

CTN-0094 Demographics Analysis

Harshi

Quarto

Quarto enables you to weave together content and executable code into a finished document. To learn more about Quarto see <https://quarto.org>.

Running Code

When you click the **Render** button a document will be generated that includes both content and the output of embedded code. You can embed code like this:

```
1 + 1
```

```
[1] 2
```

You can add options to executable code like this

```
[1] 4
```

The `echo: false` option disables the printing of code (only output is displayed).

Introduction

- Illicit opioid use is a major public health crisis (OpenAI, 2025).
- Medications for opioid use disorder reduce mortality.
- CTN-0094 harmonizes data from CTN-0027, CTN-0030, and CTN-0051.
- This report summarizes baseline demographics across projects.

Methods

Software versions

- **R:** R version 4.5.2 (2025-10-31)
 - **public.ctn0094data:** 1.0.6
 - **tidyverse:** 2.0.0
 - **gtsummary:** 2.4.0
-

Load & Prepare Data

```
library(tidyverse)

-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.1     v stringr   1.5.2
v ggplot2   4.0.0     v tibble    3.3.0
v lubridate 1.9.4     v tidyr    1.3.1
v purrr     1.1.0
-- 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 non-conflicting
```

```
library(public.ctn0094data)
library(gtsummary)
library(labelled)
library(glue)
library(scales)
```

Attaching package: 'scales'

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

discard

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

```
col_factor

# Load data
demographics <- demographics
randomization <- randomization

# Raw size
n_raw <- nrow(demographics)

# First randomization per person (earliest date)
first_rand <- randomization |>
  arrange(who, when) |>
  group_by(who) |>
  slice(1) |>
  ungroup() |>
  mutate(
    project = case_when(
      which == 27 ~ "CTN-0027",
      which == 30 ~ "CTN-0030",
      which == 51 ~ "CTN-0051",
      TRUE ~ paste0("CTN-", which)
    )
  ) |>
  select(who, project)

# Join project numbers
demo_joined <- demographics |>
  left_join(first_rand, by = "who")

# Missing before filtering
n_missing_age <- demo_joined |> filter(is.na(age)) |> nrow()
n_missing_sex <- demo_joined |> filter(is.na(is_male)) |> nrow()
n_missing_project <- demo_joined |> filter(is.na(project)) |> nrow()

# Final analysis dataset
analysis <- demo_joined |>
  filter(!is.na(age), !is.na(is_male), !is.na(project)) |>
  mutate(
    is_male_factor = factor(if_else(is_male == 1, "Male", "Not Male")),
    project = factor(project)
```

```

)

n_analysis <- nrow(analysis)
n_excluded <- n_raw - n_analysis

# Theme
ctn_theme <- theme_minimal(base_size = 14) +
  theme(
    plot.title = element_text(face = "bold"),
    axis.title = element_text(face = "bold")
  )

# Sex proportions
sex_by_project <- analysis |>
  group_by(project) |>
  summarise(prop_male = mean(is_male_factor == "Male"), .groups = "drop")

max_diff <- ifelse(nrow(sex_by_project) > 1,
                    max(sex_by_project$prop_male) - min(sex_by_project$prop_male),
                    NA)

max_diff_fmt <- ifelse(is.na(max_diff), "Not applicable", percent(max_diff, accuracy = 0.1))

# Chi-square only if valid
if (length(unique(analysis$project)) > 1 && length(unique(analysis$is_male_factor)) > 1) {
  chi_table <- table(analysis$project, analysis$is_male_factor)
  chi_test <- chisq.test(chi_table)
  p_chi_fmt <- scales::pvalue(chi_test$p.value)
} else {
  p_chi_fmt <- "Cannot compute - insufficient groups"
}

```

Handling Missing Data

Of the 3560 participants:

- 208 missing age

Characteristic	CTN-1 N = 2,492 ¹	p-value ²
age	33 (26, 43)	
is_male_factor		
Not Male	2,492 (100%)	

¹Median (Q1, Q3); n (%)
²NA

- 4 missing sex
 - 1068 missing project assignment
 - **1068 were excluded** from final analysis
-

Table 1 — Demographics Summary

```
analysis |>
  select(age, is_male_factor, project) |>
 tbl_summary(by = project) |>
  add_p()
```

The following errors were returned during `add_p()`:

- x For variable `age` (`project`) and "statistic", "p.value", and "parameter" statistics: all observations are in the same group
- x For variable `is_male_factor` (`project`) and "statistic", "p.value", and "parameter" statistics: 'x' and 'y' must have at least 2 levels

Results

Sex differences across projects

- Maximum difference in male proportion: **Not applicable**
 - Chi-square result: **Cannot compute — insufficient groups**
-

Figure 1 — Age Histogram

```
ggplot(analysis, aes(age, fill = project)) +  
  geom_histogram(alpha = 0.5, bins = 30, position = "identity") +  
  facet_wrap(~project) +  
  ctn_theme +  
  labs(title = "Age Distribution by Project", x = "Age", y = "Count")
```

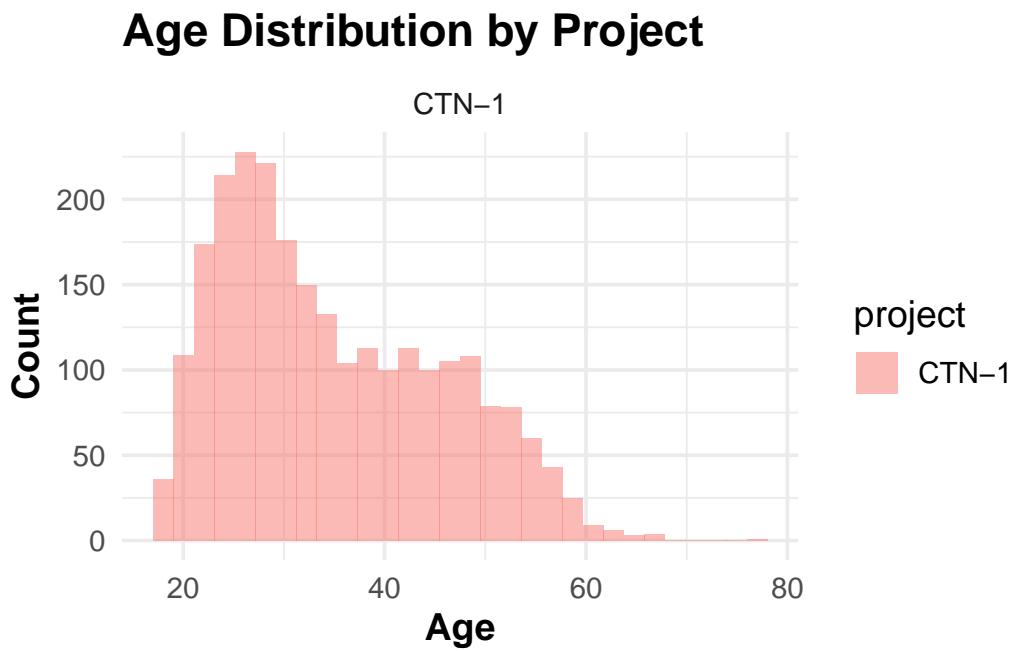


Figure 2 — Age Boxplot

```
ggplot(analysis, aes(project, age, fill = project)) +  
  geom_boxplot() +  
  ctn_theme +  
  labs(title = "Age by Project (Boxplot)", x = "Project", y = "Age")
```

Age by Project (Boxplot)

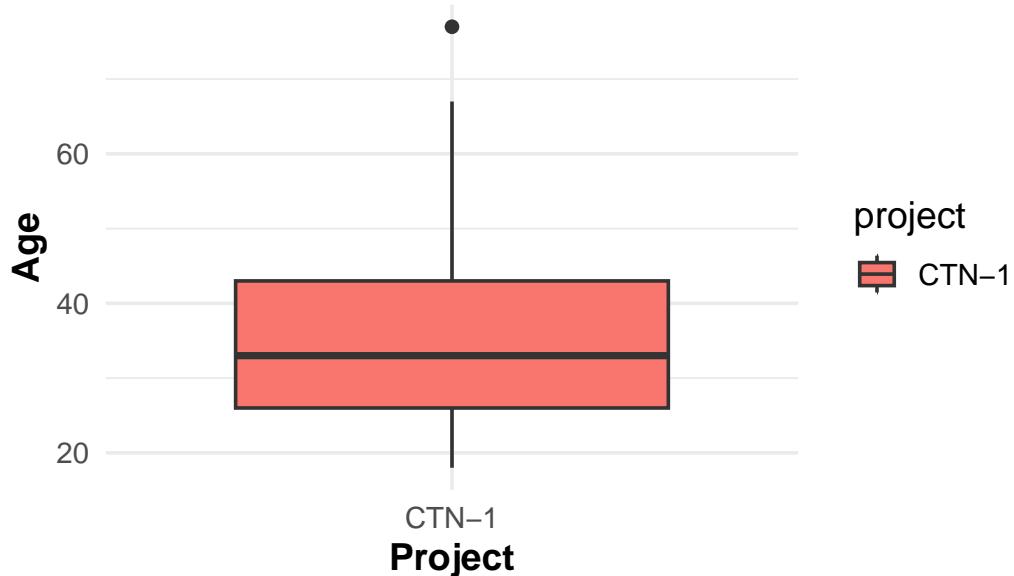


Figure 3 — Age Violin Plot

```
ggplot(analysis, aes(project, age, fill = project)) +  
  geom_violin() +  
  ctn_theme +  
  labs(title = "Age by Project (Violin Plot)", x = "Project", y = "Age")
```

Age by Project (Violin Plot)

