Water Treatment Manuscript

Alex Thawani

library(tidyverse)

── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
✔ dplyr 1.1.4 ✔ readr 2.1.5  
✔ forcats 1.0.0 ✔ stringr 1.5.1  
✔ ggplot2 3.5.1 ✔ tibble 3.2.1  
✔ lubridate 1.9.4 ✔ tidyr 1.3.1  
✔ purrr 1.0.2   
── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
✖ dplyr::filter() masks stats::filter()  
✖ dplyr::lag() masks stats::lag()  
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(gtsummary)

df <- read\_csv("/Users/alex/Documents/GitHub/R\_Projects/Cholera\_study\_project\_2024/water\_treatment.csv",show\_col\_types = FALSE)

New names:  
• `` -> `...1`

## Demographics

df |> select("age\_group","sex","level\_education","ethnicity","religion","water\_treatment\_category") |>   
 tbl\_summary(  
 by = water\_treatment\_category  
 ) |>   
 add\_ci() |>   
 add\_p()

The following errors were returned during `add\_p()`:  
✖ For variable `ethnicity` (`water\_treatment\_category`) and "estimate",  
 "p.value", "conf.low", and "conf.high" statistics: FEXACT error 7(location).  
 LDSTP=18510 is too small for this problem, (pastp=73.7218,  
 ipn\_0:=ipoin[itp=13]=93, stp[ipn\_0]=72.3388). Increase workspace or consider  
 using 'simulate.p.value=TRUE'

| **Characteristic** | **Does Not Practice** N = 301*1* | **95% CI***2* | **Practices** N = 494*1* | **95% CI***2* | **p-value***3* |
| --- | --- | --- | --- | --- | --- |
| age\_group |  |  |  |  | 0.3 |
| >50 | 63 (21%) | 17%, 26% | 111 (22%) | 19%, 26% |  |
| 18-30 | 119 (40%) | 34%, 45% | 170 (34%) | 30%, 39% |  |
| 31-50 | 119 (40%) | 34%, 45% | 213 (43%) | 39%, 48% |  |
| sex |  |  |  |  | <0.001 |
| female | 209 (69%) | 64%, 75% | 398 (81%) | 77%, 84% |  |
| male | 92 (31%) | 25%, 36% | 96 (19%) | 16%, 23% |  |
| level\_education |  |  |  |  | 0.005 |
| none | 54 (18%) | 14%, 23% | 54 (11%) | 8.4%, 14% |  |
| primary | 191 (63%) | 58%, 69% | 308 (62%) | 58%, 67% |  |
| secondary | 49 (16%) | 12%, 21% | 121 (24%) | 21%, 29% |  |
| tertiary | 7 (2.3%) | 1.0%, 4.9% | 11 (2.2%) | 1.2%, 4.1% |  |
| ethnicity |  |  |  |  |  |
| chewa | 161 (53%) | 48%, 59% | 231 (47%) | 42%, 51% |  |
| lomwe | 25 (8.3%) | 5.6%, 12% | 40 (8.1%) | 5.9%, 11% |  |
| ngonde | 2 (0.7%) | 0.12%, 2.6% | 2 (0.4%) | 0.07%, 1.6% |  |
| ngoni | 38 (13%) | 9.2%, 17% | 63 (13%) | 10%, 16% |  |
| nyanja | 6 (2.0%) | 0.81%, 4.5% | 8 (1.6%) | 0.75%, 3.3% |  |
| other\_enthni | 3 (1.0%) | 0.26%, 3.1% | 6 (1.2%) | 0.49%, 2.8% |  |
| sena | 0 (0%) | 0.00%, 1.6% | 3 (0.6%) | 0.16%, 1.9% |  |
| tonga | 10 (3.3%) | 1.7%, 6.2% | 14 (2.8%) | 1.6%, 4.8% |  |
| tumbuka | 6 (2.0%) | 0.81%, 4.5% | 5 (1.0%) | 0.37%, 2.5% |  |
| yao | 50 (17%) | 13%, 21% | 122 (25%) | 21%, 29% |  |
| religion |  |  |  |  | 0.028 |
| Christianity | 233 (77%) | 72%, 82% | 359 (73%) | 68%, 77% |  |
| muslim | 55 (18%) | 14%, 23% | 124 (25%) | 21%, 29% |  |
| others\_regligion | 13 (4.3%) | 2.4%, 7.5% | 11 (2.2%) | 1.2%, 4.1% |  |
| *1*n (%) | | | | | |
| *2*CI = Confidence Interval | | | | | |
| *3*Pearson's Chi-squared test | | | | | |

## Water treatment Practice

practice <- c(  
 "health\_risk\_associated\_with\_drinking\_untreated\_water",  
 "treat\_drinking\_water",  
 "difficulties\_accessing\_water\_treatment\_products",  
 "often\_treat\_water",  
 "drinking\_water\_treated",  
 "traditional\_beliefs\_binary",  
 "water\_treatment\_category")

df|> select(practice) |>   
 tbl\_summary(  
 by = water\_treatment\_category,  
 missing = "no",  
 type = list(all\_categorical() ~ "categorical")  
 )|>   
 add\_stat\_label() |>  
 add\_ci() |>   
 add\_p()

Warning: Using an external vector in selections was deprecated in tidyselect 1.1.0.  
ℹ Please use `all\_of()` or `any\_of()` instead.  
 # Was:  
 data %>% select(practice)  
  
 # Now:  
 data %>% select(all\_of(practice))  
  
See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.

| **Characteristic** | **Does Not Practice** N = 301 | **95% CI***1* | **Practices** N = 494 | **95% CI***1* | **p-value***2* |
| --- | --- | --- | --- | --- | --- |
| health\_risk\_associated\_with\_drinking\_untreated\_water, n (%) |  |  |  |  | <0.001 |
| no | 57 (19%) | 15%, 24% | 22 (4.5%) | 2.9%, 6.8% |  |
| yes | 244 (81%) | 76%, 85% | 472 (96%) | 93%, 97% |  |
| treat\_drinking\_water, n (%) |  |  |  |  | <0.001 |
| no | 288 (96%) | 93%, 98% | 11 (2.2%) | 1.2%, 4.1% |  |
| yes | 13 (4.3%) | 2.4%, 7.5% | 483 (98%) | 96%, 99% |  |
| difficulties\_accessing\_water\_treatment\_products, n (%) |  |  |  |  | 0.005 |
| no | 2 (15%) | 2.7%, 46% | 265 (55%) | 50%, 59% |  |
| yes | 11 (85%) | 54%, 97% | 218 (45%) | 41%, 50% |  |
| often\_treat\_water, n (%) |  |  |  |  | 0.019 |
| Always | 5 (38%) | 15%, 68% | 276 (57%) | 53%, 62% |  |
| donot\_know | 1 (7.7%) | 0.40%, 38% | 2 (0.4%) | 0.07%, 1.7% |  |
| Never | 1 (7.7%) | 0.40%, 38% | 5 (1.0%) | 0.38%, 2.5% |  |
| Sometimes | 6 (46%) | 20%, 74% | 200 (41%) | 37%, 46% |  |
| drinking\_water\_treated, n (%) |  |  |  |  | <0.001 |
| 0 | 288 (96%) | 93%, 98% | 11 (2.2%) | 1.2%, 4.1% |  |
| 1 | 13 (4.3%) | 2.4%, 7.5% | 483 (98%) | 96%, 99% |  |
| traditional\_beliefs\_binary, n (%) |  |  |  |  | 0.9 |
| 0 | 31 (47%) | 35%, 60% | 75 (48%) | 40%, 56% |  |
| 1 | 35 (53%) | 40%, 65% | 81 (52%) | 44%, 60% |  |
| *1*CI = Confidence Interval | | | | | |
| *2*Pearson's Chi-squared test; Fisher's exact test | | | | | |