STA 540 Case #1

2026-01-25

```r
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   4.0.1     ✔ 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
```

```r
library(dplyr)
library(stringr)
library(emmeans)
```

```
## Welcome to emmeans.
## Caution: You lose important information if you filter this package's results.
## See '? untidy'
```

```r
library(tidyr)
library(stringr)
library(purrr)
library(readr)
library(tibble)
library(gt)
library(ggplot2)
```

Filtering to get 254 participants

```r
df <- read_csv("/Users/josephong/Downloads/CTN_FINAL.csv", show_col_types = FALSE)

df_analysis <- df |>
  filter(PO_FLAG == "Include") |>
  filter(is.na(WAVE) | WAVE != 3)
```

Replicating Table 1

```r
N_total <- nrow(df_analysis)

# Median (IQR) formatter
med_iqr <- function(x, digits = 0) {
  med <- median(x, na.rm = TRUE)
  q1  <- quantile(x, 0.25, na.rm = TRUE)
  q3  <- quantile(x, 0.75, na.rm = TRUE)
  paste0(round(med, digits), " (", round(q1, digits), "-", round(q3, digits), ")")
}

# Age
age_tbl <- tibble(
  Characteristic = "Age in years, median (IQR)",
  Value = med_iqr(df_analysis$Q3_1, digits = 0)
)

# Ethnicity
eth_n <- sum(df_analysis$Q5_1 == 1, na.rm = TRUE)
eth_pct <- round(100 * eth_n / N_total, 1)

eth_tbl <- tibble(
  Characteristic = c("Ethnicity, n (%)", "  Hispanic/Latinx"),
  Value = c(NA_character_, paste0(eth_n, " (", eth_pct, ")"))
)

# Race
df_race <- df_analysis |>
  mutate(
    race = case_when(
      str_length(as.character(Q5_3)) > 2 ~ "Multiracial",
      Q5_3 == 25 ~ "American Indian or Alaskan Native",
      Q5_3 == 24 ~ "Black or African American",
      Q5_3 == 23 ~ "White",
      Q5_3 == 28 ~ "Other",
      TRUE       ~ NA_character_
    )
  )

race_tbl <- df_race |>
  filter(!is.na(race)) |>
  count(race) |>
  mutate(
    pct = round(100 * n / sum(n), 1),
    Value = paste0(n, " (", pct, ")")
  ) |>
  arrange(match(
    race,
    c("American Indian or Alaskan Native",
      "Black or African American",
      "White",
      "Other",
      "Multiracial")
  )) |>
  transmute(
    Characteristic = paste0("  ", race),
    Value
  )

race_header <- tibble(
  Characteristic = "Race, n (%)",
  Value = NA_character_
)

race_tbl <- bind_rows(race_header, race_tbl)

# History of PrEP uptake
prep_tbl <- df_analysis |>
  mutate(
    prep_history = case_when(
      Q6_2 == 3 ~ "Never taken PrEP",
      Q6_2 == 2 ~ "In the past 6 months",
      Q6_2 == 1 ~ "Taken PrEP >6 months ago",
      TRUE ~ NA_character_
    )
  ) |>
  count(prep_history) |>
  mutate(
    pct = round(100 * n / sum(n), 1),
    Value = paste0(n, " (", pct, ")")
  ) |>
  arrange(match(
    prep_history,
    c("Never taken PrEP",
      "In the past 6 months",
      "Taken PrEP >6 months ago")
  )) |>
  transmute(
    Characteristic = paste0("  ", prep_history),
    Value
  )

prep_header <- tibble(
  Characteristic = "History of PrEP uptake, n (%)",
  Value = NA_character_
)

prep_tbl <- bind_rows(prep_header, prep_tbl)

# Number of male sex partners
partners_tbl <- tibble(
  Characteristic = "Number of male sex partners in the past 90 days, median (IQR)",
  Value = med_iqr(df_analysis$Q11_2, digits = 0)
)

# Condom use
condom_tbl <- df_analysis |>
  mutate(
    condom_use = case_when(
      Q11_3 == 1 ~ "Never",
      Q11_3 == 2 ~ "Sometimes",
      Q11_3 == 3 ~ "About half the time",
      Q11_3 == 4 ~ "Most of the time",
      Q11_3 == 5 ~ "Always",
      TRUE ~ NA_character_
```

```r
      )
    ) |>
    count(condom_use) |>
    mutate(
      pct = round(100 * n / sum(n), 1),
      Value = paste0(n, " (", pct, ")")
    ) |>
    arrange(match(
      condom_use,
      c("Never", "Sometimes", "About half the time", "Most of the time", "Always")
    )) |>
    transmute(
      Characteristic = paste0("  ", condom_use),
      Value
    )

condom_header <- tibble(
  Characteristic = "Condom use, n (%)",
  Value = NA_character_
)

condom_tbl <- bind_rows(condom_header, condom_tbl)

# Condomless receptive anal sex
cra_n <- sum(df_analysis$Q11_4 == 1, na.rm = TRUE)
cra_pct <- round(100 * cra_n / N_total, 1)

cra_tbl <- tibble(
  Characteristic = "Condomless receptive anal sex in the past 90 days, n (%)",
  Value = paste0(cra_n, " (", cra_pct, ")")
)

# Ever tested HIV
ever_n <- sum(df_analysis$Q11_5 == 1, na.rm = TRUE)
ever_pct <- round(100 * ever_n / N_total, 1)

ever_tbl <- tibble(
  Characteristic = "Ever tested for HIV during lifetime, n (%)",
  Value = paste0(ever_n, " (", ever_pct, ")")
)

# Months since last HIV test
months_tbl_val <- df_analysis |>
  filter(LAST_HIV_TEST_MONTHS != 9999) |>
  summarise(Value = med_iqr(LAST_HIV_TEST_MONTHS, digits = 0)) |>
  pull(Value)

months_tbl <- tibble(
  Characteristic = c("If tested for HIV, median (IQR)", "  Months since last HIV test"),
  Value          = c(NA_character_, months_tbl_val)
)

# If not tested
not_tested_n <- sum(df_analysis$Q11_5 == 2, na.rm = TRUE)
not_tested_pct <- round(100 * not_tested_n / N_total, 1)

not_tested_tbl <- tibble(
  Characteristic = "If not tested for HIV, n (%)",
  Value = paste0(not_tested_n, " (", not_tested_pct, ")")
)

# Reasons among not tested
df_not_tested <- df_analysis |>
  filter(Q11_5 == 2)

reasons_tbl <- df_not_tested |>
  mutate(
    reason = case_when(
      Q11_7 == 1 ~ "Unlikely to be exposed to HIV",
      Q11_7 == 2 ~ "Afraid of testing HIV-positive",
      Q11_7 == 3 ~ "Did not want to think about HIV/HIV-positive",
      Q11_7 == 4 ~ "Worried about names being reported if positive",
      Q11_7 == 5 ~ "Dislike for needles",
      Q11_7 == 6 ~ "Unable to trust that the results will be confidential",
      Q11_7 == 7 ~ "Unaware of where to get tested",
      Q11_7 == 8 ~ "Other reasons",
      TRUE ~ NA_character_
    )
  ) |>
  count(reason) |>
  mutate(
    pct = round(100 * n / sum(n), 1),
    Value = paste0(n, " (", pct, ")")
  ) |>
  arrange(match(
    reason,
    c(
      "Unlikely to be exposed to HIV",
      "Afraid of testing HIV-positive",
      "Did not want to think about HIV/HIV-positive",
      "Worried about names being reported if positive",
      "Dislike for needles",
      "Unable to trust that the results will be confidential",
      "Unaware of where to get tested",
      "Other reasons"
    )
  )) |>
  transmute(
    Characteristic = paste0("  ", reason),
    Value
  )

reasons_header <- tibble(
  Characteristic = "Main reasons cited by the 63 participants for not getting tested, n (%)",
  Value = NA_character_
)

reasons_tbl <- bind_rows(reasons_header, reasons_tbl)

# Combining into Table 1
table1 <- bind_rows(
  age_tbl,
  eth_tbl,
  race_tbl,
  prep_tbl,
  partners_tbl,
```

```
    condom_tbl,
    cra_tbl,
    ever_tbl,
    months_tbl,
    not_tested_tbl,
    reasons_tbl
)

# -------------------------------------------------------------------------------
# Everything below is to make Table 1 look pretty now

table1_gt <- table1 |>
    mutate(
        is_header = is.na(Value) | Value == "",
        indent = str_detect(Characteristic, "^\\s{2,}"),
        Characteristic_clean = str_replace(Characteristic, "^\\s+", "")
    ) |>
    select(Characteristic_clean, Value, is_header, indent) |>   # keep only what gt needs
    gt() |>
    sub_missing(columns = Value, missing_text = "") |>
    cols_label(
        Characteristic_clean = "Characteristic",
        Value = "Value"
    ) |>
    cols_width(
        Characteristic_clean ~ pct(70),
        Value ~ pct(30)
    ) |>
    tab_header(
        title = "Table 1.",
        subtitle = md("Summary of National Institute on Drug Abuse Clinical Trials Network Social Media PrEP Study, 2
020 (N = 254)")
    ) |>
    # bold section headers (use the logical column inside gt data)
    tab_style(
        style = cell_text(weight = "bold"),
        locations = cells_body(
            rows = is_header,
            columns = Characteristic_clean
        )
    ) |>
    # indent sub-rows (also uses logical col inside gt data)
    tab_style(
        style = cell_text(indent = px(18)),
        locations = cells_body(
            rows = indent,
            columns = Characteristic_clean
        )
    ) |>
    # horizontal rules
    tab_style(
        style = cell_borders(sides = "bottom", color = "grey80", weight = px(1)),
        locations = cells_body(rows = TRUE)
    ) |>
    # hide helper columns
    cols_hide(columns = c(is_header, indent)) |>
    opt_table_outline() |>
    opt_row_striping() |>
    tab_options(
        table.font.size = px(12),
        heading.title.font.size = px(14),
        heading.subtitle.font.size = px(11),
        data_row.padding = px(4)
    )

table1_gt
```

Table 1.	
Summary of National Institute on Drug Abuse Clinical Trials Network Social Media PrEP Study, 2020 (N = 254)	
Characteristic	**Value**
Age in years, median (IQR)	25 (21-27)
Ethnicity, n (%)	
Hispanic/Latinx	66 (26)
Race, n (%)	
American Indian or Alaskan Native	1 (0.4)
Black or African American	196 (78.4)
White	28 (11.2)
Other	14 (5.6)
Multiracial	11 (4.4)
History of PrEP uptake, n (%)	
Never taken PrEP	232 (91.3)
Taken PrEP >6 months ago	22 (8.7)
Number of male sex partners in the past 90 days, median (IQR)	4 (3-6)
Condom use, n (%)	
Never	36 (14.2)
Sometimes	108 (42.5)
About half the time	37 (14.6)
Most of the time	68 (26.8)
Always	5 (2)
Condomless receptive anal sex in the past 90 days, n (%)	210 (82.7)
Ever tested for HIV during lifetime, n (%)	191 (75.2)
If tested for HIV, median (IQR)	
Months since last HIV test	11 (6-21)
If not tested for HIV, n (%)	63 (24.8)
Main reasons cited by the 63 participants for not getting tested, n (%)	
Unlikely to be exposed to HIV	8 (12.7)
Afraid of testing HIV-positive	26 (41.3)
Did not want to think about HIV/HIV-positive	8 (12.7)
Worried about names being reported if positive	3 (4.8)
Dislike for needles	5 (7.9)
Unable to trust that the results will be confidential	3 (4.8)
Unaware of where to get tested	7 (11.1)
Other reasons	3 (4.8)

Primary Analysis

```r
# Map to actual site names (within-wave)
df_pp <- df_analysis |>
  mutate(
    ordered60 = as.integer(str_trim(ORA_REDEEMED) == "Yes"),
    wave2 = case_when(
      WAVE %in% c(1, 4) ~ 1L,
      WAVE == 2 ~ 2L,
      TRUE ~ NA_integer_
    ),

    SITE = str_replace_all(SITE, "Jack'd", "Jack'd"),
    SITE_clean = str_squish(str_to_lower(SITE))
  ) |>
  filter(!is.na(wave2)) |>
  mutate(
    SITE = str_replace_all(SITE, "Jack'd", "Jack'd"),
    wave_label = if_else(wave2 == 1L, "1a", "2"),
    t_days = if_else(wave2 == 1L, 70L, 38L),
    platform_group = case_when(
      SITE %in% c("Facebook", "Instagram") ~ "Social media site",
      SITE %in% c("Grindr", "Jack'd") ~ "Dating app",
      SITE %in% c("Google", "Bing") ~ "Information search site",
      TRUE ~ NA_character_
    )
  ) |>
  filter(!is.na(platform_group))


df_sites <- df_pp |>
  mutate(
    Wave = factor(wave2, levels = c(1, 2), labels = c("Wave 1*", "Wave 2")),
    Site = case_when(
      Wave == "Wave 1*" & str_detect(SITE_clean, "facebook") ~ "Facebook",
      Wave == "Wave 1*" & str_detect(SITE_clean, "grindr")   ~ "Grindr",
      Wave == "Wave 1*" & str_detect(SITE_clean, "google")   ~ "Google",
      Wave == "Wave 2"  & str_detect(SITE_clean, "instagram") ~ "Instagram",
      Wave == "Wave 2"  & str_detect(SITE_clean, "jack")      ~ "Jack'd",
      Wave == "Wave 2"  & str_detect(SITE_clean, "bing")      ~ "Bing",
      TRUE ~ NA_character_
    )
  ) |>
  filter(!is.na(Site))

# Aggregate the counts per site within wave
site_long <- df_sites |>
  group_by(Wave, Site) |>
  summarise(
    o = sum(ordered60, na.rm = TRUE),
    t_days = first(t_days),
    .groups = "drop"
  )

analyze_one_wave_pretty <- function(wave_name, allowed_sites, days_exposure) {

  dat <- site_long |>
    filter(Wave == wave_name) |>
    mutate(Site = factor(Site, levels = allowed_sites)) |>
    complete(Site, fill = list(o = 0)) |>
    mutate(
      Wave = wave_name,
      t_days = days_exposure
    )

  # Fitting the formula
  fit <- glm(
    o ~ Site + offset(log(t_days)),
    family = poisson(link = "log"),
    data = dat
  )

  # Rates/day (t_days = 1 => offset 0)
  emm_rates <- emmeans(
    fit,
    ~ Site,
    at = list(t_days = 1),
    type = "response"
  )

  rates_tbl <- as.data.frame(emm_rates) |>
    transmute(
      Site,
      rate_per_day = rate,
      lower_95 = asymp.LCL,
      upper_95 = asymp.UCL
    ) |>
    left_join(
      dat |> select(Site, o, t_days),
      by = "Site"
    ) |>
    mutate(
      Wave = wave_name,
      raw_rate_per_day = o / t_days
    ) |>
    select(
      Wave, Site, o, t_days,
      raw_rate_per_day,
      rate_per_day, lower_95, upper_95
    ) |>
    arrange(Site)

  # Pairwise comparisons
  emm_link <- emmeans(fit, ~ Site, at = list(t_days = 1))  # link scale = log(rate)

  # Create all pairwise contrasts but then flip any "X - Bing" to "Bing - X"
  pw_tbl <- as.data.frame(pairs(emm_link)) |>
    mutate(
      Wave = wave_name,

      # If contrast is "X - Bing", flip to "Bing - X" by multiplying estimate by -1
      flip = str_detect(contrast, " - Bing$"),

      contrast = if_else(
        flip,
        str_replace(contrast, "^(.*) - Bing$", "Bing - \\1"),
```

```
        contrast
    ),

    estimate = if_else(flip, -estimate, estimate),
    z.ratio  = if_else(flip, -z.ratio,  z.ratio),

    # p-value stays the same under sign flip
    rate_ratio = exp(estimate),

    # BH adjustment within wave
    p_adj_BH = p.adjust(p.value, method = "BH")
  ) |>
  transmute(
    Wave,
    contrast,
    rate_ratio,
    SE,
    z.ratio,
    p_value = p.value,
    p_adj_BH
  )

# Pretty gt tables
rates_gt <- rates_tbl |>
  gt() |>
  tab_header(title = paste0("Primary outcome rates per day — ", wave_name)) |>
  fmt_number(columns = c(raw_rate_per_day, rate_per_day, lower_95, upper_95), decimals = 3) |>
  cols_label(
    o = "Orders (o)",
    t_days = "Days (t)",
    raw_rate_per_day = "Raw rate (o/t)",
    rate_per_day = "Model rate/day",
    lower_95 = "95% LCL",
    upper_95 = "95% UCL"
  )

pw_gt <- pw_tbl |>
  gt() |>
  tab_header(title = paste0("Pairwise site comparisons — ", wave_name)) |>
  fmt_number(columns = c(rate_ratio), decimals = 3) |>
  fmt_number(columns = c(p_value, p_adj_BH), decimals = 3) |>
  cols_label(
    contrast = "Comparison",
    rate_ratio = "Rate ratio",
    p_value = "p-value",
    p_adj_BH = "BH-adjusted p"
  )

list(
  data_used = dat,
  fit = fit,
  rates_tbl = rates_tbl,
  pairwise_tbl = pw_tbl,
  rates_gt = rates_gt,
  pairwise_gt = pw_gt
)
}

res_w1 <- analyze_one_wave_pretty(
  wave_name = "Wave 1*",
  allowed_sites = c("Facebook", "Google", "Grindr"),
  days_exposure = 70
)

res_w2 <- analyze_one_wave_pretty(
  wave_name = "Wave 2",
  allowed_sites = c("Instagram", "Jack'd", "Bing"),
  days_exposure = 38
)

# Show tables
res_w1$rates_gt
```

Primary outcome rates per day — Wave 1*

Wave	Site	Orders (o)	Days (t)	Raw rate (o/t)	Model rate/day	95% LCL	95% UCL
Wave 1*	Facebook	13	70	0.186	0.186	0.108	0.320
Wave 1*	Google	17	70	0.243	0.243	0.151	0.391
Wave 1*	Grindr	9	70	0.129	0.129	0.067	0.247

```
res_w1$pairwise_gt
```

Pairwise site comparisons — Wave 1*

Wave	Comparison	Rate ratio	SE	z.ratio	p-value	BH-adjusted p
Wave 1*	Facebook - Google	0.765	0.3684381	-0.7281115	0.747	0.747
Wave 1*	Facebook - Grindr	1.444	0.4336291	0.8480169	0.673	0.747
Wave 1*	Google - Grindr	1.889	0.4122313	1.5427959	0.271	0.747

```
res_w2$rates_gt
```

Primary outcome rates per day — Wave 2

Wave	Site	Orders (o)	Days (t)	Raw rate (o/t)	Model rate/day	95% LCL	95% UCL
Wave 2	Instagram	13	38	0.342	0.342	0.199	0.589
Wave 2	Jack'd	125	38	3.289	3.289	2.761	3.920
Wave 2	Bing	0	38	0.000	0.000	0.000	Inf

```
res_w2$pairwise_gt
```

Pairwise site comparisons — Wave 2

Wave	Comparison	Rate ratio	SE	z.ratio	p-value	BH-adjusted p
Wave 2	Instagram - Jack'd	0.104	2.914156e-01	-7.7667909566	0.000	0.000
Wave 2	Bing - Instagram	0.000	4.224717e+04	-0.0005886202	1.000	1.000
Wave 2	Bing - Jack'd	0.000	4.224717e+04	-0.0006421945	1.000	1.000

Secondary Outcome Analysis

Table (a)

```r
df254 <- df_analysis |>
  mutate(
    ORA_REDEEMED = str_trim(ORA_REDEEMED),
    ordered_grp = case_when(
      ORA_REDEEMED == "Yes" ~ "Ordered test kit",
      ORA_REDEEMED %in% c("No", "Over 60 days") ~ "Did not order test kit",
      TRUE ~ NA_character_
    )
  ) |>
  filter(!is.na(ordered_grp))

# Helper: build one substance block
make_substance_block <- function(data, name, levels_vec) {

  counts <- data |>
    filter(!is.na(category)) |>
    count(ordered_grp, category, name = "n") |>
    complete(
      ordered_grp = c("Ordered test kit", "Did not order test kit"),
      category = levels_vec,
      fill = list(n = 0)
    ) |>
    group_by(ordered_grp) |>
    mutate(
      N = sum(n),
      nN  = sprintf("%d/%d", n, N),
      pct = sprintf("%.1f", 100 * n / N)
    ) |>
    ungroup()

  tab_mat <- counts |>
    select(ordered_grp, category, n) |>
    pivot_wider(names_from = category, values_from = n, values_fill = 0) |>
    arrange(match(ordered_grp, c("Ordered test kit", "Did not order test kit"))) |>
    select(-ordered_grp) |>
    as.matrix()

  pval_fmt <- sub("^0", "", sprintf("%.3f", fisher.test(tab_mat)$p.value))

  counts |>
    pivot_wider(
      names_from = ordered_grp,
      values_from = c(nN, pct),
      names_glue = "{.value}_{ordered_grp}"
    ) |>
    mutate(
      Substance = name,
      `P-value` = ifelse(row_number() == 1, pval_fmt, "")
    ) |>
    transmute(
      Substance,
      Category = category,
      `Ordered test kit n/N` = `nN_Ordered test kit`,
      `Ordered test kit %`   = `pct_Ordered test kit`,
      `Did not order test kit n/N` = `nN_Did not order test kit`,
      `Did not order test kit %`   = `pct_Did not order test kit`,
      `P-value`
    )
}

# Count 1 responses across columns
ones_count <- function(df, cols) {
  rowSums(df[, cols, drop = FALSE] == 1, na.rm = TRUE)
}

# Defining substances in one place
spec <- list(
  Alcohol = list(
    use = "Q13_1",
    cols = c("Q13_1","Q13_2","Q13_3","Q13_4"),
    levels = c("None", "Problem use", "High Risk Substance Use"),
    classify = function(use, score, missing) case_when(
      missing ~ NA_character_,
      !use ~ "None",
      score >= 2 ~ "High Risk Substance Use",
      score == 1 ~ "Problem use",
      TRUE ~ "None"
    )
  ),
  Cannabis = list(
    use = "Q13_5",
    cols = c("Q13_5","Q13_6","Q13_7"),
    levels = c("None", "Problem use", "High Risk Substance Use"),
    classify = function(use, score, missing) case_when(
      missing ~ NA_character_,
      !use ~ "None",
      score >= 2 ~ "High Risk Substance Use",
      score == 1 ~ "Problem use",
      TRUE ~ "None"
    )
  ),
  Stimulants = list(
    use = "Q13_8",
    cols = c("Q13_8","Q13_9","Q13_10"),
    levels = c("None", "Problem Use/High Risk Substance Use"),
    classify = function(use, score, missing) case_when(
      missing ~ NA_character_,
      !use ~ "None",
      score >= 1 ~ "Problem Use/High Risk Substance Use",
      TRUE ~ "None"
    )
  ),
  Opioid = list(
    use = "Q13_11",
    cols = c("Q13_11","Q13_12","Q13_13","Q13_14","Q13_15","Q13_16"),
    levels = c("None", "Problem Use/High Risk Substance Use"),
    classify = function(use, score, missing) case_when(
      missing ~ NA_character_,
      score >= 1 ~ "Problem Use/High Risk Substance Use",
      !use ~ "None",
      TRUE ~ "None"
```

```r
      )
    ),
    Sedative = list(
      use = "Q13_17",
      cols = c("Q13_17","Q13_18","Q13_19"),
      levels = c("None", "Problem use/High Risk Substance Use", "Missing"),
      classify = function(use, score, missing) case_when(
        missing ~ "Missing",
        score >= 1 ~ "Problem use/High Risk Substance Use",
        !use ~ "None",
        TRUE ~ "None"
      )
    ),
    `Prescribed stimulant` = list(
      use = "Q13_20",
      cols = c("Q13_21","Q13_22"),
      levels = c("None", "Problem use/High Risk Substance Use", "Missing"),
      classify = function(use, score, missing) case_when(
        missing ~ "Missing",
        score >= 1 ~ "Problem use/High Risk Substance Use",
        !use ~ "None",
        TRUE ~ "None"
      )
    )
  )
)

# Building tbl_a
tbl_a <- bind_rows(lapply(names(spec), function(name) {

  s <- spec[[name]]
  use_col <- s$use
  cols <- s$cols

  use_vec <- df254[[use_col]] == 1
  missing_vec <- is.na(df254[[use_col]])
  score_vec <- ones_count(df254, cols)

  category <- s$classify(use_vec, score_vec, missing_vec)

  make_substance_block(
    data = df254 |> transmute(ordered_grp, category = category),
    name = name,
    levels_vec = s$levels
  )
}))

# Gt table
substance_gt <- tbl_a |>
  mutate(Category = paste0("   ", Category)) |>
  gt(groupname_col = "Substance") |>
  tab_header(
    title = "Appendix Table A.",
    subtitle = md("Substance use by HIV self-test kit ordering status")
  ) |>
  cols_label(
    Category = "",
    `Ordered test kit n/N` = md("Ordered<br>n/N"),
    `Ordered test kit %`   = md("Ordered<br>%"),
    `Did not order test kit n/N` = md("Not ordered<br>n/N"),
    `Did not order test kit %`   = md("Not ordered<br>%"),
    `P-value` = "P-value"
  ) |>
  cols_align("left", Category) |>
  cols_align("right", c(
    `Ordered test kit n/N`, `Ordered test kit %`,
    `Did not order test kit n/N`, `Did not order test kit %`
  )) |>
  cols_align("center", `P-value`) |>
  tab_style(cell_text(weight = "bold"), cells_row_groups()) |>
  tab_style(cell_text(weight = "bold"), cells_body(rows = `P-value` != "")) |>
  opt_table_outline() |>
  tab_options(
    table.font.size = px(12),
    data_row.padding = px(4),
    row_group.padding = px(6)
  ) |>
  tab_style(
    cell_borders(sides = "bottom", color = "transparent"),
    cells_body(rows = TRUE)
  )

substance_gt
```

Appendix Table A.					
Substance use by HIV self-test kit ordering status					
	Ordered n/N	Ordered %	Not ordered n/N	Not ordered %	P-value
Alcohol					
High Risk Substance Use	**NA**	**NA**	**40/77**	**51.9**	**.350**
None	NA	NA	16/77	20.8	
Problem use	NA	NA	21/77	27.3	
High Risk Substance Use	79/177	44.6	NA	NA	
None	52/177	29.4	NA	NA	
Problem use	46/177	26.0	NA	NA	
Cannabis					
High Risk Substance Use	**NA**	**NA**	**33/77**	**42.9**	**.096**
None	NA	NA	32/77	41.6	
Problem use	NA	NA	12/77	15.6	
High Risk Substance Use	51/177	28.8	NA	NA	
None	89/177	50.3	NA	NA	
Problem use	37/177	20.9	NA	NA	
Stimulants					
None	**NA**	**NA**	**69/77**	**89.6**	**.299**
Problem Use/High Risk Substance Use	NA	NA	8/77	10.4	
None	166/177	93.8	NA	NA	
Problem Use/High Risk Substance Use	11/177	6.2	NA	NA	
Opioid					

Appendix Table A.

Substance use by HIV self-test kit ordering status

	Ordered n/N	Ordered %	Not ordered n/N	Not ordered %	P-value
None	**NA**	**NA**	**75/77**	**97.4**	**1.000**
Problem Use/High Risk Substance Use	NA	NA	2/77	2.6	
None	173/177	97.7	NA	NA	
Problem Use/High Risk Substance Use	4/177	2.3	NA	NA	
Sedative					
Missing	**NA**	**NA**	**0/77**	**0.0**	**.766**
None	NA	NA	72/77	93.5	
Problem use/High Risk Substance Use	NA	NA	5/77	6.5	
Missing	0/177	0.0	NA	NA	
None	168/177	94.9	NA	NA	
Problem use/High Risk Substance Use	9/177	5.1	NA	NA	
Prescribed stimulant					
Missing	**NA**	**NA**	**0/77**	**0.0**	**.641**
None	NA	NA	75/77	97.4	
Problem use/High Risk Substance Use	NA	NA	2/77	2.6	
Missing	0/177	0.0	NA	NA	
None	174/177	98.3	NA	NA	
Problem use/High Risk Substance Use	3/177	1.7	NA	NA	

Table (b)

```r
stage_levels <- tibble::tibble(
  stage = c("Precontemplation", "Contemplation", "Determination", "Action", "Maintenance"),
  description = c(
    "I do not see any need to regularly test for HIV",
    "I think I should get tested for HIV regularly, but I am not sure",
    "I am ready to start getting regularly tested for HIV",
    "I am trying to get tested regularly for HIV",
    "I have been getting tested regularly over the past few years"
  )
)

# Make analysis df with ordered_grp and stage
df254 <- df_analysis |>
  mutate(
    ordered_grp = case_when(
      str_trim(ORA_REDEEMED) == "Yes" ~ "Ordered test kit",
      str_trim(ORA_REDEEMED) %in% c("No", "Over 60 days") ~ "Did not order test kit",
      TRUE ~ NA_character_
    ),
    Q15_1 = na_if(Q15_1, 9999),
    stage = dplyr::recode(
      as.character(Q15_1),
      `1` = "Precontemplation",
      `2` = "Contemplation",
      `3` = "Determination",
      `4` = "Action",
      `5` = "Maintenance",
      .default = NA_character_
    )
  ) |>
  filter(!is.na(ordered_grp))

# N by group (fixed denominators)
N_by_group <- df254 |>
  count(ordered_grp, name = "N")

# Counts + formatting
counts <- df254 |>
  count(ordered_grp, stage, name = "n") |>
  complete(ordered_grp, stage = stage_levels$stage, fill = list(n = 0)) |>
  left_join(N_by_group, by = "ordered_grp") |>
  mutate(
    nN  = sprintf("%d/%d", n, N),
    pct = sprintf("%.1f", 100 * n / N)
  )

# Fisher p-value
tab_mat <- counts |>
  select(ordered_grp, stage, n) |>
  pivot_wider(names_from = stage, values_from = n, values_fill = 0) |>
  arrange(match(ordered_grp, c("Ordered test kit", "Did not order test kit"))) |>
  select(-ordered_grp) |>
  as.matrix()

pval_fmt <- sub("^0", "", sprintf("%.3f", fisher.test(tab_mat)$p.value))

# Final table
stage_tbl <- counts |>
  left_join(stage_levels, by = "stage") |>
  select(stage, description, ordered_grp, nN, pct) |>
  pivot_wider(
    names_from = ordered_grp,
    values_from = c(nN, pct),
    names_glue = "{.value}_{ordered_grp}"
  ) |>
  arrange(match(stage, stage_levels$stage)) |>
  mutate(`P-value` = ifelse(row_number() == 1, pval_fmt, "")) |>
  transmute(
    Stage = stage,
    Description = description,
    `Ordered test kit n/N` = `nN_Ordered test kit`,
    `Ordered test kit %`   = `pct_Ordered test kit`,
    `Did not order test kit n/N` = `nN_Did not order test kit`,
    `Did not order test kit %`   = `pct_Did not order test kit`,
    `P-value`
  )

stage_gt <- stage_tbl |>
  gt() |>
  tab_header(
    title = "Stage of change for HIV testing, by test-kit ordering",
    subtitle = md("Values are n/N and % within ordering group; Fisher's exact test p-value shown once.")
  )

stage_gt
```

Stage of change for HIV testing, by test-kit ordering

Values are n/N and % within ordering group; Fisher's exact test p-value shown once.

Stage	Description	Ordered test kit n/N	Ordered test kit %	Did not order test kit n/N	Did not order test kit %	P-value
Precontemplation	I do not see any need to regularly test for HIV	7/177	4.0	2/77	2.6	.251
Contemplation	I think I should get tested for HIV regularly, but I am not sure	31/177	17.5	8/77	10.4	
Determination	I am ready to start getting regularly tested for HIV	87/177	49.2	49/77	63.6	
Action	I am trying to get tested regularly for HIV	40/177	22.6	12/77	15.6	
Maintenance	I have been getting tested regularly over the past few years	12/177	6.8	6/77	7.8	

Table (c)

```r
# Define group
df254 <- df_analysis |>
  mutate(
    ORA_REDEEMED = str_trim(ORA_REDEEMED),
    grp = case_when(
      ORA_REDEEMED == "Yes" ~ "Ordered",
      ORA_REDEEMED %in% c("No", "Over 60 days") ~ "Not ordered",
      TRUE ~ NA_character_
    ),
    grp = factor(grp, levels = c("Ordered", "Not ordered"))
  ) |>
  filter(!is.na(grp))

# Build 2-row block
make_block <- function(var, label) {

  tab <- df254 |>
    transmute(
      grp,
      resp = recode(as.character(.data[[var]]),
        `1` = "Agree",
        `2` = "Disagree",
        .default = NA_character_
      )
    ) |>
    filter(!is.na(resp)) |>
    count(grp, resp, name = "n") |>
    complete(grp, resp = c("Agree", "Disagree"), fill = list(n = 0)) |>
    group_by(grp) |>
    mutate(
      N   = sum(n),
      nN  = sprintf("%d/%d", n, N),
      pct = sprintf("%.1f", 100 * n / N)
    ) |>
    ungroup()

  # Fisher p-value
  mat <- tab |>
    select(grp, resp, n) |>
    pivot_wider(names_from = resp, values_from = n, values_fill = 0) |>
    arrange(grp) |>
    select(Agree, Disagree) |>
    as.matrix()

  pval <- sub("^0", "", sprintf("%.3f", fisher.test(mat)$p.value))

  # Wide display table
  out <- tab |>
    select(resp, grp, nN, pct) |>
    pivot_wider(
      names_from  = grp,
      values_from = c(nN, pct),
      names_glue  = "{.value}_{grp}"
    ) |>
    arrange(match(resp, c("Agree", "Disagree"))) |>
    mutate(
      Question = label,
      `P-value` = if_else(resp == "Agree", pval, "")
    ) |>
    transmute(
      Question,
      resp,
      `Ordered n/N`     = nN_Ordered,
      `Ordered %`       = pct_Ordered,
      `Not ordered n/N` = `nN_Not ordered`,
      `Not ordered %`   = `pct_Not ordered`,
      `P-value`
    )

  out
}

# Building the full Appendix C table
qs <- tibble::tibble(
  var = c("Q15_3", "Q15_4", "Q15_5", "Q15_6", "Q15_7"),
  label = c(
    "Getting tested for HIV helps people feel better",
    "Getting tested for HIV helps people from getting HIV",
    "People in my life would leave if I had HIV",
    "People who tested positive for HIV should hide it from others",
    "I would rather not know if I have HIV"
  )
)

attitudes_tbl <- purrr::map2_dfr(qs$var, qs$label, make_block)

# GT Table
attitudes_pretty <- attitudes_tbl |>
  mutate(
    Question_disp = if_else(resp == "Agree", Question, ""),
    Response      = paste0("  ", resp),
    `Ordered %`       = paste0(`Ordered %`, "%"),
    `Not ordered %`   = paste0(`Not ordered %`, "%"),
    block_break = resp == "Agree" & row_number() > 1
  ) |>
  select(
    Question_disp, Response,
    `Ordered n/N`, `Ordered %`,
    `Not ordered n/N`, `Not ordered %`,
    `P-value`,
    block_break
  )

attitudes_gt <- attitudes_pretty |>
  gt() |>
  cols_label(
    Question_disp = "Question",
    Response = "",
    `Ordered n/N` = md("Ordered<br>n/N"),
    `Ordered %`   = md("Ordered<br>%"),
    `Not ordered n/N` = md("Not ordered<br>n/N"),
    `Not ordered %`   = md("Not ordered<br>%"),
    `P-value` = "P-value"
  ) |>
```

```
tab_header(
  title = "Appendix C.",
  subtitle = md("Attitudes toward HIV testing by HIV self-test kit ordering status (Fisher's exact test).")
) |>
cols_align(align = "left", columns = c(Question_disp, Response)) |>
cols_align(
  align = "center",
  columns = c(`Ordered n/N`, `Ordered %`,
              `Not ordered n/N`, `Not ordered %`,
              `P-value`)
) |>
tab_style(
  style = cell_text(weight = "bold"),
  locations = cells_body(rows = Response == "   Agree", columns = Question_disp)
) |>
tab_style(
  style = cell_borders(sides = "top", color = "grey75", weight = px(2)),
  locations = cells_body(rows = block_break)
) |>
cols_hide(columns = block_break) |>
opt_table_outline() |>
tab_options(
  table.font.size = px(12),
  heading.title.font.size = px(14),
  heading.subtitle.font.size = px(11),
  data_row.padding = px(4)
)

attitudes_gt
```

Appendix C.					
Attitudes toward HIV testing by HIV self-test kit ordering status (Fisher's exact test).					
Question		Ordered n/N	Ordered %	Not ordered n/N	Not ordered %	P-value
---	---	---	---	---	---	---
Getting tested for HIV helps people feel better	Agree	169/177	95.5%	70/77	90.9%	.160
	Disagree	8/177	4.5%	7/77	9.1%	
Getting tested for HIV helps people from getting HIV	Agree	148/177	83.6%	62/76	81.6%	.717
	Disagree	29/177	16.4%	14/76	18.4%	
People in my life would leave if I had HIV	Agree	59/175	33.7%	37/77	48.1%	.035
	Disagree	116/175	66.3%	40/77	51.9%	
People who tested positive for HIV should hide it from others	Agree	18/177	10.2%	9/77	11.7%	.825
	Disagree	159/177	89.8%	68/77	88.3%	
I would rather not know if I have HIV	Agree	27/177	15.3%	15/77	19.5%	.463
	Disagree	150/177	84.7%	62/77	80.5%	

Table (d)

```r
df254 <- df_analysis |>
  mutate(
    ORA_REDEEMED = str_trim(ORA_REDEEMED),
    ordered_grp = case_when(
      ORA_REDEEMED == "Yes" ~ "Ordered test kit",
      ORA_REDEEMED %in% c("No", "Over 60 days") ~ "Did not order test kit",
      TRUE ~ NA_character_
    ),
    ordered_grp = factor(ordered_grp, levels = c("Ordered test kit", "Did not order test kit"))
  ) |>
  filter(!is.na(ordered_grp))

items <- tibble(
  var = c("Q94_1","Q94_5","Q94_6","Q94_7","Q94_8","Q94_9","Q94_10","Q94_11","Q94_12","Q94_13"),
  label = c(
    "I am less threatened by the idea of being HIV positive than I used to be",
    "I am less worried about HIV infection than I used to be",
    "I think HIV/AIDS is less of a problem than it used to be",
    "I think HIV/AIDS is a less serious threat than it used to be because of new HIV/AIDS treatments",
    "I am much less concerned about becoming HIV positive myself because of new HIV/AIDS treatments",
    "I think that condom use during sex is less necessary now that new HIV/AIDS treatments are available",
    "I think that someone who is HIV positive now needs to care less about condom use",
    "I think that the need for condom use is less than it used to be, because you can always start new treatment
s",
    "I think that someone who is HIV positive and uses new HIV/AIDS treatments can be cured",
    "I think that new HIV/AIDS treatments can eradicate the virus from your body"
  )
)

hiv_treatment_tbl <- bind_rows(lapply(seq_len(nrow(items)), function(i) {

  var   <- items$var[i]
  label <- items$label[i]

  d <- df254 |>
    transmute(
      ordered_grp,
      value = as.numeric(.data[[var]])
    ) |>
    filter(!is.na(ordered_grp))

  x_ordered <- d |>
    filter(ordered_grp == "Ordered test kit") |>
    pull(value)

  x_not <- d |>
    filter(ordered_grp == "Did not order test kit") |>
    pull(value)

  # Mean (SD) strings, ignoring NAs
  x_ordered2 <- x_ordered[!is.na(x_ordered)]
  x_not2     <- x_not[!is.na(x_not)]

  ordered_str <- if (length(x_ordered2) == 0) NA_character_ else sprintf("%.1f (%.1f)", mean(x_ordered2), sd(x_or
dered2))
  not_str     <- if (length(x_not2) == 0)     NA_character_ else sprintf("%.1f (%.1f)", mean(x_not2), sd(x_not2))

  # Wilcoxon test on complete cases
  dd_complete <- d |> filter(!is.na(value))

  pval <- NA_real_
  if (n_distinct(dd_complete$ordered_grp) == 2 && nrow(dd_complete) > 0) {
    pval <- wilcox.test(value ~ ordered_grp, data = dd_complete, exact = FALSE)$p.value
  }

  pval_str <- if (is.na(pval)) NA_character_ else sub("^0", "", sprintf("%.3f", pval))

  tibble(
    Statements = label,
    `Ordered test kit Mean (SD)` = ordered_str,
    `Did not order test kit Mean (SD)` = not_str,
    `P-value (Wilcoxon rank test)` = pval_str
  )
}))

# -----------------------------------
# Everything below is for making the table look pretty

hiv_gt <- hiv_treatment_tbl |>
  gt() |>
  tab_header(
    title = "Appendix: HIV treatment perceptions by test-kit ordering status",
    subtitle = md("Values are mean (SD). P-values from Wilcoxon rank-sum tests.")
  ) |>
  cols_label(
    Statements = "Statements",
    `Ordered test kit Mean (SD)` = md("Ordered test kit<br>Mean (SD)"),
    `Did not order test kit Mean (SD)` = md("Did not order test kit<br>Mean (SD)"),
    `P-value (Wilcoxon rank test)` = md("P-value<br>(Wilcoxon)")
  ) |>
  # alignment
  cols_align(align = "left", columns = Statements) |>
  cols_align(
    align = "center",
    columns = c(`Ordered test kit Mean (SD)`,
                `Did not order test kit Mean (SD)`,
                `P-value (Wilcoxon rank test)`)
  ) |>
  # column widths
  cols_width(
    Statements ~ pct(55),
    `Ordered test kit Mean (SD)` ~ pct(16),
    `Did not order test kit Mean (SD)` ~ pct(18),
    `P-value (Wilcoxon rank test)` ~ pct(11)
  ) |>
  # making it look like a paper table
  opt_table_outline() |>
  opt_row_striping() |>
  tab_style(
    style = cell_borders(sides = "bottom", color = "grey90", weight = px(1)),
    locations = cells_body(rows = TRUE)
  ) |>
  tab_options(
    table.font.size = px(12),
```

```
    heading.title.font.size = px(14),
    heading.subtitle.font.size = px(11),
    data_row.padding = px(5)
  )

hiv_gt
```

Appendix: HIV treatment perceptions by test-kit ordering status			
Values are mean (SD). P-values from Wilcoxon rank-sum tests.			
Statements	Ordered test kit Mean (SD)	Did not order test kit Mean (SD)	P-value (Wilcoxon)
I am less threatened by the idea of being HIV positive than I used to be	4.1 (2.0)	3.8 (2.2)	.421
I am less worried about HIV infection than I used to be	3.7 (2.0)	3.5 (1.8)	.400
I think HIV/AIDS is less of a problem than it used to be	3.9 (2.1)	3.6 (2.1)	.413
I think HIV/AIDS is a less serious threat than it used to be because of new HIV/AIDS treatments	4.2 (2.0)	3.8 (2.1)	.225
I am much less concerned about becoming HIV positive myself because of new HIV/AIDS treatments	3.4 (2.0)	3.0 (1.9)	.274
I think that condom use during sex is less necessary now that new HIV/AIDS treatments are available	2.4 (1.8)	2.2 (1.5)	.971
I think that someone who is HIV positive now needs to care less about condom use	1.9 (1.7)	2.5 (2.1)	.064
I think that the need for condom use is less than it used to be, because you can always start new treatments	2.5 (1.9)	2.5 (2.0)	.767
I think that someone who is HIV positive and uses new HIV/AIDS treatments can be cured	3.2 (1.9)	3.6 (2.2)	.199
I think that new HIV/AIDS treatments can eradicate the virus from your body	3.1 (1.9)	3.8 (2.1)	.029

Table (e)

```r
likert_levels <- tibble::tibble(
  value = 1:7,
  Response = c(
    "Strongly agree","Agree","Somewhat agree","Neither agree nor disagree",
    "Somewhat disagree","Disagree","Strongly disagree"
  )
)

items <- tibble::tibble(
  var = c("Q14_2", "Q14_3", "Q14_4", "Q14_5"),
  label = c(
    "I feel afraid of people living with HIV/AIDS",
    "I could not be friends with someone who has HIV/AIDS",
    "People who get HIV/AIDS through sex or drug use got what they deserve",
    "I feel anger toward people with HIV/AIDS"
  )
)

# Long once (all items)
long <- df254 |>
  select(ordered_grp, all_of(items$var)) |>
  pivot_longer(
    cols = all_of(items$var),
    names_to = "var",
    values_to = "value_raw"
  ) |>
  mutate(
    ordered_grp = factor(ordered_grp, levels = c("Ordered test kit", "Did not order test kit")),
    value = na_if(as.character(value_raw), "9999"),
    value = suppressWarnings(as.integer(value))
  ) |>
  left_join(items, by = "var") |>
  filter(!is.na(ordered_grp), !is.na(value))

# Counts and % for all items
counts <- long |>
  count(var, label, ordered_grp, value, name = "n") |>
  complete(
    var,
    ordered_grp = levels(long$ordered_grp),
    value = 1:7,
    fill = list(n = 0)
  ) |>
  group_by(var, ordered_grp) |>
  mutate(N = sum(n)) |>
  ungroup() |>
  left_join(likert_levels, by = "value") |>
  mutate(
    nN  = sprintf("%d/%d", n, N),
    pct = sprintf("%.1f", 100 * n / N)
  )

# Wilcoxon p-values once per item
pvals <- long |>
  group_by(var) |>
  summarise(p = wilcox.test(value ~ ordered_grp, exact = FALSE)$p.value, .groups = "drop") |>
  mutate(p_fmt = paste0(sub("^0", "", sprintf("%.3f", p)), "‡")) |>
  select(var, p_fmt)

# Final 7-row blocks per statement
stigma_tbl <- counts |>
  select(var, label, Response, ordered_grp, nN, pct) |>
  pivot_wider(
    names_from  = ordered_grp,
    values_from = c(nN, pct),
    names_glue  = "{.value}_{ordered_grp}"
  ) |>
  arrange(match(var, items$var), match(Response, likert_levels$Response)) |>
  left_join(pvals, by = "var") |>
  group_by(var) |>
  mutate(`P-value` = if_else(row_number() == 1, p_fmt, "")) |>
  ungroup() |>
  transmute(
    Statements = label,
    Response,
    `Ordered test kit n/N` = `nN_Ordered test kit`,
    `Ordered test kit %`   = `pct_Ordered test kit`,
    `Did not order test kit n/N` = `nN_Did not order test kit`,
    `Did not order test kit %`   = `pct_Did not order test kit`,
    `P-value`
  )

# GT Table
stigma_pretty <- stigma_tbl |>
  mutate(
    is_first = Response == "Strongly agree",
    block_break = is_first & row_number() > 1,
    Statements_disp = if_else(is_first, Statements, ""),
    Response_disp = paste0("   ", Response),
    `Ordered test kit %` = paste0(`Ordered test kit %`, "%"),
    `Did not order test kit %` = paste0(`Did not order test kit %`, "%")
  ) |>
  select(
    Statements_disp, Response_disp,
    `Ordered test kit n/N`, `Ordered test kit %`,
    `Did not order test kit n/N`, `Did not order test kit %`,
    `P-value`,
    is_first, block_break
  )

stigma_gt <- stigma_pretty |>
  gt() |>
  tab_header(
    title = "HIV stigma attitudes by HIV self-test kit ordering status",
    subtitle = md("Likert responses shown as n/N and % within group. Wilcoxon rank-sum test p-value shown once per statement (‡).")
  ) |>
  cols_label(
    Statements_disp = "Statements",
    Response_disp = "Response",
    `Ordered test kit n/N` = "n/N",
    `Ordered test kit %`   = "%",
    `Did not order test kit n/N` = "n/N",
    `Did not order test kit %`   = "%",
```

```
    `P-value` = "P-value"
  ) |>
  tab_spanner(label = "Ordered test kit", columns = c(`Ordered test kit n/N`, `Ordered test kit %`)) |>
  tab_spanner(label = "Did not order test kit", columns = c(`Did not order test kit n/N`, `Did not order test kit
%`)) |>
  cols_align(align = "left", columns = c(Statements_disp, Response_disp)) |>
  cols_align(
    align = "center",
    columns = c(`Ordered test kit n/N`, `Ordered test kit %`,
                `Did not order test kit n/N`, `Did not order test kit %`,
                `P-value`)
  ) |>
  tab_style(
    style = cell_text(weight = "bold"),
    locations = cells_body(rows = is_first, columns = Statements_disp)
  ) |>
  tab_style(
    style = cell_borders(sides = "top", color = "grey70", weight = px(2)),
    locations = cells_body(rows = block_break)
  ) |>
  cols_hide(columns = c(is_first, block_break)) |>
  opt_table_outline()

stigma_gt
```

HIV stigma attitudes by HIV self-test kit ordering status						
Likert responses shown as n/N and % within group. Wilcoxon rank-sum test p-value shown once per statement (‡).						
		Ordered test kit		Did not order test kit		P-value
Statements	Response	n/N	%	n/N	%	
I feel afraid of people living with HIV/AIDS	Strongly agree	12/177	6.8%	7/77	9.1%	.613‡
	Agree	9/177	5.1%	2/77	2.6%	
	Somewhat agree	21/177	11.9%	9/77	11.7%	
	Neither agree nor disagree	20/177	11.3%	9/77	11.7%	
	Somewhat disagree	11/177	6.2%	6/77	7.8%	
	Disagree	32/177	18.1%	17/77	22.1%	
	Strongly disagree	72/177	40.7%	27/77	35.1%	
I could not be friends with someone who has HIV/AIDS	Strongly agree	1/177	0.6%	4/77	5.2%	.032‡
	Agree	NA	NA%	0/77	0.0%	
	Agree	1/177	0.6%	NA	NA%	
	Somewhat agree	3/177	1.7%	1/77	1.3%	
	Neither agree nor disagree	9/177	5.1%	6/77	7.8%	
	Somewhat disagree	7/177	4.0%	1/77	1.3%	
	Disagree	26/177	14.7%	19/77	24.7%	
	Strongly disagree	130/177	73.4%	46/77	59.7%	
NA	Strongly agree	NA	NA%	0/77	0.0%	.332‡
People who get HIV/AIDS through sex or drug use got what they deserve	Strongly agree	2/177	1.1%	NA	NA%	
	Agree	2/177	1.1%	1/77	1.3%	
	Somewhat agree	3/177	1.7%	2/77	2.6%	
	Neither agree nor disagree	8/177	4.5%	7/77	9.1%	
	Somewhat disagree	6/177	3.4%	2/77	2.6%	
	Disagree	24/177	13.6%	12/77	15.6%	
	Strongly disagree	132/177	74.6%	53/77	68.8%	
NA	Strongly agree	NA	NA%	0/77	0.0%	.213‡
I feel anger toward people with HIV/AIDS	Strongly agree	1/177	0.6%	NA	NA%	
	Agree	0/177	0.0%	0/77	0.0%	
	Somewhat agree	0/177	0.0%	0/77	0.0%	
	Neither agree nor disagree	11/177	6.2%	5/77	6.5%	
	Somewhat disagree	NA	NA%	0/77	0.0%	
	Somewhat disagree	3/177	1.7%	NA	NA%	
	Disagree	19/177	10.7%	16/77	20.8%	
	Strongly disagree	143/177	80.8%	56/77	72.7%	

Table (f)

```r
mistrust_levels <- tibble::tibble(
  resp = 1:4,
  Response = c("Strongly agree", "Agree", "Disagree", "Strongly disagree")
)

items_mistrust <- tibble::tibble(
  var = paste0("Q16_", 1:7),
  label = c(
    "You'd better be cautious when dealing with health care organizations",
    "Patients have sometimes been deceived or misled by health care organizations",
    "When health care organizations make mistakes they usually cover it up",
    "Health care organizations have sometimes done harmful experiments on patients without their knowledge",
    "Health care organizations don't always keep your information totally private",
    "Sometimes I wonder if health care organizations really know what they are doing",
    "Mistakes are common in health care organizations"
  )
)

# Long once and recoding to 1..4
long <- df254 |>
  select(ordered_grp, all_of(items_mistrust$var)) |>
  pivot_longer(
    cols = all_of(items_mistrust$var),
    names_to = "var",
    values_to = "raw"
  ) |>
  mutate(
    ordered_grp = factor(ordered_grp, levels = c("Ordered test kit", "Did not order test kit")),
    raw = na_if(as.character(raw), "9999"),
    raw = suppressWarnings(as.numeric(raw)),
    resp = case_when(
      # Q16_1 special coding: 28,30,33,34 -> 1..4
      var == "Q16_1" & raw == 28 ~ 1L,
      var == "Q16_1" & raw == 30 ~ 2L,
      var == "Q16_1" & raw == 33 ~ 3L,
      var == "Q16_1" & raw == 34 ~ 4L,

      # Q16_2..Q16_7: 1,2,6,7 -> 1..4
      var != "Q16_1" & raw == 1 ~ 1L,
      var != "Q16_1" & raw == 2 ~ 2L,
      var != "Q16_1" & raw == 6 ~ 3L,
      var != "Q16_1" & raw == 7 ~ 4L,

      TRUE ~ NA_integer_
    )
  ) |>
  left_join(items_mistrust, by = "var") |>
  filter(!is.na(ordered_grp), !is.na(resp))

# Counts and percents (all items at once)
counts <- long |>
  count(var, label, ordered_grp, resp, name = "n") |>
  complete(
    var,
    ordered_grp = levels(long$ordered_grp),
    resp = 1:4,
    fill = list(n = 0)
  ) |>
  group_by(var, ordered_grp) |>
  mutate(N = sum(n)) |>
  ungroup() |>
  left_join(mistrust_levels, by = "resp") |>
  mutate(
    nN  = sprintf("%d/%d", n, N),
    pct = sprintf("%.1f", 100 * n / N)
  )

# Wilcoxon p-values
pvals <- long |>
  group_by(var) |>
  summarise(p = wilcox.test(resp ~ ordered_grp, exact = FALSE)$p.value, .groups = "drop") |>
  mutate(p_fmt = paste0(sub("^0", "", sprintf("%.3f", p)), "‡")) |>
  select(var, p_fmt)

# Final 4-row blocks per statement
mistrust_tbl <- counts |>
  select(var, label, Response, ordered_grp, nN, pct) |>
  pivot_wider(
    names_from  = ordered_grp,
    values_from = c(nN, pct),
    names_glue  = "{.value}_{ordered_grp}"
  ) |>
  arrange(match(var, items_mistrust$var), match(Response, mistrust_levels$Response)) |>
  left_join(pvals, by = "var") |>
  group_by(var) |>
  mutate(`P-value` = if_else(row_number() == 1, p_fmt, "")) |>
  ungroup() |>
  transmute(
    Statements = label,
    Response,
    `Ordered test kit n/N` = `nN_Ordered test kit`,
    `Ordered test kit %`   = `pct_Ordered test kit`,
    `Did not order test kit n/N` = `nN_Did not order test kit`,
    `Did not order test kit %`   = `pct_Did not order test kit`,
    `P-value`
  )

# GT Table
mistrust_pretty <- mistrust_tbl |>
  mutate(
    is_first = Response == "Strongly agree",
    block_break = is_first & row_number() > 1,
    Statements_disp = if_else(is_first, Statements, ""),
    Response_disp = paste0("   ", Response),
    `Ordered test kit %` = paste0(`Ordered test kit %`, "%"),
    `Did not order test kit %` = paste0(`Did not order test kit %`, "%")
  ) |>
  select(
    Statements_disp, Response_disp,
    `Ordered test kit n/N`, `Ordered test kit %`,
    `Did not order test kit n/N`, `Did not order test kit %`,
    `P-value`,
    is_first, block_break
  )
```

```r
mistrust_gt <- mistrust_pretty |>
  gt() |>
  tab_header(
    title = "Medical mistrust items by HIV self-test kit ordering status",
    subtitle = md("Responses shown as n/N and % within group. Wilcoxon rank-sum test p-value shown once per state
ment (‡).")
  ) |>
  cols_label(
    Statements_disp = "Statements",
    Response_disp = "Response",
    `Ordered test kit n/N` = "n/N",
    `Ordered test kit %`   = "%",
    `Did not order test kit n/N` = "n/N",
    `Did not order test kit %`   = "%",
    `P-value` = "P-value"
  ) |>
  tab_spanner(label = "Ordered test kit", columns = c(`Ordered test kit n/N`, `Ordered test kit %`)) |>
  tab_spanner(label = "Did not order test kit", columns = c(`Did not order test kit n/N`, `Did not order test kit
%`)) |>
  tab_style(
    style = cell_text(weight = "bold"),
    locations = cells_body(rows = is_first, columns = Statements_disp)
  ) |>
  tab_style(
    style = cell_borders(sides = "top", color = "grey70", weight = px(2)),
    locations = cells_body(rows = block_break)
  ) |>
  cols_hide(columns = c(is_first, block_break)) |>
  opt_table_outline() |>
  tab_options(
    table.font.size = px(12),
    heading.title.font.size = px(14),
    heading.subtitle.font.size = px(11),
    data_row.padding = px(4)
  )

mistrust_gt
```

Medical mistrust items by HIV self-test kit ordering status

Responses shown as n/N and % within group. Wilcoxon rank-sum test p-value shown once per statement (‡).

Statements	Response	Ordered test kit		Did not order test kit		P-value
		n/N	%	n/N	%	
You'd better be cautious when dealing with health care organizations	Strongly agree	37/177	20.9%	17/76	22.4%	.503‡
	Agree	66/177	37.3%	32/76	42.1%	
	Disagree	42/177	23.7%	14/76	18.4%	
	Strongly disagree	32/177	18.1%	13/76	17.1%	
Patients have sometimes been deceived or misled by health care organizations	Strongly agree	32/176	18.2%	11/77	14.3%	.413‡
	Agree	83/176	47.2%	34/77	44.2%	
	Disagree	33/176	18.8%	22/77	28.6%	
	Strongly disagree	28/176	15.9%	10/77	13.0%	
When health care organizations make mistakes they usually cover it up	Strongly agree	29/174	16.7%	10/76	13.2%	.222‡
	Agree	83/174	47.7%	29/76	38.2%	
	Disagree	36/174	20.7%	30/76	39.5%	
	Strongly disagree	26/174	14.9%	7/76	9.2%	
Health care organizations have sometimes done harmful experiments on patients without their knowledge	Strongly agree	30/176	17.0%	14/77	18.2%	.413‡
	Agree	73/176	41.5%	25/77	32.5%	
	Disagree	51/176	29.0%	26/77	33.8%	
	Strongly disagree	22/176	12.5%	12/77	15.6%	
Health care organizations don't always keep your information totally private	Strongly agree	36/175	20.6%	11/76	14.5%	.371‡
	Agree	64/175	36.6%	29/76	38.2%	
	Disagree	47/175	26.9%	23/76	30.3%	
	Strongly disagree	28/175	16.0%	13/76	17.1%	
Sometimes I wonder if health care organizations really know what they are doing	Strongly agree	22/176	12.5%	9/77	11.7%	.965‡
	Agree	66/176	37.5%	28/77	36.4%	
	Disagree	60/176	34.1%	30/77	39.0%	
	Strongly disagree	28/176	15.9%	10/77	13.0%	
Mistakes are common in health care organizations	Strongly agree	21/176	11.9%	8/75	10.7%	.638‡
	Agree	89/176	50.6%	36/75	48.0%	
	Disagree	46/176	26.1%	23/75	30.7%	
	Strongly disagree	20/176	11.4%	8/75	10.7%	

Compelling figure (Bonus)

```r
# Keeping only sites with observed orders
plot_data_clean <- bind_rows(
  res_w1$rates_tbl,
  res_w2$rates_tbl
) |>
  filter(o > 0)

plot_data_clean <- plot_data_clean |>
  group_by(Wave) |>
  mutate(
    Site = reorder(Site, rate_per_day)
  ) |>
  ungroup()

ggplot(plot_data_clean,
       aes(x = Site,
           y = rate_per_day,
           color = Site)) +

  geom_point(size = 3) +

  geom_errorbar(
    aes(ymin = lower_95, ymax = upper_95),
    width = 0.12,
    linewidth = 0.9
  ) +

  facet_wrap(~ Wave, scales = "free_x") +

  scale_y_continuous(
    "Orders per day",
    limits = c(0, NA),
    expand = expansion(mult = c(0, 0.1))
  ) +

  labs(
    title = "HIV Self-Test Kit Ordering Rates by Recruitment Platform",
    subtitle = "Poisson regression estimates with 95% confidence intervals",
    x = "Recruitment platform"
  ) +

  theme_minimal(base_size = 14) +

  theme(
    legend.position = "none",
    plot.title = element_text(face = "bold"),
    axis.text.x = element_text(angle = 20, hjust = 1),
    panel.grid.minor = element_blank()
  )
```

HIV Self-Test Kit Ordering Rates by Recruitment Platform

Poisson regression estimates with 95% confidence intervals