vctr, double	What is your current employment status?
6	Employ_Nurse	haven_labelled, vctrs_vctr, double	If currently employed, are you working as a nurse?
7	Nurse_Years	numeric	How many years have you been practicing as a nurse (since completing your initial nursing program)?
8	Recruited_How	haven_labelled, vctrs_vctr, double	How did you learn about this opportunity? - Selected Choice
9	Prior_SANE	haven_labelled, vctrs_vctr, double	Previous SANE training (numeric)
10	Trained1_Year	numeric	Year training was completed (numeric)
11	Trained2_Year	numeric	What year(s) did you complete your didactic and clinical skills SANE training?
12	Disability	haven_labelled, vctrs_vctr, double	Do you have a disability ?
13	Sex	haven_labelled, vctrs_vctr, double	What was your sex assigned at birth?

(continued)

Position	Name	Class	Label
14	Pronoun	haven_labelled, vctrs_vctr, double	What are your pronouns? - Selected Choice
15	Age	numeric	How old are you?
16	Race_Ethnicity	haven_labelled, vctrs_vctr, double	How would you describe your race/ethnicity? (Select all that apply) - Selected Choice
17	Income	haven_labelled, vctrs_vctr, double	Income level (numeric)
18	LGBTQ	haven_labelled, vctrs_vctr, double	Do you identify as a member of the LGBTQIA2S+ community?
19	NPQ_1	haven_labelled, vctrs_vctr, double	I am happy.
20	NPQ_2	haven_labelled, vctrs_vctr, double	I am preoccupied with more than one person I care for as a nurse
21	NPQ_3	haven_labelled, vctrs_vctr, double	I get satisfaction from being able to care for people.
22	NPQ_4	haven_labelled, vctrs_vctr, double	I feel connected to others.
23	NPQ_5	haven_labelled, vctrs_vctr, double	I jump or am startled by unexpected sounds.
24	NPQ_6	haven_labelled, vctrs_vctr, double	I feel invigorated after working with those I care for.
25	NPQ_7	haven_labelled, vctrs_vctr, double	I find it difficult to separate my personal life from my life as a nurse.
26	NPQ_8	haven_labelled, vctrs_vctr, double	I am not as productive at work because I am losing sleep over traumatic experiences of a person I care for as a nurse.
27	NPQ_9	haven_labelled, vctrs_vctr, double	I think that I might have been affected by the traumatic stress of those I care for as a nurse.
28	NPQ_10	haven_labelled, vctrs_vctr, double	I feel trapped by my job as a nurse.
29	NPQ_11	haven_labelled, vctrs_vctr, double	Because of my nursing, I have felt ”on edge” about various things.
30	NPQ_12	haven_labelled, vctrs_vctr, double	I like my work as a nurse.
31	NPQ_13	haven_labelled, vctrs_vctr, double	I feel depressed because of the traumatic experiences of the people I care for as a nurse.
32	NPQ_14	haven_labelled, vctrs_vctr, double	I feel as though I am experiencing the trauma of someone I have cared for as a nurse.
33	NPQ_15	haven_labelled, vctrs_vctr, double	I have beliefs that sustain me.
34	NPQ_16	haven_labelled, vctrs_vctr, double	I am pleased with how I am able to keep up with nursing techniques and protocols.
35	NPQ_17	haven_labelled, vctrs_vctr, double	I am the person I always wanted to be.
36	NPQ_18	haven_labelled, vctrs_vctr, double	My work makes me feel satisfied.
37	NPQ_19	haven_labelled, vctrs_vctr, double	I feel worn out because of my work as a nurse.
38	NPQ_20	haven_labelled, vctrs_vctr, double	I have happy thoughts and feelings about those I care for as a nurse and how I could help them.
39	NPQ_21	haven_labelled, vctrs_vctr, double	I feel overwhelmed because my case work load seems endless.
40	NPQ_22	haven_labelled, vctrs_vctr, double	I believe I can make a difference through my work.
41	NPQ_23	haven_labelled, vctrs_vctr, double	I avoid certain activities or situations because they remind me of frightening experiences of the people I care for as a nurse.
42	NPQ_24	haven_labelled, vctrs_vctr, double	I am proud of what I can do to nurse.
43	NPQ_25	haven_labelled, vctrs_vctr, double	As a result of my nursing, I have intrusive, frightening thoughts.
44	NPQ_26	haven_labelled, vctrs_vctr, double	I feel ”bogged down” by the system.
45	NPQ_27	haven_labelled, vctrs_vctr, double	I have thoughts that I am a ”success” as a nurse.

(continued)

Position	Name	Class	Label
46	NPQ_28	haven_labelled, vctrs_vctr, double	I can't recall important parts of my work with trauma victims.
47	NPQ_29	haven_labelled, vctrs_vctr, double	I am a very caring person.
48	NPQ_30	haven_labelled, vctrs_vctr, double	I am happy that I chose to do this work.
49	NPQ_1_r	haven_labelled, vctrs_vctr, double	I am happy. (reverse coded)
50	NPQ_4_r	haven_labelled, vctrs_vctr, double	I feel connected to others. (reverse coded)
51	NPQ_15_r	haven_labelled, vctrs_vctr, double	I have beliefs that sustain me. (reverse coded)
52	NPQ_17_r	haven_labelled, vctrs_vctr, double	I am the person I always wanted to be. (reverse coded)
53	NPQ_29_r	haven_labelled, vctrs_vctr, double	I am a very caring person. (reverse coded)
54	Motivation_ProfGrowth	haven_labelled, vctrs_vctr, double	What are your primary motivations for seeking SANE training (Select all that apply)? - Selected Choice
55	Motivation_PersonalConn	haven_labelled, vctrs_vctr, double	What are your primary motivations for seeking SANE training (Select all that apply)? - Selected Choice
56	Motivation_NeedSANE	haven_labelled, vctrs_vctr, double	What are your primary motivations for seeking SANE training (Select all that apply)? - Selected Choice
57	Motivation_PersonalLearn	haven_labelled, vctrs_vctr, double	What are your primary motivations for seeking SANE training (Select all that apply)? - Selected Choice
58	Motivation_Humanitarian	haven_labelled, vctrs_vctr, double	What are your primary motivations for seeking SANE training (Select all that apply)? - Selected Choice
59	Motivation_EmpRequire	haven_labelled, vctrs_vctr, double	What are your primary motivations for seeking SANE training (Select all that apply)? - Selected Choice
60	Motivation_Other	haven_labelled, vctrs_vctr, double	What are your primary motivations for seeking SANE training (Select all that apply)? - Selected Choice
61	Plan_After_Training	haven_labelled, vctrs_vctr, double	Which best describes your plans following the training?
62	Train_Goal_Fac_Bar_1	haven_labelled, vctrs_vctr, double	I will try to learn as much as I can from this course
63	Train_Goal_Fac_Bar_2	haven_labelled, vctrs_vctr, double	I am interested in learning this training material
64	Train_Goal_Fac_Bar_3	haven_labelled, vctrs_vctr, double	My supervisor is supportive of me attending this training
65	Train_Goal_Fac_Bar_4	haven_labelled, vctrs_vctr, double	I am willing to invest effort to improve skills and competencies related to becoming a forensic nurse
66	Train_Goal_Fac_Bar_5	haven_labelled, vctrs_vctr, double	I have time in my schedule to focus on this course
67	Train_Goal_Fac_Bar_6	haven_labelled, vctrs_vctr, double	I will get more from this training than most people
68	Train_Goal_Fac_Bar_7	haven_labelled, vctrs_vctr, double	I am motivated to learn the skills needed to be a forensic nurse examiner
69	Train_Goal_Fac_Bar_8	haven_labelled, vctrs_vctr, double	I have a lot of work responsibilities
70	Train_Goal_Fac_Bar_9	haven_labelled, vctrs_vctr, double	I am willing to exert considerable effort in this training in order to improve my skills
71	Train_Goal_Fac_Bar_10	haven_labelled, vctrs_vctr, double	I have a lot of family obligations right now
72	Tech_Basic_1	haven_labelled, vctrs_vctr, double	Have you ever taken an online course before?
73	Tech_Basic_2	haven_labelled, vctrs_vctr, double	Do you have regular access to a computer (e.g., desktop, laptop, tablet) at home?
74	Tech_Basic_3	haven_labelled, vctrs_vctr, double	Do you have reliable access to high-speed internet at home?
75	Tech_Feel_1	haven_labelled, vctrs_vctr, double	I usually get very anxious about using a computer
76	Tech_Feel_2	haven_labelled, vctrs_vctr, double	I'm not the type to do well with computers
77	Tech_Feel_3	haven_labelled, vctrs_vctr, double	I have a lot of self-confidence when it comes to working with computers

(continued)

Position	Name	Class	Label
78	Tech_Feel_4	haven_labelled, vctrs_vctr, double	I like working with computers
79	Tech_Feel_5	haven_labelled, vctrs_vctr, double	The online course will take less time than a traditional course in a classroom
80	ProQOL_CS	numeric	sum scores compassion from proqol
81	ProQOL_BO	numeric	sum scores burnout from proqol
82	ProQOL_STS	numeric	sum scores sec traumatic stress
83	CSW_EvalDate	Date	CSW Eval Date
84	CSW_Sat_1	haven_labelled, vctrs_vctr, double	The pacing of the clinical training was satisfactory.
85	CSW_Sat_2	haven_labelled, vctrs_vctr, double	The clinical concepts were covered_x000D_adequately.
86	CSW_Sat_3	haven_labelled, vctrs_vctr, double	There was enough time to ask_x000D_questions during the clinical training.
87	CSW_Sat_4	haven_labelled, vctrs_vctr, double	There was enough time to practice_x000D_different components of the medical_x000D_forensic exam.
88	CSW_Sat_5	haven_labelled, vctrs_vctr, double	The clinical training met my_x000D_expectations.
89	CSW_Sat_6	haven_labelled, vctrs_vctr, double	I learned a considerable amount during_x000D_the clinical training.
90	CSW_Sat_7	haven_labelled, vctrs_vctr, double	The clinical training helped me feel_x000D_more confident as a forensic examiner.
91	Finished_CSW	haven_labelled, vctrs_vctr, double	Enrolled people who completed didactic and CSW
92	Finished_Didactic	haven_labelled, vctrs_vctr, double	Did the participant complete the didactic?
93	Completed_Didactic_NotCSW	numeric	NULL
94	Started_CSW	haven_labelled, vctrs_vctr, double	Enrolled people who completed didactic and started CSW
95	Didactic_Sat_1	haven_labelled, vctrs_vctr, double	There was enough time to learn the material.
96	Didactic_Sat_2	haven_labelled, vctrs_vctr, double	The topics were covered adequately.
97	Didactic_Sat_3	haven_labelled, vctrs_vctr, double	The weekly discussion board helped me understand the material better.
98	Didactic_Sat_4	haven_labelled, vctrs_vctr, double	The online training met my expectations.
99	Didactic_Sat_5	haven_labelled, vctrs_vctr, double	I learned a considerable amount during the online training.
100	Didactic_Sat_6	haven_labelled, vctrs_vctr, double	I am satisfied with the amount I have learned in this course.
101	Didactic_Sat_7	haven_labelled, vctrs_vctr, double	Accounting for the materials you learned in this online course, please rate your response to the following question:I have improved knowledge and skills to care for sexual assault patients.
102	Didactic_EvalDate	Date	Didactic evaluation date.
103	Applied_Date	POSIXct, POSIXt	Date application applied for program.
104	Eligible	haven_labelled, vctrs_vctr, double	Eligible for the program.
105	Enrolled	haven_labelled, vctrs_vctr, double	Whether or not patrticipants were eligible, accepted, and started the program.
106	Enrolled_Date	Date	Date particpant was offered enrollment in the program. All participants in each cohort assisgned same date.
107	Cohort	numeric	Cohort the participant participated in
108	Didactic_SDate	Date	Date when participant started the didactic training. All participants in cohort assigned samed date.
109	Didactic_Status	haven_labelled, vctrs_vctr, double	Did the participant complete the didactic training?
110	Didactic_EDate	Date	Date when participant completed the didactic training.
111	Attrited	haven_labelled, vctrs_vctr, double	NULL
112	Exit_Date	Date	Participants program exit date.

(continued)

Position	Name	Class	Label
113	Stage_Raw	haven_labelled, vctrs_vctr, double	Stage Reached by Participant (Raw Values)
114	Started_Mod_1	haven_labelled, vctrs_vctr, double	Overview of Forensic Nursing and Sexual Violence - did the participant start this module?
115	Finished_Mod_1	haven_labelled, vctrs_vctr, double	Overview of Forensic Nursing and Sexual Violence - did the participant finish this module?
116	Started_Mod_2	haven_labelled, vctrs_vctr, double	Victim Responses and Crisis Intervention - did the participant start this module?
117	Finished_Mod_2	haven_labelled, vctrs_vctr, double	Victim Responses and Crisis Intervention - did the participant finish this module?
118	Started_Mod_3	haven_labelled, vctrs_vctr, double	Collaborating with Community Agencies - did the participant start this module?
119	Finished_Mod_3	haven_labelled, vctrs_vctr, double	Collaborating with Community Agencies - did the participant finish this module?
120	Started_Mod_4	haven_labelled, vctrs_vctr, double	Medical-Forensic History Taking - did the participant start this module?
121	Finished_Mod_4	haven_labelled, vctrs_vctr, double	Medical-Forensic History Taking - did the participant finish this module?
122	Started_Mod_5	haven_labelled, vctrs_vctr, double	Observing and Assessing Physical Examination Findings - did the participant start this module?
123	Finished_Mod_5	haven_labelled, vctrs_vctr, double	Observing and Assessing Physical Examination Findings-did the participant finish this module?
124	Started_Mod_6	haven_labelled, vctrs_vctr, double	Medical-Forensic Specimen Collection - did the participant start this module?
125	Finished_Mod_6	haven_labelled, vctrs_vctr, double	Medical-Forensic Specimen Collection - did the participant finish this module?
126	Started_Mod_7	haven_labelled, vctrs_vctr, double	Medical-Forensic Photography - did the participant start this module?
127	Finished_Mod_7	haven_labelled, vctrs_vctr, double	Medical-Forensic Photography - did the participant finish this module?
128	Started_Mod_8	haven_labelled, vctrs_vctr, double	Sexually Transmitted Infection Testing and Prophylaxis - did the participant start this module?
129	Finished_Mod_8	haven_labelled, vctrs_vctr, double	Sexually Transmitted Infection Testing and Prophylaxis - did the participant finish this module?
130	Started_Mod_9	haven_labelled, vctrs_vctr, double	Pregnancy Risk Evaluation and Care - did the participant start this module?
131	Finished_Mod_9	haven_labelled, vctrs_vctr, double	Pregnancy Risk Evaluation and Care- did the participant finish this module?
132	Started_Mod_10	haven_labelled, vctrs_vctr, double	Medical-Forensic Documentation- did the participant start this module?
133	Finished_Mod_10	haven_labelled, vctrs_vctr, double	Medical-Forensic Documentation- did the participant finish this module?
134	Started_Mod_11	haven_labelled, vctrs_vctr, double	Discharge and Follow-up Planning - did the participant start this module?
135	Finished_Mod_11	haven_labelled, vctrs_vctr, double	Discharge and Follow-up Planning- did the participant finish this module?
136	Started_Mod_12	haven_labelled, vctrs_vctr, double	Legal Considerations and Judicial Proceedings - did the participant start this module?
137	Finished_Mod_12	haven_labelled, vctrs_vctr, double	Legal Considerations and Judicial Proceedings- did the participant finish this module?
138	Mod_1_SDate	Date	Date when module was scheduled to start for the participants. all participants if started given same date.
139	Mod_2_SDate	Date	Date when module was scheduled to start for the participants. all participants if started given same date.
140	Mod_3_SDate	Date	Date when module was scheduled to start for the participants. all participants if started given same date.
141	Mod_4_SDate	Date	Date when module was scheduled to start for the participants. all participants if started given same date.
142	Mod_5_SDate	Date	Date when module was scheduled to start for the participants. all participants if started given same date.

(continued)

Position	Name	Class	Label
143	Mod_6_SDate	Date	Date when module was scheduled to start for the participants. all participants if started given same date.
144	Mod_7_SDate	Date	Date when module was scheduled to start for the participants. all participants if started given same date.
145	Mod_8_SDate	Date	Date when module was scheduled to start for the participants. all participants if started given same date.
146	Mod_9_SDate	Date	Date when module was scheduled to start for the participants. all participants if started given same date.
147	Mod_10_SDate	Date	Date when module was scheduled to start for the participants. all participants if started given same date.
148	Mod_11_SDate	Date	Date when module was scheduled to start for the participants. all participants if started given same date.
149	Mod_12_SDate	Date	Date when module was scheduled to start for the participants. all participants if started given same date.
150	Stage_Reached	haven_labelled, vctrs_vctr, double	Maximum Stage Reached by Participant (Recoded)
151	Setting	haven_labelled, vctrs_vctr, double	How would you describe the primary setting you practice nursing in? (recoded)
152	Train_Goal_Fac_Bar_5R	haven_labelled, vctrs_vctr, double	I have time in my schedule to focus on this course
153	Barrier_WR	haven_labelled, vctrs_vctr, double	I have a lot of work responsibilities.
154	Barrier_FO	haven_labelled, vctrs_vctr, double	I have a lot of family obligations right now.

3.4 Compare Stage_Raw to Binary Stage Indicators

Here we are comparing the stage variables against the binary indicators associated with specific stages to ensure logical consistency between variables. This helped us detect and correct some problems with the data. We leave it here now to document that we did it.

In the crosstab below check that:

- All (and only) applicants with `Stage_Raw` = `Applied` should have `Eligible` = 0.
- All (and only) applicants with `Stage_Raw` ≥ `Eligible` should have `Eligible` = 1.

```
```{r}
#| label: xtab-Stage-Raw-Eligible

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

```
                   Eligible
as_factor(Stage_Raw)   0   1 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
```

In the crosstab below check that:

- All (and only) applicants with `Stage_Raw` $\leq$ `Eligible` should have `Enrolled` $= 0$.
- All (and only) applicants with `Stage_Raw` $\geq$ `Enrolled` should have `Enrolled` $= 1$.

```{r}
#| label: xtab-Stage-Raw-Enrolled

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

```
                   Enrolled
as_factor(Stage_Raw)   0   1 Sum
        Applied      170   0 170
        Eligible      73   0  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          243 254 497
```

In the crosstab below check that:

- All (and only) applicants with `Stage_Raw` $\leq$ `Eligible` should have `Started_Mod_1` $=$ `<NA>` (missing).
- All applicants with `Stage_Raw` $=$ `Enrolled` should have `Started_Mod_1` $= 0$.
- Only applicants with `Stage_Raw` $\geq$ `Started_DT` should have `Started_Mod_1` $= 1$.

Here we are using starting didactic module 1 as a rough proxy for starting the didactic training component.

```{r}
#| label: xtab-Stage-Raw-Started-Mod-1

xtabs(~as_factor(Stage_Raw) + Started_Mod_1, addNA = TRUE,
      data = Applicants) %>%
  addmargins()
```

```
                   Started_Mod_1
as_factor(Stage_Raw)   0   1 <NA> Sum
        Applied        0   0  170 170
        Eligible       0   0   73  73
        Enrolled      17   0    0  17
        Started DT     3  58    0  61
        Finished DT    0  16    0  16
        Started CSW    0   1    0   1
        Finished CSW   0 159    0 159
        Sum           20 234  243 497
```

In the crosstab below check that:

- All (and only) applicants with `Stage_Raw` $\leq$ `Started DT` should have `Finished_Didactic` $=$ 0.
- All (and only) applicants with `Stage_Raw` $\geq$ `Finished DT` should have `Finished_Didactic` $= 1$.

```{r}
#| label: xtab-Stage-Raw-Finished-Didactic

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

```
                  Finished_Didactic
as_factor(Stage_Raw)   0   1 Sum
        Applied      170   0 170
        Eligible      73   0  73
        Enrolled      17   0  17
        Started DT    61   0  61
        Finished DT    0  16  16
        Started CSW    0   1   1
        Finished CSW   0 159 159
        Sum          321 176 497
```

In the crosstab below check that:

- All (and only) applicants with `Stage_Raw` $\leq$ `Finished DT` should have `Started_CSW` $= 0$.
- All (and only) applicants with `Stage_Raw` $\geq$ `Started CSW` should have `Started_CSW` $= 1$.

```
```{r}
#| label: xtab-Stage-Raw-CSW-Incomplete

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

```
                  Started_CSW
as_factor(Stage_Raw)   0   1 Sum
        Applied      170   0 170
        Eligible      73   0  73
        Enrolled      17   0  17
        Started DT    61   0  61
        Finished DT   16   0  16
        Started CSW    0   1   1
        Finished CSW   0 159 159
        Sum          337 160 497
```

In the crosstab below check that:

- All (and only) applicants with `Stage_Raw` $\leq$ `Started CSW` should have `Finished_CSW` $= 0$.
- All (and only) applicants with `Stage_Raw` $=$ `Finished CSW` should have `Finished_CSW` $= 1$.

```
```{r}
#| label: xtab-Stage-Raw-Finished-CSW

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

```
                  Finished_CSW
as_factor(Stage_Raw)   0   1 Sum
        Applied      170   0 170
        Eligible      73   0  73
        Enrolled      17   0  17
        Started DT    61   0  61
        Finished DT   16   0  16
        Started CSW    1   0   1
        Finished CSW   0 159 159
        Sum          338 159 497
```

3.5 Create Eligible Applicants Datasets

For research question RQ1, we need to compare eligible applicants who enrolled versus did not enroll. Subsetting the `Applicants` dataset to create an `Eligible_Applicants` dataset representing the intended population relevant to that research question will simplify analyses later.

Tip

The expected uses for the `Eligible_Applicants` data are:

- Compute enrollment rate and associated confidence interval.
- Conduct simple inferential analyses comparing eligible vs enrolled applicants to answer RQ1.
- Serve as the source from which the `Enrolled_Applicants` dataset is derived.

```{r}
#| label: create-Eligible-Applicants

Eligible_Applicants <- Applicants %>%
  filter(Stage_Raw >= 2)
```

Next we need to assess whether these eligible applicants all have complete data on the variables to be used in our RQ1 analyses. Table 6 shows an overview of missingness with respect to cases, variables, and values.

```{r}
#| label: tbl-Eligible-Applicants-cvv-missingness
#| tbl-cap: Cases, Variables, and Values Missingness for RQ1 Among Eligible
#|          Applicants

 Eligible_Applicants %>%
  select(ID, Enrolled, Education, License, Setting, Employ_Status, Employ_Nurse,
         Nurse_Years, Prior_SANE, Motivation_PersonalConn, Motivation_NeedSANE,
         Barrier_FO, Barrier_WR, ProQOL_BO, ProQOL_CS, ProQOL_STS) %>%
  cvv_missingness() %>%
  kable(format = "latex", booktabs = TRUE, digits = 1,
        col.names = c("Subset", rep(c("Count", "Percent"), times = 3))) %>%
  kable_styling() %>%
  add_header_above(header = c(" ", "Cases" = 2, "Variables" = 2, "Values" = 2))
```

Table 6: Cases, Variables, and Values Missingness for RQ1 Among Eligible Applicants

	Cases		Variables		Values	
Subset	Count	Percent	Count	Percent	Count	Percent
Complete	313	95.7	12	75	5212	99.6
Incomplete	14	4.3	4	25	20	0.4
All	327	100.0	16	100	5232	100.0

We can see that only a few eligible applicants have any missing data on the variables we want to use in the models. Table 7 shows variable-specific rates of missingness.

```{r}
#| label: tbl-Eligible-Applicants-var-missingness
#| tbl-cap: Variable-Specific Missingness for RQ1 Among Eligible Applicants

 Eligible_Applicants %>%
  select(ID, Enrolled, Education, License, Setting, Employ_Status, Employ_Nurse,
         Nurse_Years, Prior_SANE, Motivation_PersonalConn, Motivation_NeedSANE,
         Barrier_FO, Barrier_WR, ProQOL_BO, ProQOL_CS, ProQOL_STS) %>%
  var_missingness() %>%
  kable(format = "latex", booktabs = TRUE, digits = 1,
        col.names = c("Position", "Name", "N_Total", "N", "%", "N", "%")) %>%
  kable_styling() %>%
  add_header_above(header = c(" " = 3, "Valid Values" = 2, "Missing Values" = 2))
```

Table 7: Variable-Specific Missingness for RQ1 Among Eligible Applicants

			Valid Values		Missing Values	
Position	Name	N_Total	N	%	N	%
1	ID	327	327	100.0	0	0.0
2	Enrolled	327	327	100.0	0	0.0
3	Education	327	327	100.0	0	0.0
4	License	327	327	100.0	0	0.0
5	Setting	327	327	100.0	0	0.0
6	Employ_Status	327	327	100.0	0	0.0
7	Employ_Nurse	327	316	96.6	11	3.4
8	Nurse_Years	327	327	100.0	0	0.0
9	Prior_SANE	327	327	100.0	0	0.0
10	Motivation_PersonalConn	327	327	100.0	0	0.0
11	Motivation_NeedSANE	327	327	100.0	0	0.0
12	Barrier_FO	327	327	100.0	0	0.0
13	Barrier_WR	327	327	100.0	0	0.0
14	ProQOL_BO	327	324	99.1	3	0.9
15	ProQOL_CS	327	324	99.1	3	0.9
16	ProQOL_STS	327	324	99.1	3	0.9

3.6 Create Enrolled Applicants Datasets

This section creates the `Enrolled_Applicants` data frame from the `Eligible_Applicants` data frame. We need this dataset because some analyses focus only on enrolled applicants. Feedback from key stakeholders indicates that our audience likely will think of attrition as dropping out the program after enrollment, rather than as the broader concept of dropping out of the pipeline that starts at application to the program.

> **Tip**
>
> The expected uses for the `Enrolled_Applicants` data are:
>
> - Stopping-ratio modeling to answering RQ2 and RQ3.
> - Descriptive analyses to answer RQ4.

```
```{r}
#| label: create-Enrolled-Applicants

Enrolled_Applicants <- Eligible_Applicants %>%
  filter(Stage_Raw >= 3)
```
```

Next, we need to assess whether these enrolled applicants all have complete data on the variables to be used in our RQ2 and RQ3 analyses. Table 8 shows an overview of missingness with respect to cases, variables, and values.

```{r}
#| label: tbl-Enrolled-Applicants-cvv-missingness
#| tbl-cap: Cases, Variables, and Values Missingness for RQ2 & RQ3 Among
#|          Enrolled Applicants

Enrolled_Applicants %>%
  select(ID, Stage_Reached, Barrier_FO, Barrier_WR, ProQOL_BO, ProQOL_CS,
         ProQOL_STS, Motivation_NeedSANE, Motivation_PersonalConn, Setting) %>%
  cvv_missingness() %>%
  kable(format = "latex", booktabs = TRUE, digits = 1,
        col.names = c("Subset", rep(c("Count", "Percent"), times = 3))) %>%
  kable_styling() %>%
  add_header_above(header = c(" ", "Cases" = 2, "Variables" = 2, "Values" = 2))
```

Table 8: Cases, Variables, and Values Missingness for RQ2 & RQ3 Among Enrolled Applicants

	Cases		Variables		Values	
Subset	Count	Percent	Count	Percent	Count	Percent
Complete	252	99.2	7	70	2534	99.8
Incomplete	2	0.8	3	30	6	0.2
All	254	100.0	10	100	2540	100.0

We can see that only a few eligible applicants have any missing data on the variables we want to use in the models. Table 9 shows variable-specific rates of missingness.

```{r}
#| label: tbl-Enrolled-Applicants-var-missingness
#| tbl-cap: Variable-Specific Missingness for RQ2 & RQ3 Among Enrolled
#|          Applicants

Enrolled_Applicants %>%
  select(ID, Stage_Reached, Barrier_FO, Barrier_WR, ProQOL_BO, ProQOL_CS,
         ProQOL_STS, Motivation_NeedSANE, Motivation_PersonalConn, Setting) %>%
  var_missingness() %>%
  kable(format = "latex", booktabs = TRUE, digits = 1,
        col.names = c("Position", "Name", "N_Total", "N", "%", "N", "%")) %>%
  kable_styling() %>%
  add_header_above(header = c(" " = 3, "Valid Values" = 2, "Missing Values" = 2))
```

Table 9: Variable-Specific Missingness for RQ2 & RQ3 Among Enrolled Applicants

			Valid Values		Missing Values	
Position	Name	N_Total	N	%	N	%
1	ID	254	254	100.0	0	0.0
2	Stage_Reached	254	254	100.0	0	0.0
3	Barrier_FO	254	254	100.0	0	0.0
4	Barrier_WR	254	254	100.0	0	0.0
5	ProQOL_BO	254	252	99.2	2	0.8
6	ProQOL_CS	254	252	99.2	2	0.8
7	ProQOL_STS	254	252	99.2	2	0.8
8	Motivation_NeedSANE	254	254	100.0	0	0.0
9	Motivation_PersonalConn	254	254	100.0	0	0.0
10	Setting	254	254	100.0	0	0.0

The missing data are all ProQOL scale scores and it appears that a handful of people simply did not complete the entire instrument. The rate of missingness is small enough that it should be ignorable

and listwise deletion will perform adequately. Therefore, below we create `Eligible_Applicants_CD` to include only those eligible applicants who have complete (non-missing) data on the variables that will be used in the stopping ratio model. Along the way, we create mean-centered versions of the continuous predictors that can be identified by a `C` prefix on the variable name‘.

```{r}
#| label: create-Enrolled-Applicants-CD

Enrolled_Applicants_CD <- Enrolled_Applicants %>%
  # Retain cases with no missing data on variables to be used in models.
  filter(if_all(c(ID, Stage_Reached, Barrier_FO, Barrier_WR, ProQOL_BO,
                  ProQOL_CS, ProQOL_STS, Motivation_NeedSANE,
                  Motivation_PersonalConn, Setting), ~ !is.na(.x))) %>%
  # Create mean-centered versions of continuous predictors.
  mutate(CBarrier_FO = Barrier_FO - mean(Barrier_FO),
         CBarrier_WR = Barrier_WR - mean(Barrier_WR),
         CProQOL_BO = ProQOL_BO - mean(ProQOL_BO),
         CProQOL_CS = ProQOL_CS - mean(ProQOL_CS),
         CProQOL_STS = ProQOL_STS - mean(ProQOL_STS))
```

3.7 Create Started DT Subset of Enrolled Applicants

When answering the part of RQ2 pertaining to specific training modules, we need to narrow our focus to the subset of `Enrolled_Applicants_CD` containing only enrolled applicants with `Stage_Reached` ≥ 2 (those who started the didactic training). Thus, below we create that additional dataset and call it `StartedDT_Applicants`.

```{r}
#| label: create-StartedDT-Applicants

StartedDT_Applicants <- Enrolled_Applicants_CD %>%
  filter(Stage_Reached >= 2)
```

3.8 Create Thresholds Dataset

This study will use a stopping-ratio model, which is also called a forward continuation-ratio model or a stage model (Fullerton & Xu, 2016; Liu & Bai, 2020). To run the planned model, we first reorganize the data from one row per person to one row per person per threshold attempted (Cole & Ananth, 2001; Fullerton & Xu, 2016). That reorganization allows us to use standard logistic regression modeling software to fit the model and flexibly choose whether predictor effects are parallel (constrained to equality across thresholds), or non-parallel (unconstrained and allowed to vary across thresholds) by either omitting or including interaction terms.

Figure 1 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"]

}
```

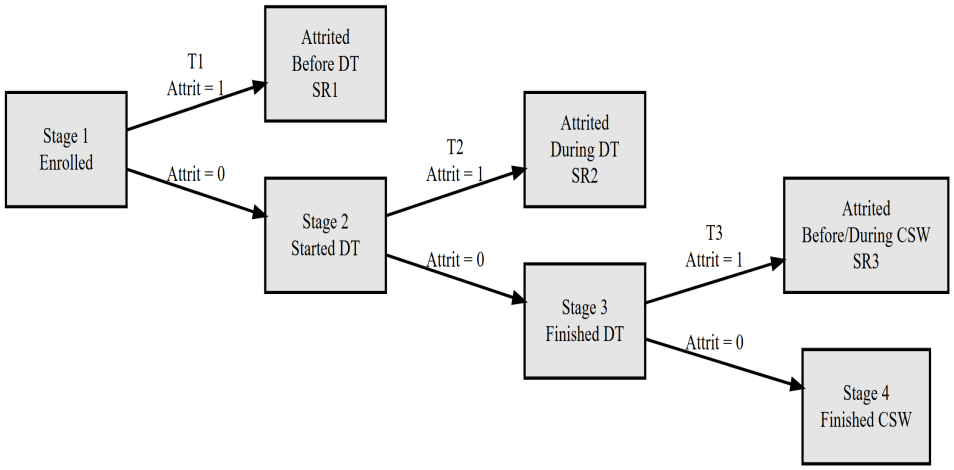

Figure 1: Stages, Thresholds (T1-T3), and Stopping Ratios (SR1-SR3) in the SANE Training Program. CSW, clinical skills workshop; DT, didactic training.

To create the person-threshold dataset `Thresholds`, we first create a data frame for each distinct threshold by filtering the `Enrolled_Applicants_CD` data by `Stage_Reached`, add a binary outcome variable called `Attrit` coded 1 for people who attrited at that threshold and 0 for people who did not, then vertically stack the threshold-specific datasets (`T1` to `T3`). Along the way, we create factor versions of the categorical variables that can be identified by the `.` suffix on the variable name.

```{r}
#| label: create-T1

T1 <- Enrolled_Applicants_CD %>%
  select(ID, Stage_Reached, Barrier_FO, Barrier_WR, ProQOL_BO, ProQOL_CS,
         ProQOL_STS, CBarrier_FO, CBarrier_WR, CProQOL_BO, CProQOL_CS,
         CProQOL_STS, Motivation_NeedSANE, Motivation_PersonalConn, Setting) %>%
  filter(Stage_Reached >= 1) %>%
  mutate(Threshold = 1,
         Attrit = if_else(Stage_Reached == Threshold, true = 1, false = 0))
```

```{r}
#| label: create-T2

T2 <- Enrolled_Applicants_CD %>%
  select(ID, Stage_Reached, Barrier_FO, Barrier_WR, ProQOL_BO, ProQOL_CS,
         ProQOL_STS, CBarrier_FO, CBarrier_WR, CProQOL_BO, CProQOL_CS,
         CProQOL_STS, Motivation_NeedSANE, Motivation_PersonalConn, Setting) %>%
  filter(Stage_Reached >= 2) %>%
  mutate(Threshold = 2,
         Attrit = if_else(Stage_Reached == Threshold, true = 1, false = 0))
```

```{r}
#| label: create-T3

T3 <- Enrolled_Applicants_CD %>%
  select(ID, Stage_Reached, Barrier_FO, Barrier_WR, ProQOL_BO, ProQOL_CS,
         ProQOL_STS, CBarrier_FO, CBarrier_WR, CProQOL_BO, CProQOL_CS,
         CProQOL_STS, Motivation_NeedSANE, Motivation_PersonalConn, Setting) %>%
  filter(Stage_Reached >= 3) %>%
  mutate(Threshold = 3,
         Attrit = if_else(Stage_Reached == Threshold, true = 1, false = 0))
```

```{r}
#| label: create-value-labels-2

Threshold_Labels <- c("T1 Before DT" = 1,
                      "T2 During DT" = 2,
                      "T3 Before/During CSW" = 3)
```

```{r}
#| label: create-Thresholds

Thresholds <- bind_rows(T1, T2, T3) %>%
  relocate(ID, Threshold, Attrit) %>%
  # Convert categorical predictors to factors with a "." variable name suffix
  mutate(Threshold. = factor(Threshold),
         Setting. = as_factor(Setting),
         Motivation_NeedSANE. = as_factor(Motivation_NeedSANE),
         Motivation_PersonalConn. = as_factor(Motivation_PersonalConn)) %>%
  set_value_labels(Threshold = Threshold_Labels,
                   Attrit = Binary_Labels)
```

Table 10 shows the sizes of datasets created above in terms of number rows and columns, with additional computed columns showing how many records stopped from proceeding to the next stage of case selection and how many continued. This allowed me to double-check that the sizes each made sense. For the `Applicants` row, the stopped cases are the ineligible applicants, while for `Eligible_Applicants` the stopped cases are those that did not enroll. For `Enrolled_Applicants`, the stopped cases are those who lacked complete data on variables required for modeling. For `Enrolled_Applicants_CD`, stopped cases were those who did not start didactic training. `T1` to `T3` each have one row per person, with the stopped cases being those that attrited at the specified threshold. The `Thresholds` dataset contains one row per person per threshold encountered because it combined the rows of `T1` to `T3`, with the stopped cases being all the rows that indicated attrition occurred.

```{r}
#| label: tbl-Dataset-Size
#| tbl-cap: Dataset Sizes

tibble(DataFrame = c("Applicants", "Eligible_Applicants",
                     "Enrolled_Applicants", "Enrolled_Applicants_CD",
                     "T1", "T2", "T3", "Thresholds"),
       N_Rows = c(nrow(Applicants), nrow(Eligible_Applicants),
                  nrow(Enrolled_Applicants), nrow(Enrolled_Applicants_CD),
                  nrow(T1), nrow(T2), nrow(T3), nrow(Thresholds)),
       N_Cols = c(ncol(Applicants), ncol(Eligible_Applicants),
                  ncol(Enrolled_Applicants), ncol(Enrolled_Applicants_CD),
                  ncol(T1), ncol(T2), ncol(T3), ncol(Thresholds)),
       Stopped = c(sum(Applicants$Stage_Raw == 1),
                   sum(Eligible_Applicants$Stage_Raw == 2),
                   sum(is.na(Enrolled_Applicants$ProQOL_BO)),
                   sum(Enrolled_Applicants_CD$Stage_Raw == 3),
                   sum(T1$Attrit), sum(T2$Attrit), sum(T3$Attrit),
                   sum(Thresholds$Attrit))) %>%
  mutate(Continued = N_Rows - Stopped) %>%
  kable(format = "latex", booktabs = TRUE) %>%
  kable_styling()
```

Table 10: Dataset Sizes

DataFrame	N_Rows	N_Cols	Stopped	Continued
Applicants	497	154	170	327
Eligible_Applicants	327	154	73	254
Enrolled_Applicants	254	154	2	252
Enrolled_Applicants_CD	252	159	17	235
T1	252	17	17	235
T2	235	17	61	174
T3	174	17	17	157
Thresholds	661	21	95	566

4 Save Data

```{r}
#| label: create-save-what

# Character vector of object names that will be saved.
save_what <- c("Applicants", "Eligible_Applicants", "Enrolled_Applicants",
               "Enrolled_Applicants_CD", "StartedDT_Applicants", "Thresholds")
```

```{r}
#| label: save-data

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

save(list = save_what, file=DataFile)
```

Table 11 shows meta-data about the data file we just saved.

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

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

Table 11: Meta-Data About the Imported Data File Saved Out

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

> **Tip**
>
> The imported data file can be loaded into R by copying the following code chunk into a script, setting the chunk option `eval: true`, and running the chunk. That will overwrite any objects already in memory that have the same names as the objects being loaded.

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

load(file=here("data/Imported_SANETP_Data.RData"))
```
```

5 References

Cole, S. R., & Ananth, C. V. (2001). Regression models for unconstrained, partially or fully constrained continuation odds ratios. *International Journal of Epidemiology*, *30*(6), 1379–1382. https://doi.org/10.1093/ije/30.6.1379

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.

Fullerton, A. S., & Xu, J. (2016). *Ordered regression models: Parallel, partial, and non-parallel alternatives*. Chapman & Hall/CRC Press. https://doi.org/10.1201/b20060

Liu, X., & Bai, H. (2020). Forward and backward continuation ratio models for ordinal response variables. *Journal of Modern Applied Statistical Methods*, *18*(2). https://doi.org/10.22237/jmasm/1604190180

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

6 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: `Import_Data.qmd`
- Output file: `Import_Data_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.

6.1 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/RtmpayCIjR/file3fa038f2361a -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)
 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)
 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)
 ellipsis       0.3.2    2021-04-29 [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)
 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)
 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)
 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)
 mvtnorm        1.3-3    2025-01-10 [1] CRAN (R 4.5.0)
 otel           0.2.0    2025-08-29 [1] CRAN (R 4.5.1)
 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)
 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)
 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)
 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)
 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)
 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)
 yaml           2.3.12   2025-12-10 [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.

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

6.2 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/Import_Data.rmarkdown
    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!