Criminal History Records Paper Step 2 (Analysis)

Steven J. Pierce

Contents

1 Purpose

This file analyzes data derived from publicly archived data files (Campbell, 2019) to produce outputs for a descriptive paper about the criminal histories of suspected serial sexual perpetrators.[1] The data come from Detroit, MI.

2 Notes on Generalized Estimating Equation (GEE) Results

Numerous sections below contain output for GEE models and estimates derived from them, such as estimated marginal means and contrasts for ratios of those means. We performed those analyses because we think the contrasts provide useful insights into how much offenders' rates of criminal activity differed between the periods before versus after the testing window. Reviewers found some of them less relevant than we did. We have left those results in the research compendium despite having removed some of them from the final paper because readers may still find them interesting and relevant. Preserving them facilitates achieving transparency regarding all the analyses we conducted.

In particular note that we ran models that compared incident counts with and without adjustments for period duration. Only the adjusted comparisons remain in the published paper.

3 Setup

Set global R chunk options (local chunk options will over-ride global options). The method for creating a size option that controls font size in code chunks and their text output is based on an answer to a question posted on stackoverflow.com.

```
# 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)
  })

# Global chunk options (over-ridden by local chunk options)
knitr::opts_chunk$set(include  = TRUE, echo = TRUE, error = TRUE,
                      message = TRUE, warning = TRUE, cfsize = "footnotesize")

# Declare location of this script relative to the project root directory.
here::i_am(path = "inst/Step_02_Analysis.Rmd")
```

```
## here() starts at S:/14-286/Analyses/SSACHR
```

Load R packages that we need to get additional functions.

```
library(here)             # for here()
library(plyr)             # For mapvalues()
```

[1] In this document, we use the terms perpetrator and offender interchangeably. Both terms should be interpreted as referring to individuals suspected of committing a crime but we occasionally omit that qualifier for the sake of brevity.

```
##
## Attaching package: 'plyr'
```

```
## The following object is masked from 'package:here':
##
##     here
```

```
library(dplyr)            # for %>%, filter(), group_by(), & summarise()
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:plyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)            # for arrange(), filter(), group_by(), mutate(),
                          # spread(), summarise(), %>%, etc.
library(rmarkdown)        # for render()
library(knitr)            # for kable()
options(kableExtra.latex.load_packages = FALSE)
library(kableExtra)       # for kable_styling(), add_header_above(), column_spec(),
```

```
##
## Attaching package: 'kableExtra'
```

```
## The following object is masked from 'package:dplyr':
##
##     group_rows
```

```
                          # collapse_rows(), and landscape()
library(descr)            # For freq().
options(descr.plot=FALSE) # Make freq() & crosstab() skip plots by default.
library(lubridate)        # For date conversion, eg. ymd(), time_length().
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(sjlabelled)       # For set_label(), get_label()
```

```
##
## Attaching package: 'sjlabelled'
```

```
## The following object is masked from 'package:dplyr':
##
##     as_label
```

```
library(haven)            # for read_spss()
```

```
##
## Attaching package: 'haven'
```

```
## The following objects are masked from 'package:sjlabelled':
##
##     as_factor, read_sas, read_spss, read_stata, write_sas, zap_labels
```

```
library(lattice)          # For xyplot(), bwplot(), etc.
library(latticeExtra)     # for layer()
library(psych)            # For describe()
library(car)              # For recode()
```

```
## Loading required package: carData
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:psych':
##
##     logit
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

```
library(SSACHR)           # for git_report(), rvlabel(), which_latex()
library(geepack)          # for geeglm()
library(emmeans)          # for emmeans()
library(texreg)           # for texreg()
```

```
## Version:  1.37.5
## Date:     2020-06-17
## Author:   Philip Leifeld (University of Essex)
##
## Consider submitting praise using the praise or praise_interactive functions.
## Please cite the JSS article in your publications -- see citation("texreg").
```

```
##
## Attaching package: 'texreg'
```

```
## The following object is masked from 'package:tidyr':
##
##     extract
```

```
library(ggplot2)          # for ggplot()
```

```
##
## Attaching package: 'ggplot2'
```

```
## The following objects are masked from 'package:psych':
##
##     %+%, alpha
```

```
## The following object is masked from 'package:latticeExtra':
##
##     layer
```

```
## The following object is masked from 'package:sjlabelled':
##
##     as_label
```

```
library(ggdist)           # for stat_halfeye()
```

Create objects to hold settings we plan to reuse often.

```
# Custom settings for use with bwplot().
my.boxes <- list(box.rectangle = list(col = "black", lwd = 2),
                 box.umbrella = list(col = "black", lwd = 2))
```

4 Load Data

The chunk below loads the data saved by the prior script *Step_01_Data_Mgt.Rmd*.

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

5 Offender-Level Summaries

We have few offender-level demographic variables that are worth reporting (we will omit summarizing height, weight, hair color, & eye color). After examining sex, race, and age, we switch focus to looking at descriptive data about the criminal history records for arrests, prosecutor charges, and adjudicated charges.

```
# Store objects with sample sizes we will reuse often later.
N_All    <- IDNEW %>% nrow()
N_Before <- IDNEW %>% filter(Years_Before > 0) %>% nrow()
N_During <- IDNEW %>% filter(Years_During > 0) %>% nrow()
N_After  <- IDNEW %>% filter(Years_After > 0) %>% nrow()

data.frame(N_All, N_Before, N_During, N_After)
```

```
##   N_All N_Before N_During N_After
## 1  1082     1041     1052    1082
```

All the CHR summaries in this section focus on variables that have been aggregated to the offender level. That is why the sample size for all of these variables is N = 1,082 (the number of offenders for whom Michigan CHR data were available and the start date for the earliest SAK testing window could be established).

5.1 Sex

```
kable(freq(as_factor(IDNEW$Sex)), format = "latex", booktabs = TRUE, digits = 2,
      caption = "Offender Sex")
```

	Frequency	Percent
Female	0	0
Male	1082	100
Total	1082	100

Table 1: Offender Sex

5.2 Race

```
kable(freq(as_factor(IDNEW$Race), user.missing = "Unknown"), format = "latex",
      booktabs = TRUE,  digits = 2, caption = "Offender Race")
```

	Frequency	Percent	Valid Percent
Asian or Pacific Islander	2	0.18	0.18
Black	1027	94.92	94.92
American Indian or Alaskan Native	2	0.18	0.18
White or Hispanic	51	4.71	4.71
Unknown	0	0.00	NA
Total	1082	100.00	100.00

Table 2: Offender Race

5.3 Age

Table 3 presents descriptive statistics for offender age at the start of the earliest testing window and at the time of CHR collection. Figure 1 shows boxplots of the distributions for both of these age variables.

```
CNames <- c("Variable", "N", "Mean", "SD", "Min", "Max", "Skew", "Kurtosis",
            "Q25", "Q50", "Q75")
IDNEW %>%
  select(AgeWDate, Age) %>%
  describe(., quant=c(.25, .50, .75)) %>%
  mutate(Variable  = c("Age at earliest SAK", "Age at CHR collection")) %>%
  select(Variable, n, mean, sd, min, max, skew, kurtosis, Q0.25, Q0.5,
         Q0.75) %>%
  kable(., format = "latex", booktabs = TRUE, digits = 2,
      col.names = CNames, row.names = FALSE,
      caption = "Offender Age")
```

Variable	N	Mean	SD	Min	Max	Skew	Kurtosis	Q25	Q50	Q75
Age at earliest SAK	1082	27.33	8.16	13.2	57.09	0.62	-0.18	20.69	26.15	32.91
Age at CHR collection	1082	45.19	8.94	22.0	77.00	0.15	-0.12	39.00	45.00	51.75

Table 3: Offender Age

```
# Figure caption.
FCap <- paste("\\label{fig:boxplot_Age}",
              "Box Plots for Offender Age. Only offenders with valid testing",
              "window start dates are included. The red line marks age 16",
              "years (the start of observed adult CHR data). CHR collection",
              "occurred on 04/15/2016 for all offfenders.",
              "CHR, criminal history record; SAK, sexual assault kit.")

IDNEW %>%
  select(OID, AgeWDate, Age) %>%
  # Rearrange data so we can easily plot multiple variables.
  pivot_longer(cols = -OID, names_to = "VName", values_to = "Age") %>%
  mutate(VName = ordered(VName, levels = c("AgeWDate", "Age"),
                         labels = c("At earliest SAK", "At CHR collection"))) %>%
  bwplot(VName ~ Age, data = ., factor = 1, xlim = c(-2, 82),
       panel = panel.mybox, par.settings = my.boxes,
       xlab = "Offender Age")  +
  latticeExtra::layer(panel.abline(v = 16, col.line = "red"))
```

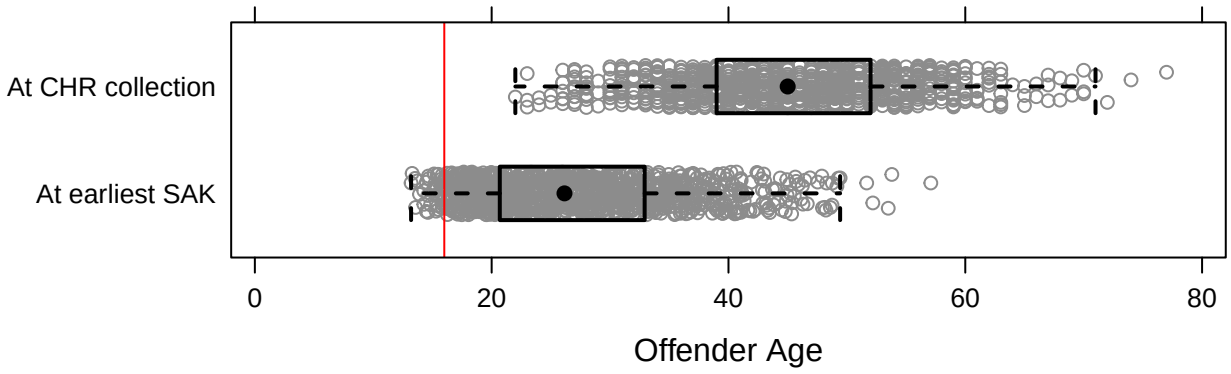

Figure 1: Box Plots for Offender Age. Only offenders with valid testing window start dates are included. The red line marks age 16 years (the start of observed adult CHR data). CHR collection occurred on 04/15/2016 for all offfenders. CHR, criminal history record; SAK, sexual assault kit.

5.4 CHR Periods

The observed portions of these offenders' adult criminal histories include incident, arrest, prosecutor charge, and judicial charge records associated with crimes that occurred between the offender's 16th birthday and April 15, 2016 (when the CHR data were collected from the official statewide CHR database). That time frame can be divided into three important periods for the purposes of this study: before, during, and after the testing window associated with the offender's earliest sexual assault kit (SAK). Here we provide descriptive data about those periods. Note that the relative timing of an offender's earliest SAK relative to their 16th birthday controls how much of the before and during periods were observed versus unobserved. Figures 2 and 3 respectively show boxplots for the date of the offenders 16th birthday and the start date of the earliest testing window (which is the date the earliest SAK was collected).

```
# Figure caption.
FCap <- paste("\\label{fig:boxplot_Age16Date}",
              "Box Plot for Date of Offender's 16th Birthday. Only offenders",
              "with valid testing window start dates are included.",
              "These dates mark the start of the observed adult criminal",
              "history record for each offender.")

bwplot(~ Age16Date, data = IDNEW, factor = 9,
       xlim = c(ymd("1948-01-01"), ymd("2012-12-31")),
       panel = panel.mybox, par.settings = my.boxes,
       xlab = "Offender's 16th Birthday")
```

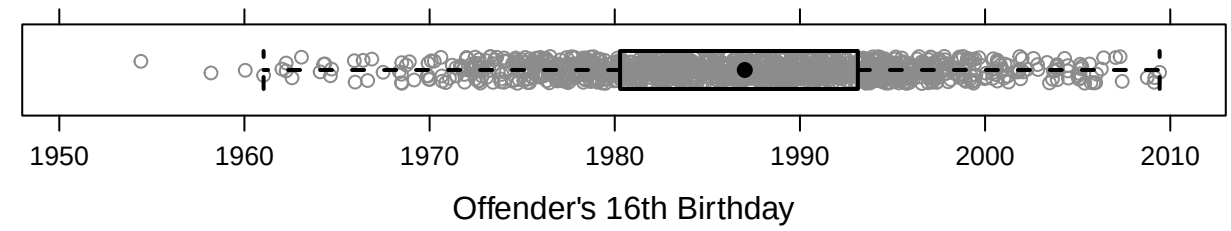

Figure 2: Box Plot for Date of Offender's 16th Birthday. Only offenders with valid testing window start dates are included. These dates mark the start of the observed adult criminal history record for each offender.

```
# Figure caption.
FCap <- paste("\\label{fig:boxplot_WDate}",
              "Box Plot for Start Date of Earliest Testing Window. Only",
              "offenders with valid testing window start dates are included.")

bwplot(~ WDate, data = IDNEW, factor = 9,
       xlim = c(ymd("1984-01-01"), ymd("2016-12-31")),
       panel = panel.mybox, par.settings = my.boxes,
       xlab = "Start of Earliest Testing Window")
```

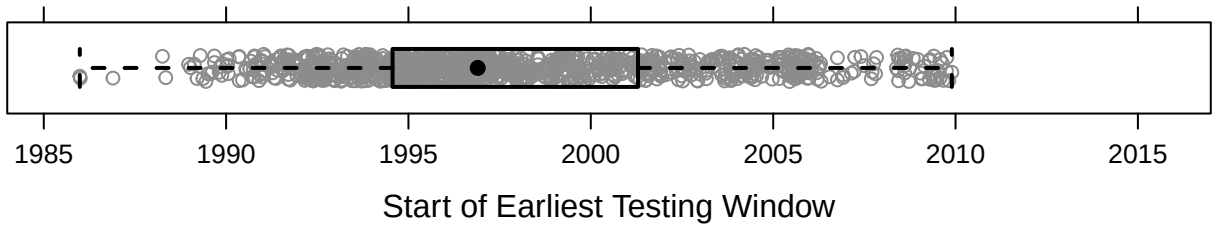

Figure 3: Box Plot for Start Date of Earliest Testing Window. Only offenders with valid testing window start dates are included.

Those dates were used in the previous data management script to compute the durations of the observed portions of adult CHR that correspond to the periods before, during, and after the earliest testing window. Table 4 shows descriptive statistics for the these durations and Figure 4 shows corresponding boxplots. These provide crucial context for interpreting the numbers of crimes observed in each of those periods and provide the denominators for translating them into incidence rates.

```
# Table caption.
TCap <- paste("Observed Period Durations (Years)")
# Footnote text.
FN <- paste0("Only offenders with valid testing window start dates are ",
             "included. ",
             "TW, earliest testing window.")

rlabs <- c("Before TW", "During TW", "After TW")

IDNEW %>%
  select(Years_Before, Years_During, Years_After) %>%
  describe(., quant=c(.25, .50, .75)) %>%
  bind_cols(data.frame(Variable = rlabs), .) %>%
  select(Variable, n, mean, sd, min, max, skew, kurtosis, Q0.25, Q0.5, Q0.75) %>%
  kable(., format = "latex", booktabs = TRUE, digits = 2, row.names = FALSE,
      col.names = CNames, caption = TCap) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

Variable	N	Mean	SD	Min	Max	Skew	Kurtosis	Q25	Q50	Q75
Before TW	1082	11.37	8.10	0.00	41.09	0.65	-0.18	4.69	10.15	16.91
During TW	1082	0.32	0.06	0.00	0.33	-5.24	25.82	0.33	0.33	0.33
After TW	1082	18.01	4.84	6.06	29.97	-0.50	-0.43	14.64	19.03	21.38

Note: Only offenders with valid testing window start dates are included. TW, earliest testing window.

Table 4: Observed Period Durations (Years)

```
# Figure caption.
FCap <- paste("\\label{fig:boxplot_windows}",
              "Box Plots for Duration of Observed Portions of Periods.",
```

```
              "Only offenders with valid testing",
              "window start dates are included. The red line marks age 16",
              "years, which is the start of observed adult CHR, which was ",
              "collected on 04/15/2016.",
              "TW, earliest testing window.")

IDNEW %>%
  select(OID, Years_Before, Years_During, Years_After) %>%
  # Rearrange data so we can easily plot multiple variables.
  pivot_longer(cols = -OID, names_to = "VName", values_to = "Duration") %>%
  mutate(VName = ordered(VName,
                         levels = rev(c("Years_Before", "Years_During",
                                        "Years_After")),
                         labels = rev(rlabs))) %>%
  bwplot(VName ~ Duration, data = ., factor = 1, xlim = c(-2, 42),
       panel = panel.mybox, par.settings = my.boxes,
       xlab = "Observed Period Duration (Years)")
```

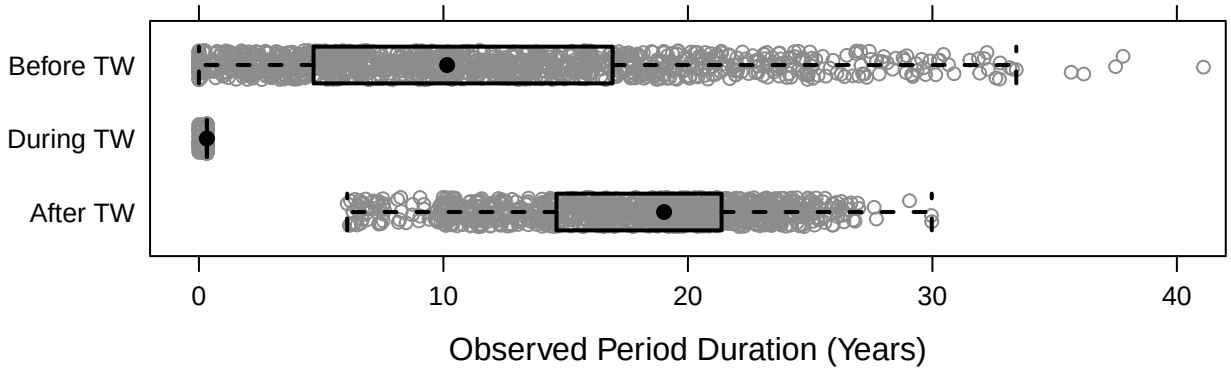

Figure 4: Box Plots for Duration of Observed Portions of Periods. Only offenders with valid testing window start dates are included. The red line marks age 16 years, which is the start of observed adult CHR, which was collected on 04/15/2016. TW, earliest testing window.

5.5 CHR Record Counts

We aggregated data to count how many CHR records of each type are associated with each offender. Here we are counting only records associated with incidents for which the offender was arrested for, charged with, or convicted of at least one offense from the 12 main crime categories. We summarize the resulting variables (*NINC12*, *NARR12*, *NCHG12*, *NJUD12*, & *NCON12*) below. Table 5 provides overall descriptive statistics for all record types, plus summaries broken down by when the associated incidents occurred relative to the testing window.

Only the data about incident records counts from Table 5 is directly reported in the manuscript. The numbers about other types of records are contextual information about the volume of data that was available.

```
# Table caption.
TCap <- paste("Offender-Level Record Counts (All Crime Categories, Overall and",
              "By When Incident Occurred Relative to Testing Window)")
# Footnote text.
FN <- paste("Only offenders with valid testing window start dates are included.",
            "Only records associated with incidents where the offender was",
            "arrested for, charged with, or convicted of at least one offense",
            "from one of the 12 main crime categories were counted.",
            "Before, during, and after refer to when the incident associated",
            "with the record occurred, not when the arrest date, charge date,",
            "adjudication date, or conviction date occurred relative to the",
            "testing window.")
```

```
BDA <- c("...Before testing window",
         "...During testing window",
         "...After testing window")
rclabs <- c("No. of incidents", BDA,
            "No. of arrest offenses", BDA,
            "No. of prosecutor charges", BDA,
            "No. of adjudicated charges (all)", BDA,
            "No. of adjudicated charges (convictions)", BDA)

# Summarize CHR record counts (both overall and broken down by IWhen).
IDNEW %>%
  rename(NINC12 = HXCat12_Any, NINC12_Before = HXCat12_Any_Before,
         NINC12_During = HXCat12_Any_During, NINC12_After = HXCat12_Any_After) %>%
  select(NINC12, NINC12_Before, NINC12_During, NINC12_After,
         NARR12, NARR12_Before, NARR12_During, NARR12_After,
         NCHG12, NCHG12_Before, NCHG12_During, NCHG12_After,
         NJUD12, NJUD12_Before, NJUD12_During, NJUD12_After,
         NCON12, NCON12_Before, NCON12_During, NCON12_After) %>%
  describe(., quant=c(.25, .50, .75)) %>%
  bind_cols(data.frame(Variable = rclabs), .) %>%
  select(Variable, n, mean, sd, min, max, skew, kurtosis, Q0.25, Q0.5, Q0.75) %>%
  kable(., format = "latex", booktabs = TRUE, digits = 2, row.names = FALSE,
        col.names = CNames, caption = TCap) %>%
  group_rows(" ", 1, 4) %>%
  group_rows(" ", 5, 8) %>%
  group_rows(" ", 9, 12) %>%
  group_rows(" ", 13, 16) %>%
  group_rows(" ", 17, 20) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
# Vectors of variables names and labels.
rcvars  <- c("NINC12", "NARR12", "NCHG12", "NJUD12", "NCON12")
rctypes <- c("Incidents", "Arrest Offenses", "Prosecutor Charges",
             "Adjudicated Charges", "Convictions")

IDNEW %>%
  rename(NINC12 = HXCat12_Any) %>%
  select(OID, NINC12, NARR12, NCHG12, NJUD12, NCON12) %>%
  # Rearrange data so we can easily plot multiple variables.
  pivot_longer(cols = -OID, names_to = "VName", values_to = "Count") %>%
  mutate(VName = ordered(VName, levels = rcvars, labels = rctypes)) ->
  RCCOUNTS
```

```
# Figure caption.
FCap <- paste("\\label{fig:boxplot_RCCOUNTS}",
              "Box Plots for Overall Record Counts by Criminal History Record",
              "Type. Only offenders with valid testing window start dates are",
              "included.",
              "Only records associated with incidents where the offender was",
              "arrested for, charged with, or convicted of at least one offense",
              "from one of the 12 main crime categories were counted.")

bwplot(VName ~ Count, data = RCCOUNTS, factor = 1, xlim = c(-4, 64),
       panel = panel.mybox, par.settings = my.boxes,
       xlab = "Overall Number of Records for Individual Offender")
```

```
# Figure caption.
FCap <- paste("\\label{fig:densityplot_RCCOUNTS}",
              "Density Plots for Overall Record Counts by Criminal History",
              "Record Type. Only offenders with valid testing window start",
              "dates are included.",
              "Only records associated with incidents where the offender was",
              "arrested for, charged with, or convicted of at least one offense",
              "from one of the 12 main crime categories were counted.",
              "Dashed red lines are normal distribution curves (for comparison).")

densityplot(~ Count | VName, data = RCCOUNTS, layout = c(5,1),
            panel = panel.mydensity, jitter.amount = 0.005,
```

Variable	N	Mean	SD	Min	Max	Skew	Kurtosis	Q25	Q50	Q75
No. of incidents	1082	7.94	5.35	1	49	1.86	6.73	4	7	10
...Before testing window	1041	2.49	2.70	0	18	1.66	3.64	0	2	4
...During testing window	1052	0.40	0.72	0	7	3.29	19.74	0	0	1
...After testing window	1082	5.15	4.73	0	49	2.36	11.63	2	4	7
No. of arrest offenses	1082	8.27	5.85	1	55	1.90	6.96	4	7	11
...Before testing window	1041	2.54	2.81	0	18	1.69	3.61	0	2	4
...During testing window	1052	0.40	0.71	0	7	3.14	17.94	0	0	1
...After testing window	1082	5.44	5.24	0	55	2.35	11.49	2	4	8
No. of prosecutor charges	1082	5.21	3.83	0	27	1.49	3.27	2	4	7
...Before testing window	1041	1.43	1.78	0	14	1.92	5.64	0	1	2
...During testing window	1052	0.39	0.81	0	8	3.94	23.98	0	0	1
...After testing window	1082	3.45	3.45	0	22	1.70	4.07	1	3	5
No. of adjudicated charges (all)	1082	10.16	7.04	0	44	1.28	1.99	5	9	13
...Before testing window	1041	2.98	3.43	0	27	1.78	4.73	0	2	4
...During testing window	1052	1.01	2.58	0	30	5.12	37.85	0	0	1
...After testing window	1082	6.31	5.97	0	40	1.50	2.85	2	5	9
No. of adjudicated charges (convictions)	1082	5.56	4.36	0	31	1.90	5.44	3	5	7
...Before testing window	1041	1.63	2.02	0	14	1.77	4.01	0	1	2
...During testing window	1052	0.49	1.51	0	20	6.96	65.47	0	0	0
...After testing window	1082	3.52	3.71	0	27	2.16	7.08	1	3	5

Note: Only offenders with valid testing window start dates are included. Only records associated with incidents where the offender was arrested for, charged with, or convicted of at least one offense from one of the 12 main crime categories were counted. Before, during, and after refer to when the incident associated with the record occurred, not when the arrest date, charge date, adjudication date, or conviction date occurred relative to the testing window.

Table 5: Offender-Level Record Counts (All Crime Categories, Overall and By When Incident Occurred Relative to Testing Window)

```
        strip = strip.custom(par.strip.text = list(cex = .8)),
        xlab = "Overall Number of Records for Offender")
```

In the subsections below, Tables 6, 10, 11, 12, and 13 respectively show the frequency distributions underlying those statistics for incidents, arrest offenses, prosecutor charges, all adjudicated charges, and adjudicated charges that were convictions.

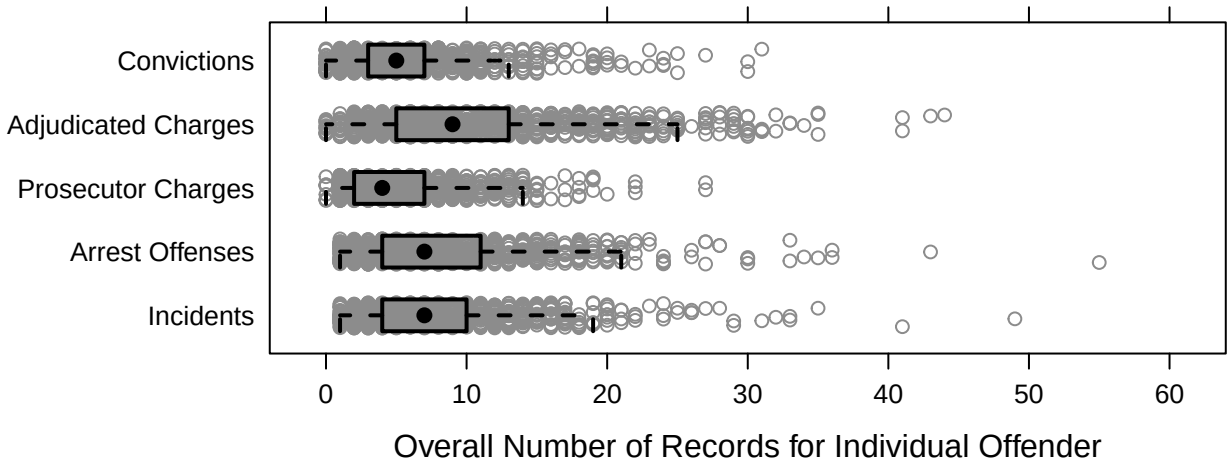

Figure 5: Box Plots for Overall Record Counts by Criminal History Record Type. Only offenders with valid testing window start dates are included. Only records associated with incidents where the offender was arrested for, charged with, or convicted of at least one offense from one of the 12 main crime categories were counted.

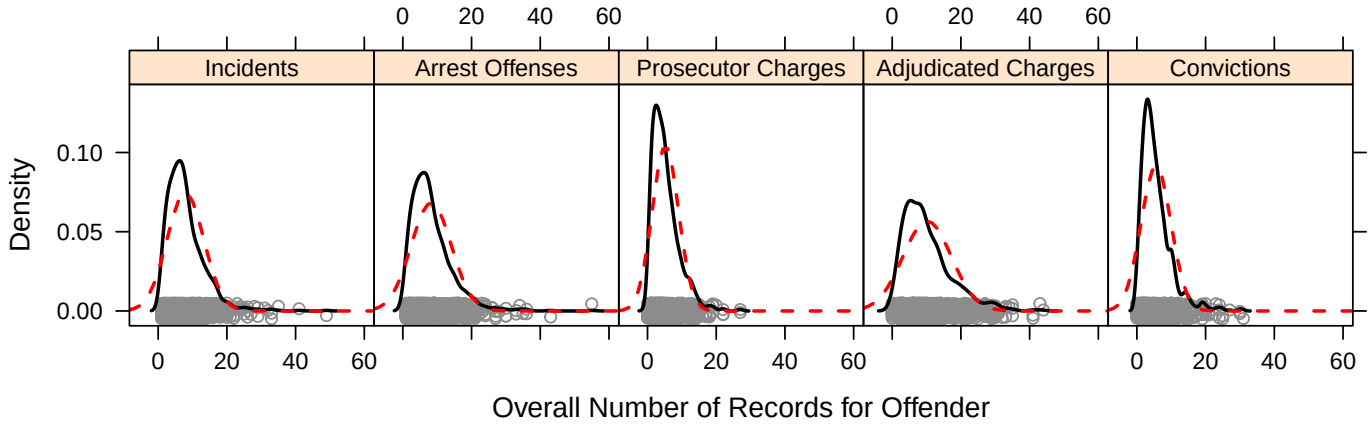

Figure 6: Density Plots for Overall Record Counts by Criminal History Record Type. Only offenders with valid testing window start dates are included. Only records associated with incidents where the offender was arrested for, charged with, or convicted of at least one offense from one of the 12 main crime categories were counted. Dashed red lines are normal distribution curves (for comparison).

5.5.1 Incident Records

```
# Table caption.
TCap <- paste("Frequency Distributions for Number of Incidents",
              "Overall and By When Incident Occurred")
# Footnote text.
FN <- paste0("Only offenders with valid testing window start dates are ",
             "included. Only incidents where the offender was arrested for, ",
             "charged with, or convicted of at least one offense from one of ",
             "the 12 main crime categories were included. ",
             "N = ", format(N_All, big.mark = ","), ". TW, earliest ",
             "testing window.")

# Get frequency distributions.
```

```
IDNEW %>%
  rename(NINC12 = HXCat12_Any, NINC12_Before = HXCat12_Any_Before,
         NINC12_During = HXCat12_Any_During, NINC12_After = HXCat12_Any_After) %>%
  select(OID, NINC12, NINC12_Before, NINC12_During, NINC12_After) %>%
  rename(All = NINC12, Before = NINC12_Before, During = NINC12_During,
         After = NINC12_After) %>%
  pivot_longer(cols = -OID, names_to = "When", values_to = "Value") %>%
  group_by(When, Value) %>%
  count() %>%
  arrange(Value) %>%
  pivot_wider(names_from = When, values_from = n, values_fill = 0) %>%
  mutate(All_p = 100*All/N_All,
         All_v = if_else(is.na(Value),
                         true = as.numeric(NA),
                         false = 100*All/N_All),
         Before_p = 100*Before/N_All,
         Before_v = if_else(is.na(Value),
                            true = as.numeric(NA),
                            false = 100*Before/N_Before),
         During_p = 100*During/N_All,
         During_v = if_else(is.na(Value),
                            true = as.numeric(NA),
                            false = 100*During/N_During),
         After_p = 100*After/N_All,
         After_v = if_else(is.na(Value),
                           true = as.numeric(NA),
                           false = 100*After/N_After)) %>%
  select(Value, All, All_p, All_v, Before, Before_p, Before_v, During,
         During_p, During_v, After, After_p, After_v) %>%
  kable(format = "latex", booktabs = TRUE, digits = 2,
        col.names = c("Value", rep(c("N", "%", "Valid %"), 4)),
        caption = TCap) %>%
  kable_styling() %>%
  add_header_above(c(" " = 1, "Overall" = 3, "Before TW" = 3, "During TW" = 3,
                     "After TW" = 3)) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

5.5.1.1 GEE Models

To more closely examine the incident counts, we fitted two generalized estimating equation (GEE) models to the long-format *IDNEWL* data, after dropping rows for unobserved periods ($Years = 0$) because the counts are missing for those observations.

Model 1 used *When* as a predictor of the period-specific counts, but did not adjust for the period durations. Model 2 supplemented *When* as a predictor with an offset term for $\log(Years)$ to adjust for the period durations. Both models used Poisson distributions, an exchangeable correlation structure, a log link function, and robust standard errors based on the sandwich estimator.

```
HXCat12_Any_gee1 <- IDNEWL %>%
  # Keep only observations for observed periods (i.e., those with Years > 0).
  filter(Years > 0) %>%
  geeglm(HXCat12_Any ~ When,
         family = poisson(link = "log"), data = .,
         id = OID, corstr = "exchangeable", std.err = "san.se")

HXCat12_Any_gee2 <- IDNEWL %>%
  filter(Years > 0) %>%
  geeglm(HXCat12_Any ~ When + offset(log(Years)),
         family = poisson(link = "log"), data = .,
         id = OID, corstr = "exchangeable", std.err = "san.se")
```

Table 7 below shows the parameter estimates for both GEE models.

```
TCap <- paste("Generalized Estimating Equation Models for Incident Counts",
              "Predicted by When Incidents Occurred")
FN <- paste ("\\item * p < .05, null hypothesis value outside 95\\% confidence",
```

	Overall			Before TW			During TW			After TW		
Value	N	%	Valid %	N	%	Valid %	N	%	Valid %	N	%	Valid %
0	0	0.00	0.00	276	25.51	26.51	714	65.99	67.87	90	8.32	8.32
1	41	3.79	3.79	210	19.41	20.17	284	26.25	27.00	126	11.65	11.65
2	72	6.65	6.65	149	13.77	14.31	41	3.79	3.90	140	12.94	12.94
3	91	8.41	8.41	127	11.74	12.20	6	0.55	0.57	135	12.48	12.48
4	87	8.04	8.04	89	8.23	8.55	3	0.28	0.29	106	9.80	9.80
5	106	9.80	9.80	52	4.81	5.00	1	0.09	0.10	82	7.58	7.58
6	98	9.06	9.06	58	5.36	5.57	1	0.09	0.10	83	7.67	7.67
7	113	10.44	10.44	28	2.59	2.69	2	0.18	0.19	72	6.65	6.65
8	87	8.04	8.04	9	0.83	0.86	0	0.00	0.00	55	5.08	5.08
9	76	7.02	7.02	14	1.29	1.34	0	0.00	0.00	39	3.60	3.60
10	50	4.62	4.62	9	0.83	0.86	0	0.00	0.00	30	2.77	2.77
11	45	4.16	4.16	8	0.74	0.77	0	0.00	0.00	28	2.59	2.59
12	44	4.07	4.07	5	0.46	0.48	0	0.00	0.00	24	2.22	2.22
13	30	2.77	2.77	1	0.09	0.10	0	0.00	0.00	18	1.66	1.66
14	31	2.87	2.87	2	0.18	0.19	0	0.00	0.00	12	1.11	1.11
15	22	2.03	2.03	2	0.18	0.19	0	0.00	0.00	9	0.83	0.83
16	21	1.94	1.94	1	0.09	0.10	0	0.00	0.00	9	0.83	0.83
17	20	1.85	1.85	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
18	7	0.65	0.65	1	0.09	0.10	0	0.00	0.00	6	0.55	0.55
19	6	0.55	0.55	0	0.00	0.00	0	0.00	0.00	4	0.37	0.37
20	8	0.74	0.74	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
21	4	0.37	0.37	0	0.00	0.00	0	0.00	0.00	2	0.18	0.18
22	3	0.28	0.28	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
23	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
24	3	0.28	0.28	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
25	2	0.18	0.18	0	0.00	0.00	0	0.00	0.00	2	0.18	0.18
26	3	0.28	0.28	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
27	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
28	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	3	0.28	0.28
29	2	0.18	0.18	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
31	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
32	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
33	2	0.18	0.18	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
35	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
40	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
41	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
49	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
NA	0	0.00	NA	41	3.79	NA	30	2.77	NA	0	0.00	NA

Note: Only offenders with valid testing window start dates are included. Only incidents where the offender was arrested for, charged with, or convicted of at least one offense from one of the 12 main crime categories were included. N = 1,082. TW, earliest testing window.

Table 6: Frequency Distributions for Number of Incidents Overall and By When Incident Occurred

```
        "interval based on z-score.",
        "\\\\\n\\item Note.",
        "The data contain up to 3 longitudinal observations (one per",
        "period) for each offender (i.e., cluster).",
        "The reference level for When was the period after the testing",
        "window.",
```

```
            "Both models used a log link function, exchangeable correlation",
            "structure, a Poisson distribution, and robust standard errors",
            "(sandwich estimator).",
            "Model 1 did not use an offset.",
            "Model 2 used log(Years) as an offset term.")

texreg(list(HXCat12_Any_gee1, HXCat12_Any_gee2), booktabs = TRUE, dcolumn = TRUE,
       threeparttable = TRUE, fontsize = "normalsize", table = TRUE,
       use.packages = FALSE, ci.force = TRUE, label = "tab:HXCat12_Any_texreg",
       stars = 0.05, caption = TCap, custom.note = FN)
```

	Model 1	Model 2
(Intercept)	1.64*	−1.25*
	[1.59; 1.69]	[−1.31; −1.20]
WhenBefore	−0.73*	−0.30*
	[−0.82; −0.64]	[−0.38; −0.23]
WhenDuring	−2.56*	1.46*
	[−2.68; −2.43]	[1.33; 1.58]
Scale parameter: gamma	2.86	2.69
Scale parameter: SE	0.17	0.15
Correlation parameter: alpha	−0.04	0.03
Correlation parameter: SE	0.02	0.02
Num. obs.	3175	3175
Num. clust.	1082	1082

* p < .05, null hypothesis value outside 95% confidence interval based on z-score.

Note. The data contain up to 3 longitudinal observations (one per period) for each offender (i.e., cluster). The reference level for When was the period after the testing window. Both models used a log link function, exchangeable correlation structure, a Poisson distribution, and robust standard errors (sandwich estimator). Model 1 did not use an offset. Model 2 used log(Years) as an offset term.

Table 7: Generalized Estimating Equation Models for Incident Counts Predicted by When Incidents Occurred

5.5.1.2 Estimated Marginal Means Table 8 provides the estimated marginal means obtained from Models 1 and 2, along with corresponding confidence intervals. The units of measurement for those means are counts per offender for Model 1 (which ignore period duration) and counts per offender-year for Model 2 (which adjust for duration and are thus considered annual incidence rates).

```
# Table caption.
TCap <- paste("Marginal Means of Incident Counts and Incidence Rates",
              "By When Incidents Occurred")

# Table footnote.
FN <- paste("Model 1 did not use an offset, so means are counts.",
            "Model 2 used log(Years) as an offset term, so means are annual",
            "incidence rates.",
            "Confidence intervals are based on z-scores.")

HXCat12_Any_emmeans1 <- emmeans(HXCat12_Any_gee1, specs = pairwise ~ When,
                            type = "response", ref = "During")
HXCat12_Any_emmeans2 <- emmeans(HXCat12_Any_gee2, specs = pairwise ~ When,
                            type = "response", ref = "During")

as_tibble(HXCat12_Any_emmeans1$emmeans) %>%
  full_join(x = ., y = as_tibble(HXCat12_Any_emmeans2$emmeans)) %>%
  mutate(Model = c(1, 1, 1, 2, 2, 2),
         When = factor(When, levels = c("Before", "During", "After"))) %>%
  arrange(Model, When) %>%
```

```
  select(When, rate, SE, df, asymp.LCL, asymp.UCL) %>%
  kable(format = "latex", booktabs = TRUE, caption = TCap, digits = 2,
        col.names = c("When", "Rate", "SE", "df", "LCL", "UCL")) %>%
  kable_styling() %>%
  add_header_above(c(" " = 4, "95% CI" = 2)) %>%
  group_rows(group_label = "Model 1: Mean per offender", start_row = 1,
             end_row = 3, italic = TRUE) %>%
  group_rows(group_label = "Model 2: Mean per offender-year",
             start_row = 4, end_row = 6, italic = TRUE) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
## Joining, by = c("When", "rate", "SE", "df", "asymp.LCL", "asymp.UCL")
```

				95% CI	
When	Rate	SE	df	LCL	UCL
Model 1: Mean per offender					
Before	2.49	0.08	Inf	2.33	2.66
During	0.40	0.02	Inf	0.36	0.45
After	5.15	0.14	Inf	4.88	5.44
Model 2: Mean per offender-year					
Before	2.13	0.06	Inf	2.01	2.27
During	12.43	0.69	Inf	11.15	13.86
After	2.89	0.08	Inf	2.74	3.06

Note:
Model 1 did not use an offset, so means are counts. Model 2 used log(Years) as an offset term, so means are annual incidence rates. Confidence intervals are based on z-scores.

Table 8: Marginal Means of Incident Counts and Incidence Rates By When Incidents Occurred

5.5.1.3 Contrasts for Ratios of Estimated Marginal Means Finally, we also estimated a set of contrasts to compare the marginal means from Models 1 and 2 (Table 9). Each of these contrasts estimates the ratios of a pair of means rather than a raw difference between means. This is because of the nature of the underlying Poisson GEE model.

```
TCap <- paste("Contrasts Estimating Ratios of Marginal Means for Incident",
              "Counts and Incidence Rates By When Incidents Occurred")

# Table footnote.
FN <- paste("Model 1 did not use an offset, so means are counts.",
            "Model 2 used log(Years) as an offset term, so means are annual",
            "incidence rates.",
            "These contrasts estimate ratios of those quantities and used",
            "Tukey's method to adjust for multiple comparisons.",
            "Confidence intervals are based on z-scores.")

as_tibble(confint(HXCat12_Any_emmeans1$contrasts)) %>%
  full_join(x = ., y = as_tibble(confint(HXCat12_Any_emmeans2$contrasts))) %>%
  kable(format = "latex", booktabs = TRUE, caption = TCap, digits = 2,
        col.names = c("Contrast", "Ratio", "SE", "df", "LCL", "UCL")) %>%
  kable_styling() %>%
  add_header_above(c(" " = 4, "95% CI" = 2)) %>%
  group_rows(group_label = "Model 1: Ratios of means per offender",
             start_row = 1, end_row = 3, italic = TRUE) %>%
  group_rows(group_label = "Model 2: Ratios of means per offender-year",
             start_row = 4, end_row = 6, italic = TRUE) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
## Joining, by = c("contrast", "ratio", "SE", "df", "asymp.LCL", "asymp.UCL")
```

Contrast	Ratio	SE	df	95% CI LCL	95% CI UCL
Model 1: Ratios of means per offender					
After / Before	2.07	0.09	Inf	1.87	2.30
After / During	12.89	0.84	Inf	11.07	15.02
Before / During	6.22	0.40	Inf	5.35	7.23
Model 2: Ratios of means per offender-year					
After / Before	1.36	0.05	Inf	1.24	1.48
After / During	0.23	0.02	Inf	0.20	0.27
Before / During	0.17	0.01	Inf	0.15	0.20

Note:
Model 1 did not use an offset, so means are counts. Model 2 used log(Years) as an offset term, so means are annual incidence rates. These contrasts estimate ratios of those quantities and used Tukey's method to adjust for multiple comparisons. Confidence intervals are based on z-scores.

Table 9: Contrasts Estimating Ratios of Marginal Means for Incident Counts and Incidence Rates By When Incidents Occurred

5.5.2 Arrest Offense Records

Output in this section is not directly used in the manuscript. It is here because it helped us to understand the volume of raw data that was being processed along the way to constructing the data sets used for the manuscript.

```
# Table caption.
TCap <- paste("Frequency Distributions for Number of Arrest Offenses (All",
              "Crime Categories) Overall and By When Incident Occurred")
# Footnote text.
FN <- paste0("N = ", format(N_All, big.mark = ","), ". TW, earliest ",
             "testing window. ",
             "Only arrest records for offenders with valid testing window start ",
             "dates and with offenses from the 12 main crime categories were ",
             "included. ",
             "Before, during, and after refer to when the incident associated ",
             "with the record occurred, not when the arrest date occurred ",
             "relative to the testing window. For example, an offender could be ",
             "arrested during the testing window for an incident that occurred ",
             "before it began. That would show up here in the before column.")

# Get frequency distributions.
IDNEW %>%
  select(OID, NARR12, NARR12_Before, NARR12_During, NARR12_After) %>%
  rename(All = NARR12, Before = NARR12_Before, During = NARR12_During,
         After = NARR12_After) %>%
  pivot_longer(cols = -OID, names_to = "When", values_to = "Value") %>%
  group_by(When, Value) %>%
  count() %>%
  arrange(Value) %>%
  pivot_wider(names_from = When, values_from = n, values_fill = 0) %>%
  mutate(All_p = 100*All/N_All,
         All_v = if_else(is.na(Value),
                         true = as.numeric(NA),
                         false = 100*All/N_All),
         Before_p = 100*Before/N_All,
         Before_v = if_else(is.na(Value),
                            true = as.numeric(NA),
                            false = 100*Before/N_Before),
         During_p = 100*During/N_All,
         During_v = if_else(is.na(Value),
                            true = as.numeric(NA),
                            false = 100*During/N_During),
```

```
       After_p = 100*After/N_All,
       After_v = if_else(is.na(Value),
                         true = as.numeric(NA),
                         false = 100*After/N_After)) %>%
select(Value, All, All_p, All_v, Before, Before_p, Before_v, During,
       During_p, During_v, After, After_p, After_v) %>%
kable(format = "latex", booktabs = TRUE, digits = 2,
      col.names = c("Value", rep(c("N", "%", "Valid %"), 4)),
      caption = TCap) %>%
kable_styling() %>%
add_header_above(c(" " = 1, "Overall" = 3, "Before TW" = 3, "During TW" = 3,
                   "After TW" = 3)) %>%
footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
       threeparttable = TRUE)
```

	Overall			Before TW			During TW			After TW		
Value	N	%	Valid %	N	%	Valid %	N	%	Valid %	N	%	Valid %
0	0	0.00	0.00	281	25.97	26.99	717	66.27	68.16	110	10.17	10.17
1	47	4.34	4.34	208	19.22	19.98	280	25.88	26.62	118	10.91	10.91
2	79	7.30	7.30	142	13.12	13.64	41	3.79	3.90	137	12.66	12.66
3	79	7.30	7.30	127	11.74	12.20	6	0.55	0.57	125	11.55	11.55
4	89	8.23	8.23	86	7.95	8.26	5	0.46	0.48	99	9.15	9.15
5	100	9.24	9.24	55	5.08	5.28	1	0.09	0.10	72	6.65	6.65
6	86	7.95	7.95	49	4.53	4.71	0	0.00	0.00	76	7.02	7.02
7	109	10.07	10.07	33	3.05	3.17	2	0.18	0.19	64	5.91	5.91
8	83	7.67	7.67	12	1.11	1.15	0	0.00	0.00	53	4.90	4.90
9	68	6.28	6.28	15	1.39	1.44	0	0.00	0.00	40	3.70	3.70
10	52	4.81	4.81	12	1.11	1.15	0	0.00	0.00	39	3.60	3.60
11	50	4.62	4.62	5	0.46	0.48	0	0.00	0.00	33	3.05	3.05
12	50	4.62	4.62	7	0.65	0.67	0	0.00	0.00	21	1.94	1.94
13	26	2.40	2.40	1	0.09	0.10	0	0.00	0.00	15	1.39	1.39
14	28	2.59	2.59	3	0.28	0.29	0	0.00	0.00	22	2.03	2.03
15	31	2.87	2.87	2	0.18	0.19	0	0.00	0.00	7	0.65	0.65
16	15	1.39	1.39	2	0.18	0.19	0	0.00	0.00	15	1.39	1.39
17	13	1.20	1.20	0	0.00	0.00	0	0.00	0.00	7	0.65	0.65
18	18	1.66	1.66	1	0.09	0.10	0	0.00	0.00	7	0.65	0.65
19	9	0.83	0.83	0	0.00	0.00	0	0.00	0.00	3	0.28	0.28
20	11	1.02	1.02	0	0.00	0.00	0	0.00	0.00	2	0.18	0.18
21	9	0.83	0.83	0	0.00	0.00	0	0.00	0.00	3	0.28	0.28
22	4	0.37	0.37	0	0.00	0.00	0	0.00	0.00	2	0.18	0.18
23	2	0.18	0.18	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
24	6	0.55	0.55	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
25	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
26	2	0.18	0.18	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
27	3	0.28	0.28	0	0.00	0.00	0	0.00	0.00	2	0.18	0.18
28	2	0.18	0.18	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
29	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
30	3	0.28	0.28	0	0.00	0.00	0	0.00	0.00	2	0.18	0.18
31	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
33	2	0.18	0.18	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
34	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
35	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
36	2	0.18	0.18	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
42	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
43	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
55	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
NA	0	0.00	NA	41	3.79	NA	30	2.77	NA	0	0.00	NA

Note: N = 1,082. TW, earliest testing window. Only arrest records for offenders with valid testing window start dates and with offenses from the 12 main crime categories were included. Before, during, and after refer to when the incident associated with the record occurred, not when the arrest date occurred relative to the testing window. For example, an offender could be arrested during the testing window for an incident that occurred before it began. That would show up here in the before column.

Table 10: Frequency Distributions for Number of Arrest Offenses (All Crime Categories) Overall and By When Incident Occurred

5.5.3 Prosecutor Charge Records

Output in this section is not directly used in the manuscript. It is here because it helped us to understand the volume of raw data that was being processed along the way to constructing the data sets used for the manuscript.

```
# Table caption.
TCap <- paste("Frequency Distributions for Number of Prosecutor Charges (All",
              "Crime Categories) Overall and By When Incident Occurred")
# Footnote text.
FN <- paste0("N = ", format(N_All, big.mark = ","), ". TW, earliest ",
             "testing window. ",
             "Only charge records for offenders with valid testing window start ",
             "dates and with charges from the 12 main crime categories were ",
             "included. ",
             "Before, during, and after refer to when the incident associated ",
             "with the record occurred, not when the charge date occurred ",
             "relative to the testing window. For example, an offender could be ",
             "charged during the testing window for an incident that occurred ",
             "before it began. That would show up here in the before column.")

# Get frequency distributions.
IDNEW %>%
  select(OID, NCHG12, NCHG12_Before, NCHG12_During, NCHG12_After) %>%
  rename(All = NCHG12, Before = NCHG12_Before, During = NCHG12_During,
         After = NCHG12_After) %>%
  pivot_longer(cols = -OID, names_to = "When", values_to = "Value") %>%
  group_by(When, Value) %>%
  count() %>%
  arrange(Value) %>%
  pivot_wider(names_from = When, values_from = n, values_fill = 0) %>%
  mutate(All_p = 100*All/N_All,
         All_v = if_else(is.na(Value),
                         true = as.numeric(NA),
                         false = 100*All/N_All),
         Before_p = 100*Before/N_All,
         Before_v = if_else(is.na(Value),
                            true = as.numeric(NA),
                            false = 100*Before/N_Before),
         During_p = 100*During/N_All,
         During_v = if_else(is.na(Value),
                            true = as.numeric(NA),
                            false = 100*During/N_During),
         After_p = 100*After/N_All,
         After_v = if_else(is.na(Value),
                           true = as.numeric(NA),
                           false = 100*After/N_After)) %>%
  select(Value, All, All_p, All_v, Before, Before_p, Before_v, During,
         During_p, During_v, After, After_p, After_v) %>%
  kable(format = "latex", booktabs = TRUE, digits = 2,
        col.names = c("Value", rep(c("N", "%", "Valid %"), 4)),
        caption = TCap) %>%
  kable_styling() %>%
  add_header_above(c(" " = 1, "Overall" = 3, "Before TW" = 3, "During TW" = 3,
                     "After TW" = 3)) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

	Overall			Before TW			During TW			After TW		
Value	N	%	Valid %	N	%	Valid %	N	%	Valid %	N	%	Valid %
0	9	0.83	0.83	414	38.26	39.77	752	69.50	71.48	189	17.47	17.47
1	136	12.57	12.57	265	24.49	25.46	244	22.55	23.19	185	17.10	17.10
2	143	13.22	13.22	136	12.57	13.06	33	3.05	3.14	162	14.97	14.97
3	144	13.31	13.31	98	9.06	9.41	10	0.92	0.95	143	13.22	13.22
4	126	11.65	11.65	66	6.10	6.34	7	0.65	0.67	92	8.50	8.50
5	128	11.83	11.83	25	2.31	2.40	2	0.18	0.19	76	7.02	7.02
6	83	7.67	7.67	18	1.66	1.73	0	0.00	0.00	63	5.82	5.82
7	80	7.39	7.39	7	0.65	0.67	3	0.28	0.29	52	4.81	4.81
8	53	4.90	4.90	6	0.55	0.58	1	0.09	0.10	22	2.03	2.03
9	43	3.97	3.97	4	0.37	0.38	0	0.00	0.00	27	2.50	2.50
10	34	3.14	3.14	0	0.00	0.00	0	0.00	0.00	20	1.85	1.85
11	20	1.85	1.85	0	0.00	0.00	0	0.00	0.00	17	1.57	1.57
12	25	2.31	2.31	0	0.00	0.00	0	0.00	0.00	11	1.02	1.02
13	20	1.85	1.85	1	0.09	0.10	0	0.00	0.00	7	0.65	0.65
14	11	1.02	1.02	1	0.09	0.10	0	0.00	0.00	4	0.37	0.37
15	8	0.74	0.74	0	0.00	0.00	0	0.00	0.00	2	0.18	0.18
16	2	0.18	0.18	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
17	3	0.28	0.28	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
18	4	0.37	0.37	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
19	4	0.37	0.37	0	0.00	0.00	0	0.00	0.00	3	0.28	0.28
20	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
22	3	0.28	0.28	0	0.00	0.00	0	0.00	0.00	3	0.28	0.28
27	2	0.18	0.18	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
NA	0	0.00	NA	41	3.79	NA	30	2.77	NA	0	0.00	NA

Note: N = 1,082. TW, earliest testing window. Only charge records for offenders with valid testing window start dates and with charges from the 12 main crime categories were included. Before, during, and after refer to when the incident associated with the record occurred, not when the charge date occurred relative to the testing window. For example, an offender could be charged during the testing window for an incident that occurred before it began. That would show up here in the before column.

Table 11: Frequency Distributions for Number of Prosecutor Charges (All Crime Categories) Overall and By When Incident Occurred

5.5.4 Adjudicated Charge Records (All)

Output in this section is not directly used in the manuscript. It is here because it helped us to understand the volume of raw data that was being processed along the way to constructing the data sets used for the manuscript.

```
# Table caption.
TCap <- paste("Frequency Distributions for Number of Adjudicated Charges (All",
              "Crime Categories) Overall and By When Incident Occurred")
# Footnote text.
FN <- paste0("N = ", format(N_All, big.mark = ","), ". TW, earliest ",
             "testing window. ",
             "Only records for offenders with valid testing window start dates ",
             "and with charges from one of the 12 main crime categories were ",
             "included. ",
             "Before, during, and after refer to when the incident associated ",
             "with the record occurred, not when the adjudication date occurred ",
             "relative to the testing window. For example, a charge could be ",
             "adjudicated during the testing window for an incident that ",
             "occurred before it began. That would show up here in the before ",
             "column.")

# Get frequency distributions.
```

```
IDNEW %>%
  select(OID, NJUD12, NJUD12_Before, NJUD12_During, NJUD12_After) %>%
  rename(All = NJUD12, Before = NJUD12_Before, During = NJUD12_During,
         After = NJUD12_After) %>%
  pivot_longer(cols = -OID, names_to = "When", values_to = "Value") %>%
  group_by(When, Value) %>%
  count() %>%
  arrange(Value) %>%
  pivot_wider(names_from = When, values_from = n, values_fill = 0) %>%
  mutate(All_p = 100*All/N_All,
         All_v = if_else(is.na(Value),
                         true = as.numeric(NA),
                         false = 100*All/N_All),
         Before_p = 100*Before/N_All,
         Before_v = if_else(is.na(Value),
                            true = as.numeric(NA),
                            false = 100*Before/N_Before),
         During_p = 100*During/N_All,
         During_v = if_else(is.na(Value),
                            true = as.numeric(NA),
                            false = 100*During/N_During),
         After_p = 100*After/N_All,
         After_v = if_else(is.na(Value),
                           true = as.numeric(NA),
                           false = 100*After/N_After)) %>%
  select(Value, All, All_p, All_v, Before, Before_p, Before_v, During,
         During_p, During_v, After, After_p, After_v) %>%
  kable(format = "latex", booktabs = TRUE, digits = 2,
        col.names = c("Value", rep(c("N", "%", "Valid %"), 4)),
        caption = TCap) %>%
  kable_styling() %>%
  add_header_above(c(" " = 1, "Overall" = 3, "Before TW" = 3, "During TW" = 3,
                     "After TW" = 3)) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

	Overall			Before TW			During TW			After TW		
Value	N	%	Valid %	N	%	Valid %	N	%	Valid %	N	%	Valid %
0	12	1.11	1.11	316	29.21	30.36	754	69.69	71.67	162	14.97	14.97
1	22	2.03	2.03	139	12.85	13.35	83	7.67	7.89	64	5.91	5.91
2	53	4.90	4.90	131	12.11	12.58	75	6.93	7.13	101	9.33	9.33
3	70	6.47	6.47	110	10.17	10.57	40	3.70	3.80	89	8.23	8.23
4	69	6.38	6.38	90	8.32	8.65	29	2.68	2.76	90	8.32	8.32
5	88	8.13	8.13	58	5.36	5.57	25	2.31	2.38	90	8.32	8.32
6	74	6.84	6.84	46	4.25	4.42	11	1.02	1.05	76	7.02	7.02
7	62	5.73	5.73	41	3.79	3.94	9	0.83	0.86	60	5.55	5.55
8	84	7.76	7.76	33	3.05	3.17	4	0.37	0.38	55	5.08	5.08
9	76	7.02	7.02	26	2.40	2.50	5	0.46	0.48	45	4.16	4.16
10	59	5.45	5.45	15	1.39	1.44	2	0.18	0.19	44	4.07	4.07
11	48	4.44	4.44	6	0.55	0.58	3	0.28	0.29	37	3.42	3.42
12	49	4.53	4.53	7	0.65	0.67	3	0.28	0.29	26	2.40	2.40
13	48	4.44	4.44	8	0.74	0.77	2	0.18	0.19	23	2.13	2.13
14	44	4.07	4.07	6	0.55	0.58	1	0.09	0.10	15	1.39	1.39
15	20	1.85	1.85	2	0.18	0.19	0	0.00	0.00	17	1.57	1.57
16	28	2.59	2.59	0	0.00	0.00	1	0.09	0.10	14	1.29	1.29
17	18	1.66	1.66	1	0.09	0.10	0	0.00	0.00	7	0.65	0.65
18	22	2.03	2.03	3	0.28	0.29	0	0.00	0.00	8	0.74	0.74
19	19	1.76	1.76	0	0.00	0.00	0	0.00	0.00	11	1.02	1.02
20	19	1.76	1.76	2	0.18	0.19	2	0.18	0.19	10	0.92	0.92
21	13	1.20	1.20	0	0.00	0.00	0	0.00	0.00	7	0.65	0.65
22	16	1.48	1.48	0	0.00	0.00	0	0.00	0.00	7	0.65	0.65
23	11	1.02	1.02	0	0.00	0.00	0	0.00	0.00	5	0.46	0.46
24	7	0.65	0.65	0	0.00	0.00	1	0.09	0.10	4	0.37	0.37
25	9	0.83	0.83	0	0.00	0.00	0	0.00	0.00	3	0.28	0.28
26	1	0.09	0.09	0	0.00	0.00	1	0.09	0.10	1	0.09	0.09
27	7	0.65	0.65	1	0.09	0.10	0	0.00	0.00	2	0.18	0.18
28	5	0.46	0.46	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
29	7	0.65	0.65	0	0.00	0.00	0	0.00	0.00	2	0.18	0.18
30	7	0.65	0.65	0	0.00	0.00	1	0.09	0.10	2	0.18	0.18
31	3	0.28	0.28	0	0.00	0.00	0	0.00	0.00	2	0.18	0.18
32	2	0.18	0.18	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
33	2	0.18	0.18	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
34	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
35	3	0.28	0.28	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
40	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
41	2	0.18	0.18	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
43	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
44	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
NA	0	0.00	NA	41	3.79	NA	30	2.77	NA	0	0.00	NA

Note: N = 1,082. TW, earliest testing window. Only records for offenders with valid testing window start dates and with charges from one of the 12 main crime categories were included. Before, during, and after refer to when the incident associated with the record occurred, not when the adjudication date occurred relative to the testing window. For example, a charge could be adjudicated during the testing window for an incident that occurred before it began. That would show up here in the before column.

Table 12: Frequency Distributions for Number of Adjudicated Charges (All Crime Categories) Overall and By When Incident Occurred

5.5.5 Adjudicated Charge Records (Convictions)

Output in this section is not directly used in the manuscript. It is here because it helped us to understand the volume of raw data that was being processed along the way to constructing the data sets used for the manuscript.

```
# Table caption.
TCap <- paste("Frequency Distributions for Number of Adjudicated Charges (All",
              "Crime Categories, Convictions) Overall and By When Incident",
              "Occurred")
# Footnote text.
FN <- paste0("N = ", format(N_All, big.mark = ","), ". TW, earliest ",
             "testing window. ",
             "Only conviction records for offenders with valid testing window ",
             "start dates and with charges from one of the 12 main crime ",
             "categories were included. ",
             "Before, during, and after refer to when the incident associated ",
             "with the record occurred, not when the adjudication date occurred ",
             "relative to the testing window. For example, an offender could be ",
             "convicted during the testing window for an incident that occurred ",
             "before it began. That would show up here in the before column.")

# Get frequency distributions.
IDNEW %>%
  select(OID, NCON12, NCON12_Before, NCON12_During, NCON12_After) %>%
  rename(All = NCON12, Before = NCON12_Before, During = NCON12_During,
         After = NCON12_After) %>%
  pivot_longer(cols = -OID, names_to = "When", values_to = "Value") %>%
  group_by(When, Value) %>%
  count() %>%
  arrange(Value) %>%
  pivot_wider(names_from = When, values_from = n, values_fill = 0) %>%
  mutate(All_p = 100*All/N_All,
         All_v = if_else(is.na(Value),
                         true = as.numeric(NA),
                         false = 100*All/N_All),
         Before_p = 100*Before/N_All,
         Before_v = if_else(is.na(Value),
                            true = as.numeric(NA),
                            false = 100*Before/N_Before),
         During_p = 100*During/N_All,
         During_v = if_else(is.na(Value),
                            true = as.numeric(NA),
                            false = 100*During/N_During),
         After_p = 100*After/N_All,
         After_v = if_else(is.na(Value),
                           true = as.numeric(NA),
                           false = 100*After/N_After)) %>%
  select(Value, All, All_p, All_v, Before, Before_p, Before_v, During,
         During_p, During_v, After, After_p, After_v) %>%
  kable(format = "latex", booktabs = TRUE, digits = 2,
        col.names = c("Value", rep(c("N", "%", "Valid %"), 4)),
        caption = TCap) %>%
  kable_styling() %>%
  add_header_above(c(" " = 1, "Overall" = 3, "Before TW" = 3, "During TW" = 3,
                     "After TW" = 3)) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

	Overall			Before TW			During TW			After TW		
Value	N	%	Valid %	N	%	Valid %	N	%	Valid %	N	%	Valid %
0	29	2.68	2.68	405	37.43	38.90	813	75.14	77.28	202	18.67	18.67
1	90	8.32	8.32	228	21.07	21.90	135	12.48	12.83	147	13.59	13.59
2	132	12.20	12.20	148	13.68	14.22	54	4.99	5.13	185	17.10	17.10
3	151	13.96	13.96	105	9.70	10.09	23	2.13	2.19	136	12.57	12.57
4	138	12.75	12.75	54	4.99	5.19	8	0.74	0.76	116	10.72	10.72
5	114	10.54	10.54	44	4.07	4.23	5	0.46	0.48	72	6.65	6.65
6	96	8.87	8.87	26	2.40	2.50	3	0.28	0.29	63	5.82	5.82
7	76	7.02	7.02	10	0.92	0.96	5	0.46	0.48	34	3.14	3.14
8	61	5.64	5.64	9	0.83	0.86	0	0.00	0.00	33	3.05	3.05
9	35	3.23	3.23	6	0.55	0.58	0	0.00	0.00	20	1.85	1.85
10	48	4.44	4.44	2	0.18	0.19	0	0.00	0.00	21	1.94	1.94
11	29	2.68	2.68	2	0.18	0.19	0	0.00	0.00	11	1.02	1.02
12	13	1.20	1.20	1	0.09	0.10	1	0.09	0.10	10	0.92	0.92
13	9	0.83	0.83	0	0.00	0.00	0	0.00	0.00	4	0.37	0.37
14	15	1.39	1.39	1	0.09	0.10	1	0.09	0.10	7	0.65	0.65
15	10	0.92	0.92	0	0.00	0.00	1	0.09	0.10	5	0.46	0.46
16	5	0.46	0.46	0	0.00	0.00	2	0.18	0.19	4	0.37	0.37
17	3	0.28	0.28	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
18	2	0.18	0.18	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
19	8	0.74	0.74	0	0.00	0.00	0	0.00	0.00	2	0.18	0.18
20	4	0.37	0.37	0	0.00	0.00	1	0.09	0.10	0	0.00	0.00
21	2	0.18	0.18	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
22	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	2	0.18	0.18
23	2	0.18	0.18	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
24	3	0.28	0.28	0	0.00	0.00	0	0.00	0.00	2	0.18	0.18
25	2	0.18	0.18	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
26	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
27	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
30	2	0.18	0.18	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
31	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
NA	0	0.00	NA	41	3.79	NA	30	2.77	NA	0	0.00	NA

Note: N = 1,082. TW, earliest testing window. Only conviction records for offenders with valid testing window start dates and with charges from one of the 12 main crime categories were included. Before, during, and after refer to when the incident associated with the record occurred, not when the adjudication date occurred relative to the testing window. For example, an offender could be convicted during the testing window for an incident that occurred before it began. That would show up here in the before column.

Table 13: Frequency Distributions for Number of Adjudicated Charges (All Crime Categories, Convictions) Overall and By When Incident Occurred

5.6 Crime Category Counts (Arrested, Charged, or Convicted)

Table 14 summarizes how many of 12 different crime categories each offender was *arrested for, charged with, or convicted of* according to the CHR data. This sums a set of binary transforms on the incident count variables associated with each crime category in *HXCat12*. It is possible to have multiple arrest offense records, charge records, or conviction records for a specific category, but we only count each distinct category once at most. The *Excluded user-missing* value is not counted as a category, so the possible range of values for these counts is 0-12. We report both overall results, and results broken down by when the incidents associated with the arrest, charge, and conviction records occurred relative to the earliest testing window. Table 15 shows the frequency distributions underlying those summaries.

```
# Table caption.
TCap <- paste("Offender-Level Crime Category Counts (Overall and By When",
              "Incident Occurred)")
# Footnote text.
FN <- paste("Only offenders with valid testing window start dates are included.",
            "Only records associated with incidents where the offender was ",
            "arrested for, charged with, or convicted of at least one offense ",
            "from one of the 12 main crime categories were counted. ",
            "Before, during, and after refer to when the incident associated",
            "with the record occurred, not when the arrest date, charge date,",
            "adjudication date, or conviction date occurred relative to the",
            "testing window.")

cclabs <- c("No. of crime categories (arrested, charged, or convicted)", BDA)

# Summarize crime category counts (both overall and broken down by IWhen).
IDNEW %>%
  select(HXCat12_Count, HXCat12_Count_Before, HXCat12_Count_During,
         HXCat12_Count_After) %>%
  describe(., quant=c(.25, .50, .75)) %>%
  bind_cols(data.frame(Variable = cclabs), .) %>%
  select(Variable, n, mean, sd, min, max, skew, kurtosis, Q0.25, Q0.5, Q0.75) %>%
  kable(., format = "latex", booktabs = TRUE, digits = 2, row.names = FALSE,
        col.names = CNames, caption = TCap) %>%
  column_spec(column = 1, width = "6 cm") %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

Variable	N	Mean	SD	Min	Max	Skew	Kurtosis	Q25	Q50	Q75
No. of crime categories (arrested, charged, or convicted)	1082	4.52	1.77	1	10	0.17	-0.38	3	4	6
...Before testing window	1041	1.92	1.72	0	8	0.77	0.05	0	2	3
...During testing window	1052	0.49	0.86	0	7	2.22	6.53	0	0	1
...After testing window	1082	3.22	1.89	0	10	0.26	-0.31	2	3	4

Note: Only offenders with valid testing window start dates are included. Only records associated with incidents where the offender was arrested for, charged with, or convicted of at least one offense from one of the 12 main crime categories were counted. Before, during, and after refer to when the incident associated with the record occurred, not when the arrest date, charge date, adjudication date, or conviction date occurred relative to the testing window.

Table 14: Offender-Level Crime Category Counts (Overall and By When Incident Occurred)

```
# Table caption.
TCap <- paste("Frequency Distributions for Number of Crime Categories for",
              "Which Offender Was Arrested, Charged, or Convicted (Overall",
              "and By When Incident Occurred)")
# Footnote text.
FN <- paste0("N = ", format(N_All, big.mark = ","), ". TW, earliest ",
            "testing window. ",
            "Only records associated with incidents where the offender was ",
            "arrested for, charged with, or convicted of at least one offense ",
            "from one of the 12 main crime categories were counted. ",
            "Before, during, and after refer to when the incident associated ",
            "with the record occurred, not when the charge date occurred ",
            "relative to the testing window. For example, an offender could be ",
            "convicted during the testing window for an incident that occurred ",
            "before it began. That would show up here in the before column.")

# Get frequency distributions.
IDNEW %>%
  select(OID, HXCat12_Count, HXCat12_Count_Before, HXCat12_Count_During,
         HXCat12_Count_After) %>%
  rename(All = HXCat12_Count, Before = HXCat12_Count_Before,
         During = HXCat12_Count_During, After = HXCat12_Count_After) %>%
  pivot_longer(cols = -OID, names_to = "When", values_to = "Value") %>%
  group_by(When, Value) %>%
```

```
count() %>%
arrange(Value) %>%
pivot_wider(names_from = When, values_from = n, values_fill = 0) %>%
mutate(All_p = 100*All/N_All,
       All_v = if_else(is.na(Value),
                       true = as.numeric(NA),
                       false = 100*All/N_All),
       Before_p = 100*Before/N_All,
       Before_v = if_else(is.na(Value),
                          true = as.numeric(NA),
                          false = 100*Before/N_Before),
       During_p = 100*During/N_All,
       During_v = if_else(is.na(Value),
                          true = as.numeric(NA),
                          false = 100*During/N_During),
       After_p = 100*After/N_All,
       After_v = if_else(is.na(Value),
                         true = as.numeric(NA),
                         false = 100*After/N_After)) %>%
select(Value, All, All_p, All_v, Before, Before_p, Before_v, During,
       During_p, During_v, After, After_p, After_v) %>%
kable(format = "latex", booktabs = TRUE, digits = 2,
      col.names = c("Value", rep(c("N", "%", "Valid %"), 4)),
      caption = TCap) %>%
kable_styling() %>%
add_header_above(c(" " = 1, "Overall" = 3, "Before TW" = 3, "During TW" = 3,
                   "After TW" = 3)) %>%
footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
       threeparttable = TRUE)
```

	Overall			Before TW			During TW			After TW		
Value	N	%	Valid %	N	%	Valid %	N	%	Valid %	N	%	Valid %
0	0	0.00	0.00	276	25.51	26.51	714	65.99	67.87	90	8.32	8.32
1	41	3.79	3.79	215	19.87	20.65	211	19.50	20.06	116	10.72	10.72
2	97	8.96	8.96	199	18.39	19.12	92	8.50	8.75	191	17.65	17.65
3	183	16.91	16.91	164	15.16	15.75	22	2.03	2.09	210	19.41	19.41
4	229	21.16	21.16	98	9.06	9.41	10	0.92	0.95	214	19.78	19.78
5	221	20.43	20.43	53	4.90	5.09	2	0.18	0.19	138	12.75	12.75
6	162	14.97	14.97	22	2.03	2.11	0	0.00	0.00	67	6.19	6.19
7	94	8.69	8.69	12	1.11	1.15	1	0.09	0.10	40	3.70	3.70
8	41	3.79	3.79	2	0.18	0.19	0	0.00	0.00	14	1.29	1.29
9	13	1.20	1.20	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
10	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
NA	0	0.00	NA	41	3.79	NA	30	2.77	NA	0	0.00	NA

Note: N = 1,082. TW, earliest testing window. Only records associated with incidents where the offender was arrested for, charged with, or convicted of at least one offense from one of the 12 main crime categories were counted. Before, during, and after refer to when the incident associated with the record occurred, not when the charge date occurred relative to the testing window. For example, an offender could be convicted during the testing window for an incident that occurred before it began. That would show up here in the before column.

Table 15: Frequency Distributions for Number of Crime Categories for Which Offender Was Arrested, Charged, or Convicted (Overall and By When Incident Occurred)

5.6.1 GEE Models

To more closely examine the crime category counts, we fitted two generalized estimating equation (GEE) models to the long-format *IDNEWL* data, after dropping rows for unobserved periods ($Years = 0$) because the counts are missing for those observations.

Model 1 used *When* as a predictor of the period-specific counts, but did not adjust for the period durations. Model 2 supplemented *When* as a predictor with an offset term for log(*Years*) to adjust for the period durations. Both models used Poisson distributions, an exchangeable correlation structure, a log link function, and robust standard errors based on the sandwich estimator.

```
cccvar_gee1 <- IDNEWL %>%
  # Keep only observations for observed periods (i.e., those with Years > 0).
  filter(Years > 0) %>%
  geeglm(HXCat12_Count ~ When,
         family = poisson(link = "log"), data = .,
         id = OID, corstr = "exchangeable", std.err = "san.se")

cccvar_gee2 <- IDNEWL %>%
  filter(Years > 0) %>%
  geeglm(HXCat12_Count ~ When + offset(log(Years)),
         family = poisson(link = "log"), data = .,
         id = OID, corstr = "exchangeable", std.err = "san.se")
```

Table 16 below shows the parameter estimates for both GEE models.

```
TCap <- paste("Generalized Estimating Equation Models for Crime Category Counts",
              "Predicted by When Incidents Occurred")
FN <- paste ("\\item * p < .05, null hypothesis value outside 95\\% confidence",
             "interval based on z-score.",
             "\\\\\n\\item Note.",
             "The data contain up to 3 longitudinal observations (one per",
             "period) for each offender (i.e., cluster).",
             "The reference level for When was the period after the testing",
             "window.",
             "Both models used a log link function, exchangeable correlation",
             "structure, a Poisson distribution, and robust standard errors",
             "(sandwich estimator).",
             "Model 1 did not use an offset.",
             "Model 2 used log(Years) as an offset term.")

texreg(list(cccvar_gee1, cccvar_gee2), booktabs = TRUE, dcolumn = TRUE,
       threeparttable = TRUE, fontsize = "normalsize", table = TRUE,
       use.packages = FALSE, ci.force = TRUE, label = "tab:cccvar_texreg",
       stars = 0.05, caption = TCap, custom.note = FN)
```

	Model 1	Model 2
(Intercept)	1.17*	−1.72*
	[1.14; 1.21]	[−1.76; −1.69]
WhenBefore	−0.52*	−0.10*
	[−0.59; −0.45]	[−0.15; −0.04]
WhenDuring	−1.88*	2.13*
	[−2.00; −1.76]	[2.02; 2.25]
Scale parameter: gamma	1.38	1.48
Scale parameter: SE	0.04	0.05
Correlation parameter: alpha	−0.07	0.03
Correlation parameter: SE	0.02	0.02
Num. obs.	3175	3175
Num. clust.	1082	1082

* p < .05, null hypothesis value outside 95% confidence interval based on z-score.

Note. The data contain up to 3 longitudinal observations (one per period) for each offender (i.e., cluster). The reference level for When was the period after the testing window. Both models used a log link function, exchangeable correlation structure, a Poisson distribution, and robust standard errors (sandwich estimator). Model 1 did not use an offset. Model 2 used log(Years) as an offset term.

Table 16: Generalized Estimating Equation Models for Crime Category Counts Predicted by When Incidents Occurred

5.6.2 Estimated Marginal Means

Table 17 provides the estimated marginal means obtained from Models 1 and 2, along with corresponding confidence intervals. The units of measurement for those means are counts per offender for Model 1 (which ignore period duration) and counts per offender-year for Model 2 (which adjust for duration and are thus considered annual incidence rates).

```
# Table caption.
TCap <- paste("Marginal Means of Crime Category Counts and Incidence Rates",
              "By When Incidents Occurred")

# Table footnote.
FN <- paste("Model 1 did not use an offset, so means are counts.",
            "Model 2 used log(Years) as an offset term, so means are annual",
            "incidence rates.",
            "Confidence intervals are based on z-scores.")

cccvar_emmeans1 <- emmeans(cccvar_gee1, specs = pairwise ~ When,
                           type = "response", ref = "During")
cccvar_emmeans2 <- emmeans(cccvar_gee2, specs = pairwise ~ When,
                           type = "response", ref = "During")

as_tibble(cccvar_emmeans1$emmeans) %>%
  full_join(x = ., y = as_tibble(cccvar_emmeans2$emmeans)) %>%
  mutate(Model = c(1, 1, 1, 2, 2, 2),
         When = factor(When, levels = c("Before", "During", "After"))) %>%
  arrange(Model, When) %>%
  select(When, rate, SE, df, asymp.LCL, asymp.UCL) %>%
  kable(format = "latex", booktabs = TRUE, caption = TCap, digits = 2,
        col.names = c("When", "Rate", "SE", "df", "LCL", "UCL")) %>%
  kable_styling() %>%
  add_header_above(c(" " = 4, "95% CI" = 2)) %>%
  group_rows(group_label = "Model 1: Mean per offender", start_row = 1,
             end_row = 3, italic = TRUE) %>%
  group_rows(group_label = "Model 2: Mean per offender-year",
             start_row = 4, end_row = 6, italic = TRUE) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
## Joining, by = c("When", "rate", "SE", "df", "asymp.LCL", "asymp.UCL")
```

				95% CI	
When	Rate	SE	df	LCL	UCL
Model 1: Mean per offender					
Before	1.91	0.05	Inf	1.81	2.02
During	0.49	0.03	Inf	0.44	0.55
After	3.22	0.06	Inf	3.11	3.34
Model 2: Mean per offender-year					
Before	1.64	0.04	Inf	1.56	1.73
During	15.28	0.83	Inf	13.74	16.99
After	1.81	0.03	Inf	1.74	1.88

Note:
Model 1 did not use an offset, so means are counts. Model 2 used log(Years) as an offset term, so means are annual incidence rates. Confidence intervals are based on z-scores.

Table 17: Marginal Means of Crime Category Counts and Incidence Rates By When Incidents Occurred

5.6.3 Contrasts for Ratios of Estimated Marginal Means

Finally, we also estimated a set of contrasts to compare the marginal means from Models 1 and 2 (Table 18). Each of these contrasts estimates the ratios of a pair of means rather than a raw difference between means. This is because of the nature of the underlying Poisson GEE model.

```
TCap <- paste("Contrasts Estimating Ratios of Marginal Means for Crime",
              "Category Counts and Incidence Rates By When Incidents",
              "Occurred")

# Table footnote.
FN <- paste("Model 1 did not use an offset, so means are counts.",
            "Model 2 used log(Years) as an offset term, so means are annual",
            "incidence rates.",
            "These contrasts estimate ratios of those quantities and used",
            "Tukey's method to adjust for multiple comparisons.",
            "Confidence intervals are based on z-scores.")

as_tibble(confint(cccvar_emmeans1$contrasts)) %>%
  full_join(x = ., y = as_tibble(confint(cccvar_emmeans2$contrasts))) %>%
  kable(format = "latex", booktabs = TRUE, caption = TCap, digits = 2,
        col.names = c("Contrast", "Ratio", "SE", "df", "LCL", "UCL")) %>%
  kable_styling() %>%
  add_header_above(c(" " = 4, "95% CI" = 2)) %>%
  group_rows(group_label = "Model 1: Ratios of means per offender",
             start_row = 1, end_row = 3, italic = TRUE) %>%
  group_rows(group_label = "Model 2: Ratios of means per offender-year",
             start_row = 4, end_row = 6, italic = TRUE) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
## Joining, by = c("contrast", "ratio", "SE", "df", "asymp.LCL", "asymp.UCL")
```

Contrast	Ratio	SE	df	95% CI LCL	95% CI UCL
Model 1: Ratios of means per offender					
After / Before	1.68	0.06	Inf	1.56	1.82
After / During	6.55	0.39	Inf	5.69	7.54
Before / During	3.89	0.24	Inf	3.38	4.48
Model 2: Ratios of means per offender-year					
After / Before	1.10	0.03	Inf	1.03	1.18
After / During	0.12	0.01	Inf	0.10	0.14
Before / During	0.11	0.01	Inf	0.09	0.12

Note:
Model 1 did not use an offset, so means are counts. Model 2 used log(Years) as an offset term, so means are annual incidence rates. These contrasts estimate ratios of those quantities and used Tukey's method to adjust for multiple comparisons. Confidence intervals are based on z-scores.

Table 18: Contrasts Estimating Ratios of Marginal Means for Crime Category Counts and Incidence Rates By When Incidents Occurred

5.7 Incident Counts by Crime Category (12 levels) and Period

In this section, we analyze variables that contain the number of unique incidents in the offender's criminal history for a given type of crime. For each category of crime, we report include counts of incidents based on a combined criterion where the offender was *arrested for, charged with, or convicted of* (any of the three events, or any combination of them) crimes classified into the relevant category.

Note that these variables were created by first aggregating *ARR*, *CHG*, and *JUD* records to the incident level to flag incidents that fit the relevant category, then aggregated again to get to the offender level incident counts. It is possible to have multiple arrest offense records, charge records, or adjudication records for convictions on a single incident, but these variables ignore that and only count each incident once.

It may be helpful to note that the order of subsections below follows the numerical coding of the categories (which was influenced by alphabetical order of the labels we chose). We grouped and presented results in a different order in the manuscript that we think makes the material easier to present to readers. So, in comparing these materials to the published paper, pay attention to the crime category labels and section headings.

The chunk below creates some objects we use to dynamically construct text later in the subsections below.

```
# Save a formatted value for total number of offenders.
N_IDNEW <- format(N_All, big.mark = ",")

# Check how many offenders had >= 1 CSC arrest, charge, or conviction in
# their CHR data for each crime category.
IDNEW %>% filter(HXCat12_1  >= 1) %>% nrow()       -> N_HXCat12_1
IDNEW %>% filter(HXCat12_2  >= 1) %>% nrow()       -> N_HXCat12_2
IDNEW %>% filter(HXCat12_3  >= 1) %>% nrow()       -> N_HXCat12_3
IDNEW %>% filter(HXCat12_4  >= 1) %>% nrow()       -> N_HXCat12_4
IDNEW %>% filter(HXCat12_5  >= 1) %>% nrow()       -> N_HXCat12_5
IDNEW %>% filter(HXCat12_6  >= 1) %>% nrow()       -> N_HXCat12_6
IDNEW %>% filter(HXCat12_7  >= 1) %>% nrow()       -> N_HXCat12_7
IDNEW %>% filter(HXCat12_8  >= 1) %>% nrow()       -> N_HXCat12_8
IDNEW %>% filter(HXCat12_9  >= 1) %>% nrow()       -> N_HXCat12_9
IDNEW %>% filter(HXCat12_10 >= 1) %>% nrow()       -> N_HXCat12_10
IDNEW %>% filter(HXCat12_11 >= 1) %>% nrow()       -> N_HXCat12_11
IDNEW %>% filter(HXCat12_12 >= 1) %>% nrow()       -> N_HXCat12_12
IDNEW %>% filter(HXCat12_Sexual >= 1) %>% nrow()   -> N_HXCat12_Sexual
IDNEW %>% filter(HXCat12_Violent >= 1) %>% nrow()  -> N_HXCat12_Violent
IDNEW %>% filter(HXCat12_Property >= 1) %>% nrow() -> N_HXCat12_Property
IDNEW %>% filter(HXCat12_Other >= 1) %>% nrow()    -> N_HXCat12_Other
```

```
# Convert those new variables to percentages.
P_HXCat12_1        <- round(100*N_HXCat12_1/N_All, digits = 0)
P_HXCat12_2        <- round(100*N_HXCat12_2/N_All, digits = 0)
P_HXCat12_3        <- round(100*N_HXCat12_3/N_All, digits = 0)
P_HXCat12_4        <- round(100*N_HXCat12_4/N_All, digits = 0)
P_HXCat12_5        <- round(100*N_HXCat12_5/N_All, digits = 0)
P_HXCat12_6        <- round(100*N_HXCat12_6/N_All, digits = 0)
P_HXCat12_7        <- round(100*N_HXCat12_7/N_All, digits = 0)
P_HXCat12_8        <- round(100*N_HXCat12_8/N_All, digits = 0)
P_HXCat12_9        <- round(100*N_HXCat12_9/N_All, digits = 0)
P_HXCat12_10       <- round(100*N_HXCat12_10/N_All, digits = 0)
P_HXCat12_11       <- round(100*N_HXCat12_11/N_All, digits = 0)
P_HXCat12_12       <- round(100*N_HXCat12_12/N_All, digits = 0)
P_HXCat12_Sexual   <- round(100*N_HXCat12_Sexual/N_All, digits = 0)
P_HXCat12_Violent  <- round(100*N_HXCat12_Violent/N_All, digits = 0)
P_HXCat12_Property <- round(100*N_HXCat12_Property/N_All, digits = 0)
P_HXCat12_Other    <- round(100*N_HXCat12_Other/N_All, digits = 0)
```

This next chunk creates some objects to store things like standardized footnotes that we will re-use.

```
# Footnote text: Descriptives tables.
FN1 <- paste("Only offenders with valid testing window start dates are",
             "included.",
             "Only incidents where the offender was arrested for, charged with,",
             "or convicted of at least one offense from one of the 12 main crime",
             "categories were counted.",
             "Before, during, and after refer to when the incident",
             "associated with the record occurred, not when the arrest date,",
             "charge date, or conviction date occurred relative to the testing",
             "window.")

# Footnote text.
FN5 <- paste0("N = ", N_IDNEW, ". TW, earliest testing window. ",
              "Only offenders with valid testing window start dates are ",
              "included. Before, during, and after refer to when the incident ",
              "associated with the record occurred, not when the arrest, ",
              "charge, or conviction date occurred relative to the testing ",
              "window. For example, an offender could be arrested during the ",
              "testing window for an incident that occurred before it began. ",
              "That would show up here in the before column. ",
              "An incident is counted if there are any arrest, charge, or ",
              "conviction records (any one type, or any combination of them ",
              "will suffice) for the specified crime category associated with ",
              "it.")

# Vector of count types
counttypes <- c("Arrested", "Charged", "Convicted",
                "Arrested, Charged, or Convicted")

# Vector of row labels for descriptives tables.
TRowLabs <- c("Incident count (arrested, charged, or convicted)", BDA)
```

5.7.1 Arson

Among these 1,082 suspected sexual offenders, there were 35 (3%) offenders who had criminal histories containing at least one incident associated with an arrest, charge, or conviction for arson. Table 19 summarizes the counts of unique arson incidents overall and by when the incident occurred, while Table 20 shows the frequency distributions underlying those summaries.

```
# Table caption.
TCap <- paste("Offender-Level Arson Incident Counts (Overall and By When",
              "Incident Occurred)")

# Summarize crime category counts (both overall and broken down by IWhen).
IDNEW %>%
  select(HXCat12_1, HXCat12_1_Before, HXCat12_1_During, HXCat12_1_After) %>%
  describe(., quant=c(.25, .50, .75)) %>%
  bind_cols(data.frame(Variable = TRowLabs), .) %>%
```

```
  select(Variable, n, mean, sd, min, max, skew, kurtosis, Q0.25, Q0.5, Q0.75) %>%
  kable(., format = "latex", booktabs = TRUE, digits = 2, row.names = FALSE,
        col.names = CNames, caption = TCap,
        linesep = c('', '', '', '\\addlinespace')) %>%
  column_spec(column = 1, width = "6 cm") %>%
  footnote(general = FN1, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

Variable	N	Mean	SD	Min	Max	Skew	Kurtosis	Q25	Q50	Q75
Incident count (arrested, charged, or convicted)	1082	0.03	0.19	0	2	5.89	37.07	0	0	0
...Before testing window	1041	0.01	0.12	0	1	8.14	64.29	0	0	0
...During testing window	1052	0.00	0.06	0	2	32.34	1045.01	0	0	0
...After testing window	1082	0.02	0.14	0	2	8.10	71.15	0	0	0

Note: Only offenders with valid testing window start dates are included. Only incidents where the offender was arrested for, charged with, or convicted of at least one offense from one of the 12 main crime categories were counted. Before, during, and after refer to when the incident associated with the record occurred, not when the arrest date, charge date, or conviction date occurred relative to the testing window.

Table 19: Offender-Level Arson Incident Counts (Overall and By When Incident Occurred)

```
# Table caption.
TCap <- paste("Frequency Distributions for Number of Arson Incidents for Which",
              "Offender Was Arrested, Charged, or Convicted (Overall and By",
              "When Incident Occurred)")

# Get frequency distributions.
IDNEW %>%
  select(OID, HXCat12_1, HXCat12_1_Before, HXCat12_1_During, HXCat12_1_After) %>%
  rename(All = HXCat12_1, Before = HXCat12_1_Before, During = HXCat12_1_During,
         After = HXCat12_1_After) %>%
  pivot_longer(cols = -OID, names_to = "When", values_to = "Value") %>%
  group_by(When, Value) %>%
  count() %>%
  arrange(Value) %>%
  pivot_wider(names_from = When, values_from = n, values_fill = 0) %>%
  mutate(All_p = 100*All/N_All,
         All_v = if_else(is.na(Value),
                         true = as.numeric(NA),
                         false = 100*All/N_All),
         Before_p = 100*Before/N_All,
         Before_v = if_else(is.na(Value),
                            true = as.numeric(NA),
                            false = 100*Before/N_Before),
         During_p = 100*During/N_All,
         During_v = if_else(is.na(Value),
                            true = as.numeric(NA),
                            false = 100*During/N_During),
         After_p = 100*After/N_All,
         After_v = if_else(is.na(Value),
                           true = as.numeric(NA),
                           false = 100*After/N_After)) %>%
  select(Value, All, All_p, All_v, Before, Before_p, Before_v, During,
         During_p, During_v, After, After_p, After_v) %>%
  kable(format = "latex", booktabs = TRUE, digits = 2,
        col.names = c("Value", rep(c("N", "%", "Valid %"), 4)),
        caption = TCap) %>%
  kable_styling() %>%
  add_header_above(c(" " = 1, "Overall" = 3, "Before TW" = 3, "During TW" = 3,
                     "After TW" = 3)) %>%
  footnote(general = FN5, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

	Overall			Before TW			During TW			After TW		
Value	N	%	Valid %	N	%	Valid %	N	%	Valid %	N	%	Valid %
0	1047	96.77	96.77	1026	94.82	98.56	1051	97.13	99.9	1063	98.24	98.24
1	33	3.05	3.05	15	1.39	1.44	0	0.00	0.0	18	1.66	1.66
2	2	0.18	0.18	0	0.00	0.00	1	0.09	0.1	1	0.09	0.09
NA	0	0.00	NA	41	3.79	NA	30	2.77	NA	0	0.00	NA

Note: N = 1,082. TW, earliest testing window. Only offenders with valid testing window start dates are included. Before, during, and after refer to when the incident associated with the record occurred, not when the arrest, charge, or conviction date occurred relative to the testing window. For example, an offender could be arrested during the testing window for an incident that occurred before it began. That would show up here in the before column. An incident is counted if there are any arrest, charge, or conviction records (any one type, or any combination of them will suffice) for the specified crime category associated with it.

Table 20: Frequency Distributions for Number of Arson Incidents for Which Offender Was Arrested, Charged, or Convicted (Overall and By When Incident Occurred)

5.7.1.1 GEE Models To more closely examine the arson incident counts, we fitted two generalized estimating equation (GEE) models to the long-format *IDNEWL* data, after dropping rows for unobserved periods (*Years* = 0) because the counts are missing for those observations.

Model 1 used *When* as a predictor of the period-specific counts, but did not adjust for the period durations. Model 2 supplemented *When* as a predictor with an offset term for log(*Years*) to adjust for the period durations. Both models used Poisson distributions, an exchangeable correlation structure, a log link function, and robust standard errors based on the sandwich estimator.

```
HXCat12_1_gee1 <- IDNEWL %>%
  filter(Years > 0) %>%
  geeglm(HXCat12_1 ~ When,
         family = poisson(link = "log"), data = .,
         id = OID, corstr = "exchangeable", std.err = "san.se")

HXCat12_1_gee2 <- IDNEWL %>%
  filter(Years > 0) %>%
  geeglm(HXCat12_1 ~ When + offset(log(Years)),
         family = poisson(link = "log"), data = .,
         id = OID, corstr = "exchangeable", std.err = "san.se")
```

Table 21 below shows the parameter estimates for both GEE models.

```
TCap <- paste("Generalized Estimating Equation Models for Arson Incident Counts",
              "Predicted by When Incidents Occurred")
FN <- paste ("\\item * p < .05, null hypothesis value outside 95\\% confidence",
             "interval based on z-score.",
             "\\\\\n\\item Note.",
             "The data contain up to 3 longitudinal observations (one per",
             "period) for each offender (i.e., cluster).",
             "The reference priod for When was after the testing window.",
             "Both models used a log link function, exchangeable correlation",
             "structure, a Poisson distribution, and robust standard errors",
             "(sandwich estimator).",
             "Model 1 did not use an offset.",
             "Model 2 used log(Years) as an offset term.")

texreg(list(HXCat12_1_gee1, HXCat12_1_gee2), booktabs = TRUE, dcolumn = TRUE,
       threeparttable = TRUE, fontsize = "normalsize", table = TRUE,
       use.packages = FALSE, ci.force = TRUE, label = "tab:HXCat12_1_texreg",
       stars = 0.05, caption = TCap, custom.note = FN)
```

	Model 1	Model 2
(Intercept)	−3.99*	−6.88*
	[−4.45; −3.54]	[−7.34; −6.43]
WhenBefore	−0.25	0.17
	[−0.93; 0.43]	[−0.51; 0.85]
WhenDuring	−2.27*	1.74
	[−4.29; −0.26]	[−0.27; 3.75]
Scale parameter: gamma	1.35	1.35
Scale parameter: SE	14.51	14.38
Correlation parameter: alpha	−0.01	−0.01
Correlation parameter: SE	0.07	0.06
Num. obs.	3175	3175
Num. clust.	1082	1082

* p < .05, null hypothesis value outside 95% confidence interval based on z-score.

Note. The data contain up to 3 longitudinal observations (one per period) for each offender (i.e., cluster). The reference priod for When was after the testing window. Both models used a log link function, exchangeable correlation structure, a Poisson distribution, and robust standard errors (sandwich estimator). Model 1 did not use an offset. Model 2 used log(Years) as an offset term.

Table 21: Generalized Estimating Equation Models for Arson Incident Counts Predicted by When Incidents Occurred

5.7.1.2 Estimated Marginal Means Table 22 provides the estimated marginal means obtained from Models 1 and 2, along with corresponding confidence intervals. The units of measurement for those means are counts per offender for Model 1 (which ignore period duration) and counts per offender-year for Model 2 (which adjust for duration and are thus considered annual incidence rates).

```
# Table caption.
TCap <- paste("Marginal Means of Arson Incident Counts and Incidence Rates",
              "By When Incidents Occurred")

# Table footnote.
FN <- paste("Model 1 did not use an offset, so means are counts.",
            "Model 2 used log(Years) as an offset term, so means are annual",
            "incidence rates.",
            "Confidence intervals are based on z-scores.")

HXCat12_1_emmeans1 <- emmeans(HXCat12_1_gee1, specs = pairwise ~ When,
                              type = "response", ref = "During")
HXCat12_1_emmeans2 <- emmeans(HXCat12_1_gee2, specs = pairwise ~ When,
                              type = "response", ref = "During")

as_tibble(HXCat12_1_emmeans1$emmeans) %>%
  full_join(x = ., y = as_tibble(HXCat12_1_emmeans2$emmeans)) %>%
  mutate(Model = c(1, 1, 1, 2, 2, 2),
         When = factor(When, levels = c("Before", "During", "After"))) %>%
  arrange(Model, When) %>%
  select(When, rate, SE, df, asymp.LCL, asymp.UCL) %>%
  kable(format = "latex", booktabs = TRUE, caption = TCap, digits = 2,
        col.names = c("When", "Rate", "SE", "df", "LCL", "UCL")) %>%
  kable_styling() %>%
  add_header_above(c(" " = 4, "95% CI" = 2)) %>%
  group_rows(group_label = "Model 1: Mean per offender", start_row = 1,
             end_row = 3, italic = TRUE) %>%
  group_rows(group_label = "Model 2: Mean per offender-year",
             start_row = 4, end_row = 6, italic = TRUE) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
## Joining, by = c("When", "rate", "SE", "df", "asymp.LCL", "asymp.UCL")
```

				95% CI	
When	Rate	SE	df	LCL	UCL
Model 1: Mean per offender					
Before	0.01	0.00	Inf	0.01	0.02
During	0.00	0.00	Inf	0.00	0.01
After	0.02	0.00	Inf	0.01	0.03
Model 2: Mean per offender-year					
Before	0.01	0.00	Inf	0.01	0.02
During	0.06	0.06	Inf	0.01	0.42
After	0.01	0.00	Inf	0.01	0.02

Note: Model 1 did not use an offset, so means are counts. Model 2 used log(Years) as an offset term, so means are annual incidence rates. Confidence intervals are based on z-scores.

Table 22: Marginal Means of Arson Incident Counts and Incidence Rates By When Incidents Occurred

5.7.1.3 Contrasts for Ratios of Estimated Marginal Means Finally, we also estimated a set of contrasts to compare the marginal means from Models 1 and 2 (Table 23). Each of these contrasts estimates the ratios of a pair of means rather than a raw difference between means. This is because of the nature of the underlying Poisson GEE model.

```
TCap <- paste("Contrasts Estimating Ratios of Marginal Means for Arson",
              "Incident Counts and Incidence Rates By When Incidents",
              "Occurred")

# Table footnote.
FN <- paste("Model 1 did not use an offset, so means are counts.",
            "Model 2 used log(Years) as an offset term, so means are annual",
            "incidence rates.",
            "These contrasts estimate ratios of those quantities and used",
            "Tukey's method to adjust for multiple comparisons.",
            "Confidence intervals are based on z-scores.")

as_tibble(confint(HXCat12_1_emmeans1$contrasts)) %>%
  full_join(x = ., y = as_tibble(confint(HXCat12_1_emmeans2$contrasts))) %>%
  kable(format = "latex", booktabs = TRUE, caption = TCap, digits = 2,
        col.names = c("Contrast", "Ratio", "SE", "df", "LCL", "UCL")) %>%
  kable_styling() %>%
  add_header_above(c(" " = 4, "95% CI" = 2)) %>%
  group_rows(group_label = "Model 1: Ratios of means per offender",
             start_row = 1, end_row = 3, italic = TRUE) %>%
  group_rows(group_label = "Model 2: Ratios of means per offender-year",
             start_row = 4, end_row = 6, italic = TRUE) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
## Joining, by = c("contrast", "ratio", "SE", "df", "asymp.LCL", "asymp.UCL")
```

Contrast	Ratio	SE	df	95% CI LCL	95% CI UCL
Model 1: Ratios of means per offender					
After / Before	1.28	0.45	Inf	0.57	2.90
After / During	9.72	9.97	Inf	0.88	107.74
Before / During	7.58	7.82	Inf	0.67	85.13
Model 2: Ratios of means per offender-year					
After / Before	0.84	0.29	Inf	0.37	1.90
After / During	0.17	0.18	Inf	0.02	1.93
Before / During	0.21	0.21	Inf	0.02	2.33

Note: Model 1 did not use an offset, so means are counts. Model 2 used log(Years) as an offset term, so means are annual incidence rates. These contrasts estimate ratios of those quantities and used Tukey's method to adjust for multiple comparisons. Confidence intervals are based on z-scores.

Table 23: Contrasts Estimating Ratios of Marginal Means for Arson Incident Counts and Incidence Rates By When Incidents Occurred

5.7.2 Assault - DV, Stalking

Among these 1,082 suspected sexual offenders, there were 288 (27%) offenders who had criminal histories containing at least one incident associated with an arrest, charge, or conviction for assault involving domestic violence and/or stalking. Table 24 summarizes the counts of unique assault incidents overall and by when the incident occurred, while Table 25 shows the frequency distributions underlying those summaries.

```
# Table caption.
TCap <- paste("Offender-Level Assault Involving Domestc Violence and/or",
              "Stalking Incident Counts (Overall and By When Incident Occurred)")

# Summarize crime category counts (both overall and broken down by IWhen).
IDNEW %>%
  select(HXCat12_2, HXCat12_2_Before, HXCat12_2_During, HXCat12_2_After) %>%
  describe(., quant=c(.25, .50, .75)) %>%
  bind_cols(data.frame(Variable = TRowLabs), .) %>%
  select(Variable, n, mean, sd, min, max, skew, kurtosis, Q0.25, Q0.5, Q0.75) %>%
  kable(., format = "latex", booktabs = TRUE, digits = 2, row.names = FALSE,
        col.names = CNames, caption = TCap,
        linesep = c('', '', '', '\\addlinespace')) %>%
  column_spec(column = 1, width = "6 cm") %>%
  footnote(general = FN1, general_title = "Note: ", footnote_as_chunk = TRUE,
           threeparttable = TRUE)
```

Variable	N	Mean	SD	Min	Max	Skew	Kurtosis	Q25	Q50	Q75
Incident count (arrested, charged, or convicted)	1082	0.42	0.85	0	6	2.54	7.50	0	0	1
...Before testing window	1041	0.05	0.23	0	3	5.75	42.36	0	0	0
...During testing window	1052	0.02	0.15	0	2	8.65	82.86	0	0	0
...After testing window	1082	0.36	0.79	0	6	2.80	9.14	0	0	0

Note: Only offenders with valid testing window start dates are included. Only incidents where the offender was arrested for, charged with, or convicted of at least one offense from one of the 12 main crime categories were counted. Before, during, and after refer to when the incident associated with the record occurred, not when the arrest date, charge date, or conviction date occurred relative to the testing window.

Table 24: Offender-Level Assault Involving Domestc Violence and/or Stalking Incident Counts (Overall and By When Incident Occurred)

```
# Table caption.
TCap <- paste("Frequency Distributions for Number of Assault Involving",
              "Domestic Violence and/or Stalking Incidents for Which",
              "Offender Was Arrested, Charged, or Convicted (Overall and By",
              "When Incident Occurred)")

# Get frequency distributions.
IDNEW %>%
  select(OID, HXCat12_2, HXCat12_2_Before, HXCat12_2_During, HXCat12_2_After) %>%
  rename(All = HXCat12_2, Before = HXCat12_2_Before, During = HXCat12_2_During,
         After = HXCat12_2_After) %>%
  pivot_longer(cols = -OID, names_to = "When", values_to = "Value") %>%
  group_by(When, Value) %>%
  count() %>%
  arrange(Value) %>%
  pivot_wider(names_from = When, values_from = n, values_fill = 0) %>%
  mutate(All_p = 100*All/N_All,
         All_v = if_else(is.na(Value),
                         true = as.numeric(NA),
                         false = 100*All/N_All),
         Before_p = 100*Before/N_All,
         Before_v = if_else(is.na(Value),
                            true = as.numeric(NA),
                            false = 100*Before/N_Before),
         During_p = 100*During/N_All,
         During_v = if_else(is.na(Value),
                            true = as.numeric(NA),
                            false = 100*During/N_During),
         After_p = 100*After/N_All,
         After_v = if_else(is.na(Value),
                           true = as.numeric(NA),
                           false = 100*After/N_After)) %>%
  select(Value, All, All_p, All_v, Before, Before_p, Before_v, During,
         During_p, During_v, After, After_p, After_v) %>%
  kable(format = "latex", booktabs = TRUE, digits = 2,
        col.names = c("Value", rep(c("N", "%", "Valid %"), 4)),
        caption = TCap) %>%
  kable_styling() %>%
  add_header_above(c(" " = 1, "Overall" = 3, "Before TW" = 3, "During TW" = 3,
                     "After TW" = 3)) %>%
  footnote(general = FN5, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

	Overall			Before TW			During TW			After TW		
Value	N	%	Valid %	N	%	Valid %	N	%	Valid %	N	%	Valid %
0	794	73.38	73.38	996	92.05	95.68	1034	95.56	98.29	836	77.26	77.26
1	182	16.82	16.82	43	3.97	4.13	16	1.48	1.52	156	14.42	14.42
2	64	5.91	5.91	1	0.09	0.10	2	0.18	0.19	55	5.08	5.08
3	28	2.59	2.59	1	0.09	0.10	0	0.00	0.00	22	2.03	2.03
4	9	0.83	0.83	0	0.00	0.00	0	0.00	0.00	10	0.92	0.92
5	4	0.37	0.37	0	0.00	0.00	0	0.00	0.00	2	0.18	0.18
6	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
NA	0	0.00	NA	41	3.79	NA	30	2.77	NA	0	0.00	NA

Note: N = 1,082. TW, earliest testing window. Only offenders with valid testing window start dates are included. Before, during, and after refer to when the incident associated with the record occurred, not when the arrest, charge, or conviction date occurred relative to the testing window. For example, an offender could be arrested during the testing window for an incident that occurred before it began. That would show up here in the before column. An incident is counted if there are any arrest, charge, or conviction records (any one type, or any combination of them will suffice) for the specified crime category associated with it.

Table 25: Frequency Distributions for Number of Assault Involving Domestic Violence and/or Stalking Incidents for Which Offender Was Arrested, Charged, or Convicted (Overall and By When Incident Occurred)

5.7.2.1 GEE Models To more closely examine the assault involving domestic violence and/or stalking incident counts, we fitted two generalized estimating equation (GEE) models to the long-format *IDNEWL* data, after dropping rows for unobserved periods ($Years = 0$) because the counts are missing for those observations.

Model 1 used *When* as a predictor of the period-specific counts, but did not adjust for the period durations. Model 2 supplemented *When* as a predictor with an offset term for log(*Years*) to adjust for the period durations. Both models used Poisson distributions, an exchangeable correlation structure, a log link function, and robust standard errors based on the sandwich estimator.

```
HXCat12_2_gee1 <- IDNEWL %>%
  filter(Years > 0) %>%
  geeglm(HXCat12_2 ~ When,
         family = poisson(link = "log"), data = .,
         id = OID, corstr = "exchangeable", std.err = "san.se")

HXCat12_2_gee2 <- IDNEWL %>%
  filter(Years > 0) %>%
  geeglm(HXCat12_2 ~ When + offset(log(Years)),
         family = poisson(link = "log"), data = .,
         id = OID, corstr = "exchangeable", std.err = "san.se")
```

Table 26 below shows the parameter estimates for both GEE models.

```
TCap <- paste("Generalized Estimating Equation Models for Assault Involving",
              "Domestic Violence and/or Stalking Incident Counts",
              "Predicted by When Incidents Occurred")
FN <- paste ("\\item * p < .05, null hypothesis value outside 95\\% confidence",
             "interval based on z-score.",
             "\\\\\n\\item Note.",
             "The data contain up to 3 longitudinal observations (one per",
             "period) for each offender (i.e., cluster).",
             "The reference priod for When was after the testing window.",
             "Both models used a log link function, exchangeable correlation",
             "structure, a Poisson distribution, and robust standard errors",
             "(sandwich estimator).",
             "Model 1 did not use an offset.",
             "Model 2 used log(Years) as an offset term.")

texreg(list(HXCat12_2_gee1, HXCat12_2_gee2), booktabs = TRUE, dcolumn = TRUE,
       threeparttable = TRUE, fontsize = "normalsize", table = TRUE,
       use.packages = FALSE, ci.force = TRUE, label = "tab:HXCat12_2_texreg",
       stars = 0.05, caption = TCap, custom.note = FN)
```

	Model 1	Model 2
(Intercept)	−1.03*	−3.91*
	[−1.16; −0.89]	[−4.05; −3.78]
WhenBefore	−2.05*	−1.63*
	[−2.37; −1.73]	[−1.94; −1.32]
WhenDuring	−2.94*	1.08*
	[−3.44; −2.44]	[0.58; 1.58]
Scale parameter: gamma	1.36	1.33
Scale parameter: SE	0.63	0.62
Correlation parameter: alpha	0.03	0.05
Correlation parameter: SE	0.02	0.04
Num. obs.	3175	3175
Num. clust.	1082	1082

* p < .05, null hypothesis value outside 95% confidence interval based on z-score.

Note. The data contain up to 3 longitudinal observations (one per period) for each offender (i.e., cluster). The reference priod for When was after the testing window. Both models used a log link function, exchangeable correlation structure, a Poisson distribution, and robust standard errors (sandwich estimator). Model 1 did not use an offset. Model 2 used log(Years) as an offset term.

Table 26: Generalized Estimating Equation Models for Assault Involving Domestic Violence and/or Stalking Incident Counts Predicted by When Incidents Occurred

5.7.2.2 Estimated Marginal Means Table 27 provides the estimated marginal means obtained from Models 1 and 2, along with corresponding confidence intervals. The units of measurement for those means are counts per offender for Model 1 (which ignore period duration) and counts per offender-year for Model 2 (which adjust for duration and are thus considered annual incidence rates).

```
# Table caption.
TCap <- paste("Marginal Means of Assault Involving Domestic Violenace and/or",
              "Stalking Incident Counts and Incidence Rates",
              "By When Incidents Occurred")

# Table footnote.
FN <- paste("Model 1 did not use an offset, so means are counts.",
            "Model 2 used log(Years) as an offset term, so means are annual",
            "incidence rates.",
            "Confidence intervals are based on z-scores.")

HXCat12_2_emmeans1 <- emmeans(HXCat12_2_gee1, specs = pairwise ~ When,
                              type = "response", ref = "During")
HXCat12_2_emmeans2 <- emmeans(HXCat12_2_gee2, specs = pairwise ~ When,
                              type = "response", ref = "During")

as_tibble(HXCat12_2_emmeans1$emmeans) %>%
  full_join(x = ., y = as_tibble(HXCat12_2_emmeans2$emmeans)) %>%
  mutate(Model = c(1, 1, 1, 2, 2, 2),
         When = factor(When, levels = c("Before", "During", "After"))) %>%
  arrange(Model, When) %>%
  select(When, rate, SE, df, asymp.LCL, asymp.UCL) %>%
  kable(format = "latex", booktabs = TRUE, caption = TCap, digits = 2,
        col.names = c("When", "Rate", "SE", "df", "LCL", "UCL")) %>%
  kable_styling() %>%
  add_header_above(c(" " = 4, "95% CI" = 2)) %>%
  group_rows(group_label = "Model 1: Mean per offender", start_row = 1,
             end_row = 3, italic = TRUE) %>%
  group_rows(group_label = "Model 2: Mean per offender-year",
             start_row = 4, end_row = 6, italic = TRUE) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
           threeparttable = TRUE)
```

```
## Joining, by = c("When", "rate", "SE", "df", "asymp.LCL", "asymp.UCL")
```

				95% CI	
When	Rate	SE	df	LCL	UCL
Model 1: Mean per offender					
Before	0.05	0.01	Inf	0.03	0.06
During	0.02	0.00	Inf	0.01	0.03
After	0.36	0.02	Inf	0.31	0.41
Model 2: Mean per offender-year					
Before	0.04	0.01	Inf	0.03	0.05
During	0.59	0.14	Inf	0.37	0.95
After	0.20	0.01	Inf	0.18	0.23

Note: Model 1 did not use an offset, so means are counts. Model 2 used log(Years) as an offset term, so means are annual incidence rates. Confidence intervals are based on z-scores.

Table 27: Marginal Means of Assault Involving Domestic Violenace and/or Stalking Incident Counts and Incidence Rates By When Incidents Occurred

5.7.2.3 Contrasts for Ratios of Estimated Marginal Means Finally, we also estimated a set of contrasts to compare the marginal means from Models 1 and 2 (Table 28). Each of these contrasts estimates the ratios of a pair of means rather than a raw difference between means. This is because of the nature of the underlying Poisson GEE model.

```
TCap <- paste("Contrasts Estimating Ratios of Marginal Means for Assault",
              "Involving Domestic Violence and/or Stalking",
              "Incident Counts and Incidence Rates By When Incidents",
              "Occurred")

# Table footnote.
FN <- paste("Model 1 did not use an offset, so means are counts.",
            "Model 2 used log(Years) as an offset term, so means are annual",
            "incidence rates.",
            "These contrasts estimate ratios of those quantities and used",
            "Tukey's method to adjust for multiple comparisons.",
            "Confidence intervals are based on z-scores.")

as_tibble(confint(HXCat12_2_emmeans1$contrasts)) %>%
  full_join(x = ., y = as_tibble(confint(HXCat12_2_emmeans2$contrasts))) %>%
  kable(format = "latex", booktabs = TRUE, caption = TCap, digits = 2,
        col.names = c("Contrast", "Ratio", "SE", "df", "LCL", "UCL")) %>%
  kable_styling() %>%
  add_header_above(c(" " = 4, "95% CI" = 2)) %>%
  group_rows(group_label = "Model 1: Ratios of means per offender",
             start_row = 1, end_row = 3, italic = TRUE) %>%
  group_rows(group_label = "Model 2: Ratios of means per offender-year",
             start_row = 4, end_row = 6, italic = TRUE) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
## Joining, by = c("contrast", "ratio", "SE", "df", "asymp.LCL", "asymp.UCL")
```

Contrast	Ratio	SE	df	95% CI LCL	95% CI UCL
Model 1: Ratios of means per offender					
After / Before	7.78	1.27	Inf	5.31	11.40
After / During	18.87	4.80	Inf	10.39	34.24
Before / During	2.43	0.68	Inf	1.26	4.67
Model 2: Ratios of means per offender-year					
After / Before	5.10	0.81	Inf	3.51	7.41
After / During	0.34	0.09	Inf	0.19	0.62
Before / During	0.07	0.02	Inf	0.03	0.13

Note:
Model 1 did not use an offset, so means are counts. Model 2 used log(Years) as an offset term, so means are annual incidence rates. These contrasts estimate ratios of those quantities and used Tukey's method to adjust for multiple comparisons. Confidence intervals are based on z-scores.

Table 28: Contrasts Estimating Ratios of Marginal Means for Assault Involving Domestic Violence and/or Stalking Incident Counts and Incidence Rates By When Incidents Occurred

5.7.3 Assault - Non-Sexual, Non-DV

Among these 1,082 suspected sexual offenders, there were 654 (60%) offenders who had criminal histories containing at least one incident associated with an arrest, charge, or conviction for a non-sexual, non-domestic violence assault. Table 29 summarizes the counts of unique assault incidents overall and by when the incident occurred, while Table 30 shows the frequency distributions underlying those summaries.

```
# Table caption.
TCap <- paste("Offender-Level Assault (Non-Sexual, Non-Domestic Violence)",
              "Incident Counts (Overall and By When Incident Occurred)")

# Summarize crime category counts (both overall and broken down by IWhen).
IDNEW %>%
  select(HXCat12_3, HXCat12_3_Before, HXCat12_3_During, HXCat12_3_After) %>%
  describe(., quant=c(.25, .50, .75)) %>%
  bind_cols(data.frame(Variable = TRowLabs), .) %>%
  select(Variable, n, mean, sd, min, max, skew, kurtosis, Q0.25, Q0.5, Q0.75) %>%
  kable(., format = "latex", booktabs = TRUE, digits = 2, row.names = FALSE,
        col.names = CNames, caption = TCap,
        linesep = c('', '', '', '\\addlinespace')) %>%
  column_spec(column = 1, width = "6 cm") %>%
  footnote(general = FN1, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
# Table caption.
TCap <- paste("Frequency Distributions for Number of Assault (Non-Sexual,",
              "Non-Domestic Violence) Incidents for Which Offender Was",
              "Arrested, Charged, or Convicted (Overall and By When Incident",
              "Occurred)")

# Get frequency distributions.
IDNEW %>%
  select(OID, HXCat12_3, HXCat12_3_Before, HXCat12_3_During, HXCat12_3_After) %>%
  rename(All = HXCat12_3, Before = HXCat12_3_Before, During = HXCat12_3_During,
         After = HXCat12_3_After) %>%
  pivot_longer(cols = -OID, names_to = "When", values_to = "Value") %>%
  group_by(When, Value) %>%
  count() %>%
  arrange(Value) %>%
  pivot_wider(names_from = When, values_from = n, values_fill = 0) %>%
```

Variable	N	Mean	SD	Min	Max	Skew	Kurtosis	Q25	Q50	Q75
Incident count (arrested, charged, or convicted)	1082	1.29	1.60	0	11	1.94	5.33	0	1	2
...Before testing window	1041	0.29	0.65	0	4	2.94	10.23	0	0	0
...During testing window	1052	0.05	0.23	0	2	4.90	25.40	0	0	0
...After testing window	1082	0.97	1.43	0	11	2.36	7.94	0	0	1

Note: Only offenders with valid testing window start dates are included. Only incidents where the offender was arrested for, charged with, or convicted of at least one offense from one of the 12 main crime categories were counted. Before, during, and after refer to when the incident associated with the record occurred, not when the arrest date, charge date, or conviction date occurred relative to the testing window.

Table 29: Offender-Level Assault (Non-Sexual, Non-Domestic Violence) Incident Counts (Overall and By When Incident Occurred)

```
mutate(All_p = 100*All/N_All,
       All_v = if_else(is.na(Value),
                       true = as.numeric(NA),
                       false = 100*All/N_All),
       Before_p = 100*Before/N_All,
       Before_v = if_else(is.na(Value),
                          true = as.numeric(NA),
                          false = 100*Before/N_Before),
       During_p = 100*During/N_All,
       During_v = if_else(is.na(Value),
                          true = as.numeric(NA),
                          false = 100*During/N_During),
       After_p = 100*After/N_All,
       After_v = if_else(is.na(Value),
                         true = as.numeric(NA),
                         false = 100*After/N_After)) %>%
select(Value, All, All_p, All_v, Before, Before_p, Before_v, During,
       During_p, During_v, After, After_p, After_v) %>%
kable(format = "latex", booktabs = TRUE, digits = 2,
      col.names = c("Value", rep(c("N", "%", "Valid %"), 4)),
      caption = TCap) %>%
kable_styling() %>%
add_header_above(c(" " = 1, "Overall" = 3, "Before TW" = 3, "During TW" = 3,
                   "After TW" = 3)) %>%
footnote(general = FN5, general_title = "Note: ", footnote_as_chunk = TRUE,
       threeparttable = TRUE)
```

	Overall			Before TW			During TW			After TW		
Value	N	%	Valid %	N	%	Valid %	N	%	Valid %	N	%	Valid %
0	428	39.56	39.56	821	75.88	78.87	1004	92.79	95.44	557	51.48	51.48
1	308	28.47	28.47	170	15.71	16.33	45	4.16	4.28	269	24.86	24.86
2	157	14.51	14.51	29	2.68	2.79	3	0.28	0.29	130	12.01	12.01
3	90	8.32	8.32	13	1.20	1.25	0	0.00	0.00	60	5.55	5.55
4	47	4.34	4.34	8	0.74	0.77	0	0.00	0.00	33	3.05	3.05
5	30	2.77	2.77	0	0.00	0.00	0	0.00	0.00	17	1.57	1.57
6	8	0.74	0.74	0	0.00	0.00	0	0.00	0.00	5	0.46	0.46
7	4	0.37	0.37	0	0.00	0.00	0	0.00	0.00	4	0.37	0.37
8	3	0.28	0.28	0	0.00	0.00	0	0.00	0.00	3	0.28	0.28
9	4	0.37	0.37	0	0.00	0.00	0	0.00	0.00	2	0.18	0.18
10	2	0.18	0.18	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
11	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
NA	0	0.00	NA	41	3.79	NA	30	2.77	NA	0	0.00	NA

Note: N = 1,082. TW, earliest testing window. Only offenders with valid testing window start dates are included. Before, during, and after refer to when the incident associated with the record occurred, not when the arrest, charge, or conviction date occurred relative to the testing window. For example, an offender could be arrested during the testing window for an incident that occurred before it began. That would show up here in the before column. An incident is counted if there are any arrest, charge, or conviction records (any one type, or any combination of them will suffice) for the specified crime category associated with it.

Table 30: Frequency Distributions for Number of Assault (Non-Sexual, Non-Domestic Violence) Incidents for Which Offender Was Arrested, Charged, or Convicted (Overall and By When Incident Occurred)

5.7.3.1 GEE Models To more closely examine the assault (non-sexual, non-domestic violence) incident counts, we fitted two generalized estimating equation (GEE) models to the long-format *IDNEWL* data, after dropping rows for unobserved periods ($Years = 0$) because the counts are missing for those observations.

Model 1 used *When* as a predictor of the period-specific counts, but did not adjust for the period durations. Model 2 supplemented *When* as a predictor with an offset term for log(*Years*) to adjust for the period durations. Both models used Poisson distributions, an exchangeable correlation structure, a log link function, and robust standard errors based on the sandwich estimator.

```
HXCat12_3_gee1 <- IDNEWL %>%
  filter(Years > 0) %>%
  geeglm(HXCat12_3 ~ When,
         family = poisson(link = "log"), data = .,
         id = OID, corstr = "exchangeable", std.err = "san.se")

HXCat12_3_gee2 <- IDNEWL %>%
  filter(Years > 0) %>%
  geeglm(HXCat12_3 ~ When + offset(log(Years)),
         family = poisson(link = "log"), data = .,
         id = OID, corstr = "exchangeable", std.err = "san.se")
```

Table 31 below shows the parameter estimates for both GEE models.

```
TCap <- paste("Generalized Estimating Equation Models for Assault (Non-Sexual,",
              "Non-Domestic Violance) Incident Counts",
              "Predicted by When Incidents Occurred")
FN <- paste ("\\item * p < .05, null hypothesis value outside 95\\% confidence",
             "interval based on z-score.",
             "\\\\\n\\item Note.",
             "The data contain up to 3 longitudinal observations (one per",
             "period) for each offender (i.e., cluster).",
```

```
            "The reference priod for When was after the testing window.",
            "Both models used a log link function, exchangeable correlation",
            "structure, a Poisson distribution, and robust standard errors",
            "(sandwich estimator).",
            "Model 1 did not use an offset.",
            "Model 2 used log(Years) as an offset term.")

texreg(list(HXCat12_3_gee1, HXCat12_3_gee2), booktabs = TRUE, dcolumn = TRUE,
       threeparttable = TRUE, fontsize = "normalsize", table = TRUE,
       use.packages = FALSE, ci.force = TRUE, label = "tab:HXCat12_3_texreg",
       stars = 0.05, caption = TCap, custom.note = FN)
```

	Model 1	Model 2
(Intercept)	−0.03	−2.92*
	[−0.12; 0.06]	[−3.01; −2.83]
WhenBefore	−1.21*	−0.80*
	[−1.38; −1.05]	[−0.95; −0.64]
WhenDuring	−2.99*	1.02*
	[−3.29; −2.69]	[0.72; 1.32]
Scale parameter: gamma	1.56	1.59
Scale parameter: SE	0.19	0.20
Correlation parameter: alpha	0.02	0.04
Correlation parameter: SE	0.02	0.02
Num. obs.	3175	3175
Num. clust.	1082	1082

* p < .05, null hypothesis value outside 95% confidence interval based on z-score.

Note. The data contain up to 3 longitudinal observations (one per period) for each offender (i.e., cluster). The reference priod for When was after the testing window. Both models used a log link function, exchangeable correlation structure, a Poisson distribution, and robust standard errors (sandwich estimator). Model 1 did not use an offset. Model 2 used log(Years) as an offset term.

Table 31: Generalized Estimating Equation Models for Assault (Non-Sexual, Non-Domestic Violance) Incident Counts Predicted by When Incidents Occurred

5.7.3.2 Estimated Marginal Means Table 32 provides the estimated marginal means obtained from Models 1 and 2, along with corresponding confidence intervals. The units of measurement for those means are counts per offender for Model 1 (which ignore period duration) and counts per offender-year for Model 2 (which adjust for duration and are thus considered annual incidence rates).

```
# Table caption.
TCap <- paste("Marginal Means of Assault (Non-Sexual, Non-Domestic Violence)",
              "Incident Counts and Incidence Rates By When Incidents Occurred")

# Table footnote.
FN <- paste("Model 1 did not use an offset, so means are counts.",
            "Model 2 used log(Years) as an offset term, so means are annual",
            "incidence rates.",
            "Confidence intervals are based on z-scores.")

HXCat12_3_emmeans1 <- emmeans(HXCat12_3_gee1, specs = pairwise ~ When,
                              type = "response", ref = "During")
HXCat12_3_emmeans2 <- emmeans(HXCat12_3_gee2, specs = pairwise ~ When,
                              type = "response", ref = "During")

as_tibble(HXCat12_3_emmeans1$emmeans) %>%
  full_join(x = ., y = as_tibble(HXCat12_3_emmeans2$emmeans)) %>%
  mutate(Model = c(1, 1, 1, 2, 2, 2),
```

```
        When = factor(When, levels = c("Before", "During", "After"))) %>%
  arrange(Model, When) %>%
  select(When, rate, SE, df, asymp.LCL, asymp.UCL) %>%
  kable(format = "latex", booktabs = TRUE, caption = TCap, digits = 2,
        col.names = c("When", "Rate", "SE", "df", "LCL", "UCL")) %>%
  kable_styling() %>%
  add_header_above(c(" " = 4, "95% CI" = 2)) %>%
  group_rows(group_label = "Model 1: Mean per offender", start_row = 1,
             end_row = 3, italic = TRUE) %>%
  group_rows(group_label = "Model 2: Mean per offender-year",
             start_row = 4, end_row = 6, italic = TRUE) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
           threeparttable = TRUE)
```

```
## Joining, by = c("When", "rate", "SE", "df", "asymp.LCL", "asymp.UCL")
```

				95% CI	
When	Rate	SE	df	LCL	UCL
Model 1: Mean per offender					
Before	0.29	0.02	Inf	0.25	0.33
During	0.05	0.01	Inf	0.04	0.06
After	0.97	0.04	Inf	0.89	1.06
Model 2: Mean per offender-year					
Before	0.25	0.02	Inf	0.22	0.28
During	1.51	0.22	Inf	1.13	2.00
After	0.54	0.02	Inf	0.50	0.60

Note: Model 1 did not use an offset, so means are counts. Model 2 used log(Years) as an offset term, so means are annual incidence rates. Confidence intervals are based on z-scores.

Table 32: Marginal Means of Assault (Non-Sexual, Non-Domestic Violence) Incident Counts and Incidence Rates By When Incidents Occurred

5.7.3.3 Contrasts for Ratios of Estimated Marginal Means Finally, we also estimated a set of contrasts to compare the marginal means from Models 1 and 2 (Table 33). Each of these contrasts estimates the ratios of a pair of means rather than a raw difference between means. This is because of the nature of the underlying Poisson GEE model.

```
TCap <- paste("Contrasts Estimating Ratios of Marginal Means for Assault",
              "(Non-Sexual, Non-Domestic Violence)",
              "Incident Counts and Incidence Rates By When Incidents",
              "Occurred")

# Table footnote.
FN <- paste("Model 1 did not use an offset, so means are counts.",
            "Model 2 used log(Years) as an offset term, so means are annual",
            "incidence rates.",
            "These contrasts estimate ratios of those quantities and used",
            "Tukey's method to adjust for multiple comparisons.",
            "Confidence intervals are based on z-scores.")

as_tibble(confint(HXCat12_3_emmeans1$contrasts)) %>%
  full_join(x = ., y = as_tibble(confint(HXCat12_3_emmeans2$contrasts))) %>%
  kable(format = "latex", booktabs = TRUE, caption = TCap, digits = 2,
        col.names = c("Contrast", "Ratio", "SE", "df", "LCL", "UCL")) %>%
  kable_styling() %>%
  add_header_above(c(" " = 4, "95% CI" = 2)) %>%
  group_rows(group_label = "Model 1: Ratios of means per offender",
             start_row = 1, end_row = 3, italic = TRUE) %>%
```

```
  group_rows(group_label = "Model 2: Ratios of means per offender-year",
             start_row = 4, end_row = 6, italic = TRUE) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
## Joining, by = c("contrast", "ratio", "SE", "df", "asymp.LCL", "asymp.UCL")
```

				95% CI	
Contrast	Ratio	SE	df	LCL	UCL
Model 1: Ratios of means per offender					
After / Before	3.37	0.28	Inf	2.77	4.09
After / During	19.95	3.04	Inf	13.96	28.53
Before / During	5.92	0.92	Inf	4.12	8.52
Model 2: Ratios of means per offender-year					
After / Before	2.22	0.17	Inf	1.85	2.66
After / During	0.36	0.06	Inf	0.25	0.52
Before / During	0.16	0.03	Inf	0.11	0.23

Note:
Model 1 did not use an offset, so means are counts. Model 2 used log(Years) as an offset term, so means are annual incidence rates. These contrasts estimate ratios of those quantities and used Tukey's method to adjust for multiple comparisons. Confidence intervals are based on z-scores.

Table 33: Contrasts Estimating Ratios of Marginal Means for Assault (Non-Sexual, Non-Domestic Violence) Incident Counts and Incidence Rates By When Incidents Occurred

5.7.4 Burglary

Among these 1,082 suspected sexual offenders, there were 350 (32%) offenders who had criminal histories containing at least one incident associated with an arrest, charge, or conviction for burglary. Table 34 summarizes the counts of unique burglary incidents overall and by when the incident occurred, while Table 35 shows the frequency distributions underlying those summaries.

```
# Table caption.
TCap <- paste("Offender-Level Burglary Incident Counts (Overall and By When",
              "Incident Occurred)")

# Summarize crime category counts (both overall and broken down by IWhen).
IDNEW %>%
  select(HXCat12_4, HXCat12_4_Before, HXCat12_4_During, HXCat12_4_After) %>%
  describe(., quant=c(.25, .50, .75)) %>%
  bind_cols(data.frame(Variable = TRowLabs), .) %>%
  select(Variable, n, mean, sd, min, max, skew, kurtosis, Q0.25, Q0.5, Q0.75) %>%
  kable(., format = "latex", booktabs = TRUE, digits = 2, row.names = FALSE,
        col.names = CNames, caption = TCap,
        linesep = c('', '', '', '\\addlinespace')) %>%
  column_spec(column = 1, width = "6 cm") %>%
  footnote(general = FN1, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
# Table caption.
TCap <- paste("Frequency Distributions for Number of Burglary Incidents for",
              "Which Offender Was Arrested, Charged, or Convicted (Overall and",
              "By When Incident Occurred)")

# Get frequency distributions.
```

Variable	N	Mean	SD	Min	Max	Skew	Kurtosis	Q25	Q50	Q75
Incident count (arrested, charged, or convicted)	1082	0.67	1.36	0	10	3.13	12.46	0	0	1
...Before testing window	1041	0.29	0.78	0	8	3.82	20.54	0	0	0
...During testing window	1052	0.04	0.34	0	7	15.13	275.51	0	0	0
...After testing window	1082	0.35	0.93	0	9	3.94	20.26	0	0	0

Note: Only offenders with valid testing window start dates are included. Only incidents where the offender was arrested for, charged with, or convicted of at least one offense from one of the 12 main crime categories were counted. Before, during, and after refer to when the incident associated with the record occurred, not when the arrest date, charge date, or conviction date occurred relative to the testing window.

Table 34: Offender-Level Burglary Incident Counts (Overall and By When Incident Occurred)

```
IDNEW %>%
  select(OID, HXCat12_4, HXCat12_4_Before, HXCat12_4_During, HXCat12_4_After) %>%
  rename(All = HXCat12_4, Before = HXCat12_4_Before, During = HXCat12_4_During,
         After = HXCat12_4_After) %>%
  pivot_longer(cols = -OID, names_to = "When", values_to = "Value") %>%
  group_by(When, Value) %>%
  count() %>%
  arrange(Value) %>%
  pivot_wider(names_from = When, values_from = n, values_fill = 0) %>%
  mutate(All_p = 100*All/N_All,
         All_v = if_else(is.na(Value),
                         true = as.numeric(NA),
                         false = 100*All/N_All),
         Before_p = 100*Before/N_All,
         Before_v = if_else(is.na(Value),
                            true = as.numeric(NA),
                            false = 100*Before/N_Before),
         During_p = 100*During/N_All,
         During_v = if_else(is.na(Value),
                            true = as.numeric(NA),
                            false = 100*During/N_During),
         After_p = 100*After/N_All,
         After_v = if_else(is.na(Value),
                           true = as.numeric(NA),
                           false = 100*After/N_After)) %>%
  select(Value, All, All_p, All_v, Before, Before_p, Before_v, During,
         During_p, During_v, After, After_p, After_v) %>%
  kable(format = "latex", booktabs = TRUE, digits = 2,
        col.names = c("Value", rep(c("N", "%", "Valid %"), 4)),
        caption = TCap) %>%
  kable_styling() %>%
  add_header_above(c(" " = 1, "Overall" = 3, "Before TW" = 3, "During TW" = 3,
                     "After TW" = 3)) %>%
  footnote(general = FN5, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

	Overall			Before TW			During TW			After TW		
Value	N	%	Valid %	N	%	Valid %	N	%	Valid %	N	%	Valid %
0	732	67.65	67.65	857	79.21	82.32	1026	94.82	97.53	872	80.59	80.59
1	187	17.28	17.28	109	10.07	10.47	22	2.03	2.09	127	11.74	11.74
2	73	6.75	6.75	45	4.16	4.32	1	0.09	0.10	43	3.97	3.97
3	36	3.33	3.33	21	1.94	2.02	1	0.09	0.10	14	1.29	1.29
4	30	2.77	2.77	5	0.46	0.48	0	0.00	0.00	17	1.57	1.57
5	5	0.46	0.46	2	0.18	0.19	0	0.00	0.00	3	0.28	0.28
6	9	0.83	0.83	0	0.00	0.00	1	0.09	0.10	4	0.37	0.37
7	1	0.09	0.09	1	0.09	0.10	1	0.09	0.10	0	0.00	0.00
8	4	0.37	0.37	1	0.09	0.10	0	0.00	0.00	1	0.09	0.09
9	3	0.28	0.28	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
10	2	0.18	0.18	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
NA	0	0.00	NA	41	3.79	NA	30	2.77	NA	0	0.00	NA

Note: N = 1,082. TW, earliest testing window. Only offenders with valid testing window start dates are included. Before, during, and after refer to when the incident associated with the record occurred, not when the arrest, charge, or conviction date occurred relative to the testing window. For example, an offender could be arrested during the testing window for an incident that occurred before it began. That would show up here in the before column. An incident is counted if there are any arrest, charge, or conviction records (any one type, or any combination of them will suffice) for the specified crime category associated with it.

Table 35: Frequency Distributions for Number of Burglary Incidents for Which Offender Was Arrested, Charged, or Convicted (Overall and By When Incident Occurred)

5.7.4.1 GEE Models To more closely examine the burglary incident counts, we fitted two generalized estimating equation (GEE) models to the long-format *IDNEWL* data, after dropping rows for unobserved periods (*Years* = 0) because the counts are missing for those observations.

Model 1 used *When* as a predictor of the period-specific counts, but did not adjust for the period durations. Model 2 supplemented *When* as a predictor with an offset term for log(*Years*) to adjust for the period durations. Both models used Poisson distributions, an exchangeable correlation structure, a log link function, and robust standard errors based on the sandwich estimator.

```
HXCat12_4_gee1 <- IDNEWL %>%
  filter(Years > 0) %>%
  geeglm(HXCat12_4 ~ When,
         family = poisson(link = "log"), data = .,
         id = OID, corstr = "exchangeable", std.err = "san.se")

HXCat12_4_gee2 <- IDNEWL %>%
  filter(Years > 0) %>%
  geeglm(HXCat12_4 ~ When + offset(log(Years)),
         family = poisson(link = "log"), data = .,
         id = OID, corstr = "exchangeable", std.err = "san.se")
```

Table 36 below shows the parameter estimates for both GEE models.

```
TCap <- paste("Generalized Estimating Equation Models for Burglary Incident Counts",
              "Predicted by When Incidents Occurred")
FN <- paste ("\\item * p < .05, null hypothesis value outside 95\\% confidence",
             "interval based on z-score.",
             "\\\\\n\\item Note.",
             "The data contain up to 3 longitudinal observations (one per",
             "period) for each offender (i.e., cluster).",
             "The reference priod for When was after the testing window.",
             "Both models used a log link function, exchangeable correlation",
```

```
            "structure, a Poisson distribution, and robust standard errors",
            "(sandwich estimator).",
            "Model 1 did not use an offset.",
            "Model 2 used log(Years) as an offset term.")

texreg(list(HXCat12_4_gee1, HXCat12_4_gee2), booktabs = TRUE, dcolumn = TRUE,
       threeparttable = TRUE, fontsize = "normalsize", table = TRUE,
       use.packages = FALSE, ci.force = TRUE, label = "tab:HXCat12_4_texreg",
       stars = 0.05, caption = TCap, custom.note = FN)
```

	Model 1	Model 2
(Intercept)	−1.05*	−3.94*
	[−1.21; −0.89]	[−4.10; −3.78]
WhenBefore	−0.17	0.26*
	[−0.38; 0.03]	[0.05; 0.46]
WhenDuring	−2.22*	1.79*
	[−2.77; −1.67]	[1.24; 2.34]
Scale parameter: gamma	2.49	2.35
Scale parameter: SE	0.95	0.94
Correlation parameter: alpha	0.10	0.13
Correlation parameter: SE	0.05	0.08
Num. obs.	3175	3175
Num. clust.	1082	1082

* p < .05, null hypothesis value outside 95% confidence interval based on z-score.

Note. The data contain up to 3 longitudinal observations (one per period) for each offender (i.e., cluster). The reference priod for When was after the testing window. Both models used a log link function, exchangeable correlation structure, a Poisson distribution, and robust standard errors (sandwich estimator). Model 1 did not use an offset. Model 2 used log(Years) as an offset term.

Table 36: Generalized Estimating Equation Models for Burglary Incident Counts Predicted by When Incidents Occurred

5.7.4.2 Estimated Marginal Means Table 37 provides the estimated marginal means obtained from Models 1 and 2, along with corresponding confidence intervals. The units of measurement for those means are counts per offender for Model 1 (which ignore period duration) and counts per offender-year for Model 2 (which adjust for duration and are thus considered annual incidence rates).

```
# Table caption.
TCap <- paste("Marginal Means of Burglary Incident Counts and Incidence Rates",
              "By When Incidents Occurred")

# Table footnote.
FN <- paste("Model 1 did not use an offset, so means are counts.",
            "Model 2 used log(Years) as an offset term, so means are annual",
            "incidence rates.",
            "Confidence intervals are based on z-scores.")

HXCat12_4_emmeans1 <- emmeans(HXCat12_4_gee1, specs = pairwise ~ When,
                              type = "response", ref = "During")
HXCat12_4_emmeans2 <- emmeans(HXCat12_4_gee2, specs = pairwise ~ When,
                              type = "response", ref = "During")

as_tibble(HXCat12_4_emmeans1$emmeans) %>%
  full_join(x = ., y = as_tibble(HXCat12_4_emmeans2$emmeans)) %>%
  mutate(Model = c(1, 1, 1, 2, 2, 2),
         When = factor(When, levels = c("Before", "During", "After"))) %>%
  arrange(Model, When) %>%
  select(When, rate, SE, df, asymp.LCL, asymp.UCL) %>%
```

```
  kable(format = "latex", booktabs = TRUE, caption = TCap, digits = 2,
        col.names = c("When", "Rate", "SE", "df", "LCL", "UCL")) %>%
  kable_styling() %>%
  add_header_above(c(" " = 4, "95% CI" = 2)) %>%
  group_rows(group_label = "Model 1: Mean per offender", start_row = 1,
             end_row = 3, italic = TRUE) %>%
  group_rows(group_label = "Model 2: Mean per offender-year",
             start_row = 4, end_row = 6, italic = TRUE) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
## Joining, by = c("When", "rate", "SE", "df", "asymp.LCL", "asymp.UCL")
```

				95% CI	
When	Rate	SE	df	LCL	UCL
Model 1: Mean per offender					
Before	0.29	0.02	Inf	0.25	0.35
During	0.04	0.01	Inf	0.02	0.06
After	0.35	0.03	Inf	0.30	0.41
Model 2: Mean per offender-year					
Before	0.25	0.02	Inf	0.22	0.30
During	1.18	0.32	Inf	0.70	2.01
After	0.20	0.02	Inf	0.17	0.23

Note: Model 1 did not use an offset, so means are counts. Model 2 used log(Years) as an offset term, so means are annual incidence rates. Confidence intervals are based on z-scores.

Table 37: Marginal Means of Burglary Incident Counts and Incidence Rates By When Incidents Occurred

5.7.4.3 Contrasts for Ratios of Estimated Marginal Means Finally, we also estimated a set of contrasts to compare the marginal means from Models 1 and 2 (Table 38). Each of these contrasts estimates the ratios of a pair of means rather than a raw difference between means. This is because of the nature of the underlying Poisson GEE model.

```
TCap <- paste("Contrasts Estimating Ratios of Marginal Means for Burglary",
              "Incident Counts and Incidence Rates By When Incidents",
              "Occurred")

# Table footnote.
FN <- paste("Model 1 did not use an offset, so means are counts.",
            "Model 2 used log(Years) as an offset term, so means are annual",
            "incidence rates.",
            "These contrasts estimate ratios of those quantities and used",
            "Tukey's method to adjust for multiple comparisons.",
            "Confidence intervals are based on z-scores.")

as_tibble(confint(HXCat12_4_emmeans1$contrasts)) %>%
  full_join(x = ., y = as_tibble(confint(HXCat12_4_emmeans2$contrasts))) %>%
  kable(format = "latex", booktabs = TRUE, caption = TCap, digits = 2,
        col.names = c("Contrast", "Ratio", "SE", "df", "LCL", "UCL")) %>%
  kable_styling() %>%
  add_header_above(c(" " = 4, "95% CI" = 2)) %>%
  group_rows(group_label = "Model 1: Ratios of means per offender",
             start_row = 1, end_row = 3, italic = TRUE) %>%
  group_rows(group_label = "Model 2: Ratios of means per offender-year",
             start_row = 4, end_row = 6, italic = TRUE) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
## Joining, by = c("contrast", "ratio", "SE", "df", "asymp.LCL", "asymp.UCL")
```

Contrast	Ratio	SE	df	95% CI LCL	95% CI UCL
Model 1: Ratios of means per offender					
After / Before	1.19	0.13	Inf	0.93	1.53
After / During	9.23	2.59	Inf	4.78	17.82
Before / During	7.76	2.14	Inf	4.07	14.80
Model 2: Ratios of means per offender-year					
After / Before	0.77	0.08	Inf	0.61	0.98
After / During	0.17	0.05	Inf	0.09	0.32
Before / During	0.22	0.06	Inf	0.11	0.41

Note:
Model 1 did not use an offset, so means are counts. Model 2 used log(Years) as an offset term, so means are annual incidence rates. These contrasts estimate ratios of those quantities and used Tukey's method to adjust for multiple comparisons. Confidence intervals are based on z-scores.

Table 38: Contrasts Estimating Ratios of Marginal Means for Burglary Incident Counts and Incidence Rates By When Incidents Occurred

5.7.5 Criminal Sexual Conduct

Among these 1,082 suspected sexual offenders, there were 431 (40%) offenders who had criminal histories containing at least one incident associated with an arrest, charge, or conviction for criminal sexual conduct. Table 39 summarizes the counts of unique criminal sexual conduct incidents overall and by when the incident occurred, while Table 40 shows the frequency distributions underlying those summaries.

```
# Table caption.
TCap <- paste("Offender-Level Criminal Sexual Conduct Incidents (Overall and",
              "By When Incident Occurred)")

# Summarize crime category counts (both overall and broken down by IWhen).
IDNEW %>%
  select(HXCat12_5, HXCat12_5_Before, HXCat12_5_During, HXCat12_5_After) %>%
  describe(., quant=c(.25, .50, .75)) %>%
  bind_cols(data.frame(Variable = TRowLabs), .) %>%
  select(Variable, n, mean, sd, min, max, skew, kurtosis, Q0.25, Q0.5, Q0.75) %>%
  kable(., format = "latex", booktabs = TRUE, digits = 2, row.names = FALSE,
        col.names = CNames, caption = TCap,
        linesep = c('', '', '', '\\addlinespace')) %>%
  column_spec(column = 1, width = "6 cm") %>%
  footnote(general = FN1, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
# Table caption.
TCap <- paste("Frequency Distributions for Number of Criminal Sexual Conduct",
              "Incidents for Which Offender Was Arrested, Charged, or",
              "Convicted (Overall and By When Incident Occurred)")

# Get frequency distributions.
IDNEW %>%
  select(OID, HXCat12_5, HXCat12_5_Before, HXCat12_5_During, HXCat12_5_After) %>%
  rename(All = HXCat12_5, Before = HXCat12_5_Before,
         During = HXCat12_5_During, After = HXCat12_5_After) %>%
  pivot_longer(cols = -OID, names_to = "When", values_to = "Value") %>%
  group_by(When, Value) %>%
  count() %>%
  arrange(Value) %>%
  pivot_wider(names_from = When, values_from = n, values_fill = 0) %>%
  mutate(All_p = 100*All/N_All,
```

Variable	N	Mean	SD	Min	Max	Skew	Kurtosis	Q25	Q50	Q75
Incident count (arrested, charged, or convicted)	1082	0.63	1.05	0	8	2.78	10.66	0	0	1
...Before testing window	1041	0.12	0.37	0	3	3.41	13.35	0	0	0
...During testing window	1052	0.22	0.54	0	7	5.10	48.27	0	0	0
...After testing window	1082	0.30	0.78	0	6	4.00	20.47	0	0	0

Note: Only offenders with valid testing window start dates are included. Only incidents where the offender was arrested for, charged with, or convicted of at least one offense from one of the 12 main crime categories were counted. Before, during, and after refer to when the incident associated with the record occurred, not when the arrest date, charge date, or conviction date occurred relative to the testing window.

Table 39: Offender-Level Criminal Sexual Conduct Incidents (Overall and By When Incident Occurred)

```
       All_v = if_else(is.na(Value),
                       true = as.numeric(NA),
                       false = 100*All/N_All),
       Before_p = 100*Before/N_All,
       Before_v = if_else(is.na(Value),
                          true = as.numeric(NA),
                          false = 100*Before/N_Before),
       During_p = 100*During/N_All,
       During_v = if_else(is.na(Value),
                          true = as.numeric(NA),
                          false = 100*During/N_During),
       After_p = 100*After/N_All,
       After_v = if_else(is.na(Value),
                         true = as.numeric(NA),
                         false = 100*After/N_After)) %>%
select(Value, All, All_p, All_v, Before, Before_p, Before_v, During,
       During_p, During_v, After, After_p, After_v) %>%
kable(format = "latex", booktabs = TRUE, digits = 2,
      col.names = c("Value", rep(c("N", "%", "Valid %"), 4)),
      caption = TCap) %>%
kable_styling() %>%
add_header_above(c(" " = 1, "Overall" = 3, "Before TW" = 3, "During TW" = 3,
                   "After TW" = 3)) %>%
footnote(general = FN5, general_title = "Note: ", footnote_as_chunk = TRUE,
       threeparttable = TRUE)
```

	Overall			Before TW			During TW			After TW		
Value	N	%	Valid %	N	%	Valid %	N	%	Valid %	N	%	Valid %
0	651	60.17	60.17	929	85.86	89.24	853	78.84	81.08	875	80.87	80.87
1	299	27.63	27.63	99	9.15	9.51	182	16.82	17.30	147	13.59	13.59
2	76	7.02	7.02	11	1.02	1.06	12	1.11	1.14	35	3.23	3.23
3	29	2.68	2.68	2	0.18	0.19	2	0.18	0.19	10	0.92	0.92
4	12	1.11	1.11	0	0.00	0.00	1	0.09	0.10	8	0.74	0.74
5	4	0.37	0.37	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
6	7	0.65	0.65	0	0.00	0.00	0	0.00	0.00	6	0.55	0.55
7	3	0.28	0.28	0	0.00	0.00	2	0.18	0.19	0	0.00	0.00
8	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
NA	0	0.00	NA	41	3.79	NA	30	2.77	NA	0	0.00	NA

Note: N = 1,082. TW, earliest testing window. Only offenders with valid testing window start dates are included. Before, during, and after refer to when the incident associated with the record occurred, not when the arrest, charge, or conviction date occurred relative to the testing window. For example, an offender could be arrested during the testing window for an incident that occurred before it began. That would show up here in the before column. An incident is counted if there are any arrest, charge, or conviction records (any one type, or any combination of them will suffice) for the specified crime category associated with it.

Table 40: Frequency Distributions for Number of Criminal Sexual Conduct Incidents for Which Offender Was Arrested, Charged, or Convicted (Overall and By When Incident Occurred)

5.7.5.1 Overlap Between During Period Incidents and Earliest Known SAKs The earliest known SAKs come from a different data source than the criminal history records. Those SAKs were tested long after most of the offenders' criminal history records had accumulated. It is possible that some or even all of the criminal sexual conduct incidents found in the offenders' criminal histories during the testing window represent the same event as the earliest known SAK that starts the testing window. It is also possible that some or all of the criminal sexual conduct incidents in the criminal histories are entirely different events that show repeat offending behavior within a short period of time. Therefore, below we assess how much overlap there is between the earliest SAKs that define when the testing windows start and the CSC incidents from the criminal histories that occur during that period.

We start by examining criminal sexual conduct incident records. On these records, the *IID* variable is a primary key used for uniquely identifying each incident, the *ESAK_IID* variable on an incident record contains the IID associated with the offender's earliest known SAK, and *IWhen* classifies each incident as occurring before, during, or after the testing window associated with the offender's earliest known SAK. When *IID* = *ESAK_IID*, the incident record overlaps with the earliest known SAK. So, Table 41 shows how many CSC incident records there are in each *IWhen* period that overlap with the earliest SAK.

```
# Create tibble with only CSC incident records.
INCEW %>%
  filter(HXCat12_5 == 1) %>%
  mutate(Overlap = (ESAK_IID == IID)) ->
  INCEW_CSC

# Create tibble with only CSC incident records during the testing window.
INCEW_CSC %>% filter(IWhen == "During") ->
  INCEW_CSCD

TCap <- paste("Number of Criminal Sexual Conduct Incidents By When Incident",
              "Occurred and Whether The Incident Overlaps The Offender's",
              "Earliest SAK")

# Crosstab
xtabs(~IWhen + Overlap, data = INCEW_CSC, addNA = TRUE) %>%
  addmargins() %>%
  as.data.frame() %>%
  pivot_wider(names_from = Overlap, values_from = Freq) %>%
  kable(., format = "latex", booktabs = TRUE, caption = TCap,
        col.names = c("IWhen", "False", "True", "Sum"),
```

```
      format.args = list(big.mark = ",")) %>%
  add_header_above(c("", "Overlap Earliest SAK" = 2, "")) %>%
  column_spec(2:4, width = "1.6cm")
```

	Overlap Earliest SAK		
IWhen	False	True	Sum
Before	127	0	127
During	66	164	230
After	320	0	320
Sum	513	164	677

Table 41: Number of Criminal Sexual Conduct Incidents By When Incident Occurred and Whether The Incident Overlaps The Offender's Earliest SAK

All of the overlap incidents should occur during the testing window and that is consistent with what we see in Table 41. Therefore, we move on to extract a few key numbers and do some additional computations to quantify the degree of overlap.

```
# No. of CSC incidents observed (all periods).
INCEW_CSC %>% nrow() -> N_ICSC
N_ICSC
```

```
## [1] 677
```

```
# No. of CSC incidents observed during testing window.
INCEW_CSCD %>% nrow() -> N_ICSCD
N_ICSCD
```

```
## [1] 230
```

```
# No. of unique offenders with >=1 CSC incident in CHR
INCEW_CSC %>% pull(OID) %>% unique() %>% length() -> N_OCSC
N_OCSC
```

```
## [1] 431
```

```
# No. of unique offenders with >=1 CSC incident in CHR during testing window
INCEW_CSCD %>% pull(OID) %>% unique() %>% length() -> N_OCSCD
N_OCSCD
```

```
## [1] 199
```

```
# No. of unique offenders whose earliest SAK overlaps a CSC incident in CHR
# There is a maximum of one overlap incident per offender, so this is also the
# number of unique overlap CSC incidents.
INCEW_CSC %>% filter(ESAK_IID == IID) %>% pull(OID) %>% unique() %>% length() ->
  N_Overlap
N_Overlap
```

```
## [1] 164
```

From an incident perspective, there were a total of 677 CSC incidents in the criminal history data, but only 230 (33.97%) of them occurred during an offender's testing window. The 164 incidents that overlap with the offender's earliest known SAK comprise 24.22% of all the CSC incidents and 71.3% of the CSC incidents that occurred during testing windows. So, most of the CSC incidents that occurred during an offender's testing window are linked to the very SAK that started the testing window itself.

Meanwhile, from an offender perspective, there are 431 offenders (39.83% of the sample) with at least one CSC incident in their criminal history records. However, only 199 (18.39%) offenders have one or more CSC incidents in their criminal history that specifically fall during their testing window. Further subsetting shows that there are 164 offenders for whom we believe the earliest known SAK can be linked to a specific CSC incident from their criminal history data. They comprise 15.16% of the full sample and 82.41% of the subset of offenders who had at least one CSC incident during the testing window according to the criminal history records.

5.7.5.2 GEE Models To more closely examine the criminal sexual conduct incident counts, we fitted two generalized estimating equation (GEE) models to the long-format *IDNEWL* data, after dropping rows for unobserved periods (*Years* = 0) because the counts are missing for those observations.

Model 1 used *When* as a predictor of the period-specific counts, but did not adjust for the period durations. Model 2 supplemented *When* as a predictor with an offset term for log(*Years*) to adjust for the period durations. Both models used Poisson distributions, an exchangeable correlation structure, a log link function, and robust standard errors based on the sandwich estimator.

```
HXCat12_5_gee1 <- IDNEWL %>%
  filter(Years > 0) %>%
  geeglm(HXCat12_5 ~ When,
         family = poisson(link = "log"), data = .,
         id = OID, corstr = "exchangeable", std.err = "san.se")

HXCat12_5_gee2 <- IDNEWL %>%
  filter(Years > 0) %>%
  geeglm(HXCat12_5 ~ When + offset(log(Years)),
         family = poisson(link = "log"), data = .,
         id = OID, corstr = "exchangeable", std.err = "san.se")
```

Table 42 below shows the parameter estimates for both GEE models.

```
TCap <- paste("Generalized Estimating Equation Models for Criminal Sexual",
              "Conduct Incident Counts",
              "Predicted by When Incidents Occurred")
FN <- paste ("\\item * p < .05, null hypothesis value outside 95\\% confidence",
             "interval based on z-score.",
             "\\\\\n\\item Note.",
             "The data contain up to 3 longitudinal observations (one per",
             "period) for each offender (i.e., cluster).",
             "The reference priod for When was after the testing window.",
             "Both models used a log link function, exchangeable correlation",
             "structure, a Poisson distribution, and robust standard errors",
             "(sandwich estimator).",
             "Model 1 did not use an offset.",
             "Model 2 used log(Years) as an offset term.")

texreg(list(HXCat12_5_gee1, HXCat12_5_gee2), booktabs = TRUE, dcolumn = TRUE,
       threeparttable = TRUE, fontsize = "normalsize", table = TRUE,
       use.packages = FALSE, ci.force = TRUE, label = "tab:HXCat12_5_texreg",
       stars = 0.05, caption = TCap, custom.note = FN)
```

	Model 1	Model 2
(Intercept)	−1.22*	−4.11*
	[−1.37; −1.06]	[−4.27; −3.96]
WhenBefore	−0.89*	−0.46*
	[−1.11; −0.66]	[−0.69; −0.24]
WhenDuring	−0.30*	3.72*
	[−0.52; −0.08]	[3.50; 3.93]
Scale parameter: gamma	1.52	1.45
Scale parameter: SE	0.11	0.12
Correlation parameter: alpha	0.05	0.06
Correlation parameter: SE	0.02	0.02
Num. obs.	3175	3175
Num. clust.	1082	1082

* p < .05, null hypothesis value outside 95% confidence interval based on z-score.

Note. The data contain up to 3 longitudinal observations (one per period) for each offender (i.e., cluster). The reference priod for When was after the testing window. Both models used a log link function, exchangeable correlation structure, a Poisson distribution, and robust standard errors (sandwich estimator). Model 1 did not use an offset. Model 2 used log(Years) as an offset term.

Table 42: Generalized Estimating Equation Models for Criminal Sexual Conduct Incident Counts Predicted by When Incidents Occurred

5.7.5.3 Estimated Marginal Means Table 43 provides the estimated marginal means obtained from Models 1 and 2, along with corresponding confidence intervals. The units of measurement for those means are counts per offender for Model 1 (which ignore period duration) and counts per offender-year for Model 2 (which adjust for duration and are thus considered annual incidence rates).

```
# Table caption.
TCap <- paste("Marginal Means of Criminal Sexual Conduct Incident Counts and",
              "Incidence Rates By When Incidents Occurred")

# Table footnote.
FN <- paste("Model 1 did not use an offset, so means are counts.",
            "Model 2 used log(Years) as an offset term, so means are annual",
            "incidence rates.",
            "Confidence intervals are based on z-scores.")

HXCat12_5_emmeans1 <- emmeans(HXCat12_5_gee1, specs = pairwise ~ When,
                          type = "response", ref = "During")
HXCat12_5_emmeans2 <- emmeans(HXCat12_5_gee2, specs = pairwise ~ When,
                          type = "response", ref = "During")

as_tibble(HXCat12_5_emmeans1$emmeans) %>%
  full_join(x = ., y = as_tibble(HXCat12_5_emmeans2$emmeans)) %>%
  mutate(Model = c(1, 1, 1, 2, 2, 2),
         When = factor(When, levels = c("Before", "During", "After"))) %>%
  arrange(Model, When) %>%
  select(When, rate, SE, df, asymp.LCL, asymp.UCL) %>%
  kable(format = "latex", booktabs = TRUE, caption = TCap, digits = 2,
        col.names = c("When", "Rate", "SE", "df", "LCL", "UCL")) %>%
  kable_styling() %>%
  add_header_above(c(" " = 4, "95% CI" = 2)) %>%
  group_rows(group_label = "Model 1: Mean per offender", start_row = 1,
             end_row = 3, italic = TRUE) %>%
  group_rows(group_label = "Model 2: Mean per offender-year",
             start_row = 4, end_row = 6, italic = TRUE) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
## Joining, by = c("When", "rate", "SE", "df", "asymp.LCL", "asymp.UCL")
```

				95% CI	
When	Rate	SE	df	LCL	UCL
Model 1: Mean per offender					
Before	0.12	0.01	Inf	0.10	0.15
During	0.22	0.02	Inf	0.19	0.25
After	0.30	0.02	Inf	0.25	0.35
Model 2: Mean per offender-year					
Before	0.10	0.01	Inf	0.09	0.13
During	6.82	0.52	Inf	5.87	7.92
After	0.17	0.01	Inf	0.14	0.19

Note: Model 1 did not use an offset, so means are counts. Model 2 used log(Years) as an offset term, so means are annual incidence rates. Confidence intervals are based on z-scores.

Table 43: Marginal Means of Criminal Sexual Conduct Incident Counts and Incidence Rates By When Incidents Occurred

5.7.5.4 Contrasts for Ratios of Estimated Marginal Means Finally, we also estimated a set of contrasts to compare the marginal means from Models 1 and 2 (Table 44). Each of these contrasts estimates the ratios of a pair of means rather than a raw difference between means. This is because of the nature of the underlying Poisson GEE model.

```
TCap <- paste("Contrasts Estimating Ratios of Marginal Means for Criminal",
              "Sexual Conduct Incident Counts and Incidence Rates By When",
              "Incidents Occurred")

# Table footnote.
FN <- paste("Model 1 did not use an offset, so means are counts.",
            "Model 2 used log(Years) as an offset term, so means are annual",
            "incidence rates.",
            "These contrasts estimate ratios of those quantities and used",
            "Tukey's method to adjust for multiple comparisons.",
            "Confidence intervals are based on z-scores.")

as_tibble(confint(HXCat12_5_emmeans1$contrasts)) %>%
  full_join(x = ., y = as_tibble(confint(HXCat12_5_emmeans2$contrasts))) %>%
  kable(format = "latex", booktabs = TRUE, caption = TCap, digits = 2,
        col.names = c("Contrast", "Ratio", "SE", "df", "LCL", "UCL")) %>%
  kable_styling() %>%
  add_header_above(c(" " = 4, "95% CI" = 2)) %>%
  group_rows(group_label = "Model 1: Ratios of means per offender",
             start_row = 1, end_row = 3, italic = TRUE) %>%
  group_rows(group_label = "Model 2: Ratios of means per offender-year",
             start_row = 4, end_row = 6, italic = TRUE) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
## Joining, by = c("contrast", "ratio", "SE", "df", "asymp.LCL", "asymp.UCL")
```

				95% CI	
Contrast	Ratio	SE	df	LCL	UCL
Model 1: Ratios of means per offender					
After / Before	2.43	0.28	Inf	1.85	3.18
After / During	1.35	0.15	Inf	1.04	1.76
Before / During	0.56	0.07	Inf	0.42	0.74
Model 2: Ratios of means per offender-year					
After / Before	1.59	0.18	Inf	1.22	2.08
After / During	0.02	0.00	Inf	0.02	0.03
Before / During	0.02	0.00	Inf	0.01	0.02

Note:
Model 1 did not use an offset, so means are counts. Model 2 used log(Years) as an offset term, so means are annual incidence rates. These contrasts estimate ratios of those quantities and used Tukey's method to adjust for multiple comparisons. Confidence intervals are based on z-scores.

Table 44: Contrasts Estimating Ratios of Marginal Means for Criminal Sexual Conduct Incident Counts and Incidence Rates By When Incidents Occurred

5.7.6 Drug Crime

Among these 1,082 suspected sexual offenders, there were 596 (55%) offenders who had criminal histories containing at least one incident associated with an arrest, charge, or conviction for drug crime. Table 45 summarizes the counts of unique drug crime incidents overall and by and when the incident occurred, while Table 46 shows the frequency distributions underlying those summaries.

```
# Table caption.
TCap <- paste("Offender-Level Drug Crime Incident Counts (Overall and By When",
              "Incident Occurred)")

# Summarize crime category counts (both overall and broken down by IWhen).
IDNEW %>%
  select(HXCat12_6, HXCat12_6_Before, HXCat12_6_During, HXCat12_6_After) %>%
  describe(., quant=c(.25, .50, .75)) %>%
  bind_cols(data.frame(Variable = TRowLabs), .) %>%
  select(Variable, n, mean, sd, min, max, skew, kurtosis, Q0.25, Q0.5, Q0.75) %>%
  kable(., format = "latex", booktabs = TRUE, digits = 2, row.names = FALSE,
        col.names = CNames, caption = TCap,
        linesep = c('', '', '', '\\addlinespace')) %>%
  column_spec(column = 1, width = "6 cm") %>%
  footnote(general = FN1, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
# Table caption.
TCap <- paste("Frequency Distributions for Number of Drug Crime Incidents for",
              "Which Offender Was Arrested, Charged, or Convicted (Overall and",
              "By When Incident Occurred)")

# Get frequency distributions.
IDNEW %>%
  select(OID, HXCat12_6, HXCat12_6_Before, HXCat12_6_During, HXCat12_6_After) %>%
  rename(All = HXCat12_6, Before = HXCat12_6_Before, During = HXCat12_6_During,
         After = HXCat12_6_After) %>%
  pivot_longer(cols = -OID, names_to = "When", values_to = "Value") %>%
  group_by(When, Value) %>%
  count() %>%
  arrange(Value) %>%
  pivot_wider(names_from = When, values_from = n, values_fill = 0) %>%
  mutate(All_p = 100*All/N_All,
```

Variable	N	Mean	SD	Min	Max	Skew	Kurtosis	Q25	Q50	Q75
Incident count (arrested, charged, or convicted)	1082	1.38	1.86	0	14	1.94	5.01	0	1	2
...Before testing window	1041	0.48	0.90	0	6	2.24	5.41	0	0	1
...During testing window	1052	0.04	0.20	0	2	5.97	38.75	0	0	0
...After testing window	1082	0.88	1.52	0	14	2.63	9.92	0	0	1

Note: Only offenders with valid testing window start dates are included. Only incidents where the offender was arrested for, charged with, or convicted of at least one offense from one of the 12 main crime categories were counted. Before, during, and after refer to when the incident associated with the record occurred, not when the arrest date, charge date, or conviction date occurred relative to the testing window.

Table 45: Offender-Level Drug Crime Incident Counts (Overall and By When Incident Occurred)

```
       All_v = if_else(is.na(Value),
                       true = as.numeric(NA),
                       false = 100*All/N_All),
       Before_p = 100*Before/N_All,
       Before_v = if_else(is.na(Value),
                          true = as.numeric(NA),
                          false = 100*Before/N_Before),
       During_p = 100*During/N_All,
       During_v = if_else(is.na(Value),
                          true = as.numeric(NA),
                          false = 100*During/N_During),
       After_p = 100*After/N_All,
       After_v = if_else(is.na(Value),
                         true = as.numeric(NA),
                         false = 100*After/N_After)) %>%
select(Value, All, All_p, All_v, Before, Before_p, Before_v, During,
       During_p, During_v, After, After_p, After_v) %>%
kable(format = "latex", booktabs = TRUE, digits = 2,
      col.names = c("Value", rep(c("N", "%", "Valid %"), 4)),
      caption = TCap) %>%
kable_styling() %>%
add_header_above(c(" " = 1, "Overall" = 3, "Before TW" = 3, "During TW" = 3,
                   "After TW" = 3)) %>%
footnote(general = FN5, general_title = "Note: ", footnote_as_chunk = TRUE,
       threeparttable = TRUE)
```

	Overall			Before TW			During TW			After TW		
Value	N	%	Valid %	N	%	Valid %	N	%	Valid %	N	%	Valid %
0	486	44.92	44.92	735	67.93	70.61	1017	93.99	96.67	654	60.44	60.44
1	238	22.00	22.00	185	17.10	17.77	32	2.96	3.04	201	18.58	18.58
2	134	12.38	12.38	67	6.19	6.44	3	0.28	0.29	95	8.78	8.78
3	86	7.95	7.95	39	3.60	3.75	0	0.00	0.00	57	5.27	5.27
4	64	5.91	5.91	11	1.02	1.06	0	0.00	0.00	37	3.42	3.42
5	33	3.05	3.05	3	0.28	0.29	0	0.00	0.00	17	1.57	1.57
6	17	1.57	1.57	1	0.09	0.10	0	0.00	0.00	7	0.65	0.65
7	9	0.83	0.83	0	0.00	0.00	0	0.00	0.00	8	0.74	0.74
8	5	0.46	0.46	0	0.00	0.00	0	0.00	0.00	2	0.18	0.18
9	6	0.55	0.55	0	0.00	0.00	0	0.00	0.00	3	0.28	0.28
10	2	0.18	0.18	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
11	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
14	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
NA	0	0.00	NA	41	3.79	NA	30	2.77	NA	0	0.00	NA

Note: N = 1,082. TW, earliest testing window. Only offenders with valid testing window start dates are included. Before, during, and after refer to when the incident associated with the record occurred, not when the arrest, charge, or conviction date occurred relative to the testing window. For example, an offender could be arrested during the testing window for an incident that occurred before it began. That would show up here in the before column. An incident is counted if there are any arrest, charge, or conviction records (any one type, or any combination of them will suffice) for the specified crime category associated with it.

Table 46: Frequency Distributions for Number of Drug Crime Incidents for Which Offender Was Arrested, Charged, or Convicted (Overall and By When Incident Occurred)

5.7.6.1 GEE Models To more closely examine the drug crime incident counts, we fitted two generalized estimating equation (GEE) models to the long-format *IDNEWL* data, after dropping rows for unobserved periods (*Years* = 0) because the counts are missing for those observations.

Model 1 used *When* as a predictor of the period-specific counts, but did not adjust for the period durations. Model 2 supplemented *When* as a predictor with an offset term for log(*Years*) to adjust for the period durations. Both models used Poisson distributions, an exchangeable correlation structure, a log link function, and robust standard errors based on the sandwich estimator.

```
HXCat12_6_gee1 <- IDNEWL %>%
  filter(Years > 0) %>%
  geeglm(HXCat12_6 ~ When,
         family = poisson(link = "log"), data = .,
         id = OID, corstr = "exchangeable", std.err = "san.se")

HXCat12_6_gee2 <- IDNEWL %>%
  filter(Years > 0) %>%
  geeglm(HXCat12_6 ~ When + offset(log(Years)),
         family = poisson(link = "log"), data = .,
         id = OID, corstr = "exchangeable", std.err = "san.se")
```

Table 47 below shows the parameter estimates for both GEE models.

```
TCap <- paste("Generalized Estimating Equation Models for Drug Crime Incident",
              "Counts Predicted by When Incidents Occurred")
FN <- paste ("\\item * p < .05, null hypothesis value outside 95\\% confidence",
             "interval based on z-score.",
             "\\\\\n\\item Note.",
             "The data contain up to 3 longitudinal observations (one per",
             "period) for each offender (i.e., cluster).",
```

```
             "The reference priod for When was after the testing window.",
             "Both models used a log link function, exchangeable correlation",
             "structure, a Poisson distribution, and robust standard errors",
             "(sandwich estimator).",
             "Model 1 did not use an offset.",
             "Model 2 used log(Years) as an offset term.")

texreg(list(HXCat12_6_gee1, HXCat12_6_gee2), booktabs = TRUE, dcolumn = TRUE,
       threeparttable = TRUE, fontsize = "normalsize", table = TRUE,
       use.packages = FALSE, ci.force = TRUE, label = "tab:HXCat12_6_texreg",
       stars = 0.05, caption = TCap, custom.note = FN)
```

	Model 1	Model 2
(Intercept)	−0.13*	−3.03*
	[−0.23; −0.03]	[−3.13; −2.92]
WhenBefore	−0.60*	−0.16*
	[−0.74; −0.45]	[−0.30; −0.02]
WhenDuring	−3.19*	0.80*
	[−3.53; −2.84]	[0.44; 1.16]
Scale parameter: gamma	1.82	1.96
Scale parameter: SE	0.28	0.32
Correlation parameter: alpha	0.07	0.10
Correlation parameter: SE	0.02	0.03
Num. obs.	3175	3175
Num. clust.	1082	1082

* p < .05, null hypothesis value outside 95% confidence interval based on z-score.

Note. The data contain up to 3 longitudinal observations (one per period) for each offender (i.e., cluster). The reference priod for When was after the testing window. Both models used a log link function, exchangeable correlation structure, a Poisson distribution, and robust standard errors (sandwich estimator). Model 1 did not use an offset. Model 2 used log(Years) as an offset term.

Table 47: Generalized Estimating Equation Models for Drug Crime Incident Counts Predicted by When Incidents Occurred

5.7.6.2 Estimated Marginal Means Table 48 provides the estimated marginal means obtained from Models 1 and 2, along with corresponding confidence intervals. The units of measurement for those means are counts per offender for Model 1 (which ignore period duration) and counts per offender-year for Model 2 (which adjust for duration and are thus considered annual incidence rates).

```
# Table caption.
TCap <- paste("Marginal Means of Drug Crime Incident Counts and Incidence Rates",
              "By When Incidents Occurred")

# Table footnote.
FN <- paste("Model 1 did not use an offset, so means are counts.",
            "Model 2 used log(Years) as an offset term, so means are annual",
            "incidence rates.",
            "Confidence intervals are based on z-scores.")

HXCat12_6_emmeans1 <- emmeans(HXCat12_6_gee1, specs = pairwise ~ When,
                              type = "response", ref = "During")
HXCat12_6_emmeans2 <- emmeans(HXCat12_6_gee2, specs = pairwise ~ When,
                              type = "response", ref = "During")

as_tibble(HXCat12_6_emmeans1$emmeans) %>%
  full_join(x = ., y = as_tibble(HXCat12_6_emmeans2$emmeans)) %>%
  mutate(Model = c(1, 1, 1, 2, 2, 2),
```

```
        When = factor(When, levels = c("Before", "During", "After"))) %>%
  arrange(Model, When) %>%
  select(When, rate, SE, df, asymp.LCL, asymp.UCL) %>%
  kable(format = "latex", booktabs = TRUE, caption = TCap, digits = 2,
        col.names = c("When", "Rate", "SE", "df", "LCL", "UCL")) %>%
  kable_styling() %>%
  add_header_above(c(" " = 4, "95% CI" = 2)) %>%
  group_rows(group_label = "Model 1: Mean per offender", start_row = 1,
             end_row = 3, italic = TRUE) %>%
  group_rows(group_label = "Model 2: Mean per offender-year",
             start_row = 4, end_row = 6, italic = TRUE) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
## Joining, by = c("When", "rate", "SE", "df", "asymp.LCL", "asymp.UCL")
```

				95% CI	
When	Rate	SE	df	LCL	UCL
Model 1: Mean per offender					
Before	0.48	0.03	Inf	0.43	0.54
During	0.04	0.01	Inf	0.03	0.05
After	0.88	0.05	Inf	0.79	0.97
Model 2: Mean per offender-year					
Before	0.42	0.02	Inf	0.37	0.47
During	1.09	0.19	Inf	0.77	1.54
After	0.49	0.03	Inf	0.44	0.54

Note: Model 1 did not use an offset, so means are counts. Model 2 used log(Years) as an offset term, so means are annual incidence rates. Confidence intervals are based on z-scores.

Table 48: Marginal Means of Drug Crime Incident Counts and Incidence Rates By When Incidents Occurred

5.7.6.3 Contrasts for Ratios of Estimated Marginal Means Finally, we also estimated a set of contrasts to compare the marginal means from Models 1 and 2 (Table 49). Each of these contrasts estimates the ratios of a pair of means rather than a raw difference between means. This is because of the nature of the underlying Poisson GEE model.

```
TCap <- paste("Contrasts Estimating Ratios of Marginal Means for Drug Crime",
              "Incident Counts and Incidence Rates By When Incidents",
              "Occurred")

# Table footnote.
FN <- paste("Model 1 did not use an offset, so means are counts.",
            "Model 2 used log(Years) as an offset term, so means are annual",
            "incidence rates.",
            "These contrasts estimate ratios of those quantities and used",
            "Tukey's method to adjust for multiple comparisons.",
            "Confidence intervals are based on z-scores.")

as_tibble(confint(HXCat12_6_emmeans1$contrasts)) %>%
  full_join(x = ., y = as_tibble(confint(HXCat12_6_emmeans2$contrasts))) %>%
  kable(format = "latex", booktabs = TRUE, caption = TCap, digits = 2,
        col.names = c("Contrast", "Ratio", "SE", "df", "LCL", "UCL")) %>%
  kable_styling() %>%
  add_header_above(c(" " = 4, "95% CI" = 2)) %>%
  group_rows(group_label = "Model 1: Ratios of means per offender",
             start_row = 1, end_row = 3, italic = TRUE) %>%
  group_rows(group_label = "Model 2: Ratios of means per offender-year",
             start_row = 4, end_row = 6, italic = TRUE) %>%
```

```
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
## Joining, by = c("contrast", "ratio", "SE", "df", "asymp.LCL", "asymp.UCL")
```

				95% CI	
Contrast	Ratio	SE	df	LCL	UCL
Model 1: Ratios of means per offender					
After / Before	1.82	0.13	Inf	1.53	2.16
After / During	24.17	4.29	Inf	15.95	36.62
Before / During	13.28	2.33	Inf	8.81	20.03
Model 2: Ratios of means per offender-year					
After / Before	1.18	0.08	Inf	1.00	1.39
After / During	0.45	0.08	Inf	0.29	0.69
Before / During	0.38	0.07	Inf	0.25	0.59

Note:
Model 1 did not use an offset, so means are counts. Model 2 used log(Years) as an offset term, so means are annual incidence rates. These contrasts estimate ratios of those quantities and used Tukey's method to adjust for multiple comparisons. Confidence intervals are based on z-scores.

Table 49: Contrasts Estimating Ratios of Marginal Means for Drug Crime Incident Counts and Incidence Rates By When Incidents Occurred

5.7.7 Homicide

Among these 1,082 suspected sexual offenders, there were 102 (9%) offenders who had criminal histories containing at least one incident associated with an arrest, charge, or conviction for homicide. Table 50 summarizes the counts of unique homicide incidents overall and by when the incident occurred, while Table 51 shows the frequency distributions underlying those summaries.

```
# Table caption.
TCap <- paste("Offender-Level Homicide Incident Counts (Overall and By When",
              "Incident Occurred)")

# Summarize crime category counts (both overall and broken down by IWhen).
IDNEW %>%
  select(HXCat12_7, HXCat12_7_Before, HXCat12_7_During, HXCat12_7_After) %>%
  describe(., quant=c(.25, .50, .75)) %>%
  bind_cols(data.frame(Variable = TRowLabs), .) %>%
  select(Variable, n, mean, sd, min, max, skew, kurtosis, Q0.25, Q0.5, Q0.75) %>%
  kable(., format = "latex", booktabs = TRUE, digits = 2, row.names = FALSE,
        col.names = CNames, caption = TCap,
        linesep = c('', '', '', '\\addlinespace')) %>%
  column_spec(column = 1, width = "6 cm") %>%
  footnote(general = FN1, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
# Table caption.
TCap <- paste("Frequency Distributions for Number of Homicide Incidents for",
              "Which Offender Was Arrested, Charged, or Convicted (Overall and",
              "By When Incident Occurred)")

# Get frequency distributions.
IDNEW %>%
  select(OID, HXCat12_7, HXCat12_7_Before, HXCat12_7_During, HXCat12_7_After) %>%
```

Variable	N	Mean	SD	Min	Max	Skew	Kurtosis	Q25	Q50	Q75
Incident count (arrested, charged, or convicted)	1082	0.11	0.37	0	3	3.85	17.13	0	0	0
...Before testing window	1041	0.03	0.17	0	2	7.07	54.81	0	0	0
...During testing window	1052	0.01	0.09	0	1	11.32	126.26	0	0	0
...After testing window	1082	0.08	0.31	0	2	4.16	18.09	0	0	0

Note: Only offenders with valid testing window start dates are included. Only incidents where the offender was arrested for, charged with, or convicted of at least one offense from one of the 12 main crime categories were counted. Before, during, and after refer to when the incident associated with the record occurred, not when the arrest date, charge date, or conviction date occurred relative to the testing window.

Table 50: Offender-Level Homicide Incident Counts (Overall and By When Incident Occurred)

```
rename(All = HXCat12_7, Before = HXCat12_7_Before, During = HXCat12_7_During,
       After = HXCat12_7_After) %>%
pivot_longer(cols = -OID, names_to = "When", values_to = "Value") %>%
group_by(When, Value) %>%
count() %>%
arrange(Value) %>%
pivot_wider(names_from = When, values_from = n, values_fill = 0) %>%
mutate(All_p = 100*All/N_All,
       All_v = if_else(is.na(Value),
                       true = as.numeric(NA),
                       false = 100*All/N_All),
       Before_p = 100*Before/N_All,
       Before_v = if_else(is.na(Value),
                          true = as.numeric(NA),
                          false = 100*Before/N_Before),
       During_p = 100*During/N_All,
       During_v = if_else(is.na(Value),
                          true = as.numeric(NA),
                          false = 100*During/N_During),
       After_p = 100*After/N_All,
       After_v = if_else(is.na(Value),
                         true = as.numeric(NA),
                         false = 100*After/N_After)) %>%
select(Value, All, All_p, All_v, Before, Before_p, Before_v, During,
       During_p, During_v, After, After_p, After_v) %>%
kable(format = "latex", booktabs = TRUE, digits = 2,
      col.names = c("Value", rep(c("N", "%", "Valid %"), 4)),
      caption = TCap) %>%
kable_styling() %>%
add_header_above(c(" " = 1, "Overall" = 3, "Before TW" = 3, "During TW" = 3,
                   "After TW" = 3)) %>%
footnote(general = FN5, general_title = "Note: ", footnote_as_chunk = TRUE,
       threeparttable = TRUE)
```

	Overall			Before TW			During TW			After TW		
Value	N	%	Valid %	N	%	Valid %	N	%	Valid %	N	%	Valid %
0	980	90.57	90.57	1016	93.90	97.60	1044	96.49	99.24	1008	93.16	93.16
1	86	7.95	7.95	23	2.13	2.21	8	0.74	0.76	62	5.73	5.73
2	13	1.20	1.20	2	0.18	0.19	0	0.00	0.00	12	1.11	1.11
3	3	0.28	0.28	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
NA	0	0.00	NA	41	3.79	NA	30	2.77	NA	0	0.00	NA

Note: N = 1,082. TW, earliest testing window. Only offenders with valid testing window start dates are included. Before, during, and after refer to when the incident associated with the record occurred, not when the arrest, charge, or conviction date occurred relative to the testing window. For example, an offender could be arrested during the testing window for an incident that occurred before it began. That would show up here in the before column. An incident is counted if there are any arrest, charge, or conviction records (any one type, or any combination of them will suffice) for the specified crime category associated with it.

Table 51: Frequency Distributions for Number of Homicide Incidents for Which Offender Was Arrested, Charged, or Convicted (Overall and By When Incident Occurred)

5.7.7.1 GEE Models To more closely examine the homicide incident counts, we fitted two generalized estimating equation (GEE) models to the long-format *IDNEWL* data, after dropping rows for unobserved periods (*Years* = 0) because the counts are missing for those observations.

Model 1 used *When* as a predictor of the period-specific counts, but did not adjust for the period durations. Model 2 supplemented *When* as a predictor with an offset term for log(*Years*) to adjust for the period durations. Both models used Poisson distributions, an exchangeable correlation structure, a log link function, and robust standard errors based on the sandwich estimator.

```
HXCat12_7_gee1 <- IDNEWL %>%
  filter(Years > 0) %>%
  geeglm(HXCat12_7 ~ When,
         family = poisson(link = "log"), data = .,
         id = OID, corstr = "exchangeable", std.err = "san.se")

HXCat12_7_gee2 <- IDNEWL %>%
  filter(Years > 0) %>%
  geeglm(HXCat12_7 ~ When + offset(log(Years)),
         family = poisson(link = "log"), data = .,
         id = OID, corstr = "exchangeable", std.err = "san.se")
```

Table 52 below shows the parameter estimates for both GEE models.

```
TCap <- paste("Generalized Estimating Equation Models for Homicide Incident Counts",
              "Predicted by When Incidents Occurred")
FN <- paste ("\\item * p < .05, null hypothesis value outside 95\\% confidence",
             "interval based on z-score.",
             "\\\\\\n\\item Note.",
             "The data contain up to 3 longitudinal observations (one per",
             "period) for each offender (i.e., cluster).",
             "The reference priod for When was after the testing window.",
             "Both models used a log link function, exchangeable correlation",
             "structure, a Poisson distribution, and robust standard errors",
             "(sandwich estimator).",
             "Model 1 did not use an offset.",
             "Model 2 used log(Years) as an offset term.")

texreg(list(HXCat12_7_gee1, HXCat12_7_gee2), booktabs = TRUE, dcolumn = TRUE,
       threeparttable = TRUE, fontsize = "normalsize", table = TRUE,
       use.packages = FALSE, ci.force = TRUE, label = "tab:HXCat12_7_texreg",
       stars = 0.05, caption = TCap, custom.note = FN)
```

	Model 1	Model 2
(Intercept)	−2.53*	−5.42*
	[−2.76; −2.30]	[−5.65; −5.19]
WhenBefore	−1.12*	−0.69*
	[−1.58; −0.66]	[−1.14; −0.24]
WhenDuring	−2.35*	1.67*
	[−3.05; −1.65]	[0.97; 2.37]
Scale parameter: gamma	1.11	1.08
Scale parameter: SE	1.30	1.28
Correlation parameter: alpha	0.05	0.05
Correlation parameter: SE	0.08	0.08
Num. obs.	3175	3175
Num. clust.	1082	1082

* p < .05, null hypothesis value outside 95% confidence interval based on z-score.

Note. The data contain up to 3 longitudinal observations (one per period) for each offender (i.e., cluster). The reference priod for When was after the testing window. Both models used a log link function, exchangeable correlation structure, a Poisson distribution, and robust standard errors (sandwich estimator). Model 1 did not use an offset. Model 2 used log(Years) as an offset term.

Table 52: Generalized Estimating Equation Models for Homicide Incident Counts Predicted by When Incidents Occurred

5.7.7.2 Estimated Marginal Means Table 53 provides the estimated marginal means obtained from Models 1 and 2, along with corresponding confidence intervals. The units of measurement for those means are counts per offender for Model 1 (which ignore period duration) and counts per offender-year for Model 2 (which adjust for duration and are thus considered annual incidence rates).

```
# Table caption.
TCap <- paste("Marginal Means of Homicide Incident Counts and Incidence Rates",
              "By When Incidents Occurred")

# Table footnote.
FN <- paste("Model 1 did not use an offset, so means are counts.",
            "Model 2 used log(Years) as an offset term, so means are annual",
            "incidence rates.",
            "Confidence intervals are based on z-scores.")

HXCat12_7_emmeans1 <- emmeans(HXCat12_7_gee1, specs = pairwise ~ When,
                              type = "response", ref = "During")
HXCat12_7_emmeans2 <- emmeans(HXCat12_7_gee2, specs = pairwise ~ When,
                              type = "response", ref = "During")

as_tibble(HXCat12_7_emmeans1$emmeans) %>%
  full_join(x = ., y = as_tibble(HXCat12_7_emmeans2$emmeans)) %>%
  mutate(Model = c(1, 1, 1, 2, 2, 2),
         When = factor(When, levels = c("Before", "During", "After"))) %>%
  arrange(Model, When) %>%
  select(When, rate, SE, df, asymp.LCL, asymp.UCL) %>%
  kable(format = "latex", booktabs = TRUE, caption = TCap, digits = 2,
        col.names = c("When", "Rate", "SE", "df", "LCL", "UCL")) %>%
  kable_styling() %>%
  add_header_above(c(" " = 4, "95% CI" = 2)) %>%
  group_rows(group_label = "Model 1: Mean per offender", start_row = 1,
             end_row = 3, italic = TRUE) %>%
  group_rows(group_label = "Model 2: Mean per offender-year",
             start_row = 4, end_row = 6, italic = TRUE) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
## Joining, by = c("When", "rate", "SE", "df", "asymp.LCL", "asymp.UCL")
```

				95% CI	
When	Rate	SE	df	LCL	UCL
Model 1: Mean per offender					
Before	0.03	0.01	Inf	0.02	0.04
During	0.01	0.00	Inf	0.00	0.02
After	0.08	0.01	Inf	0.06	0.10
Model 2: Mean per offender-year					
Before	0.02	0.00	Inf	0.02	0.03
During	0.24	0.08	Inf	0.12	0.47
After	0.04	0.01	Inf	0.04	0.06

Note: Model 1 did not use an offset, so means are counts. Model 2 used log(Years) as an offset term, so means are annual incidence rates. Confidence intervals are based on z-scores.

Table 53: Marginal Means of Homicide Incident Counts and Incidence Rates By When Incidents Occurred

5.7.7.3 Contrasts for Ratios of Estimated Marginal Means Finally, we also estimated a set of contrasts to compare the marginal means from Models 1 and 2 (Table 54). Each of these contrasts estimates the ratios of a pair of means rather than a raw difference between means. This is because of the nature of the underlying Poisson GEE model.

```
TCap <- paste("Contrasts Estimating Ratios of Marginal Means for Homicide",
              "Incident Counts and Incidence Rates By When Incidents",
              "Occurred")

# Table footnote.
FN <- paste("Model 1 did not use an offset, so means are counts.",
            "Model 2 used log(Years) as an offset term, so means are annual",
            "incidence rates.",
            "These contrasts estimate ratios of those quantities and used",
            "Tukey's method to adjust for multiple comparisons.",
            "Confidence intervals are based on z-scores.")

as_tibble(confint(HXCat12_7_emmeans1$contrasts)) %>%
  full_join(x = ., y = as_tibble(confint(HXCat12_7_emmeans2$contrasts))) %>%
  kable(format = "latex", booktabs = TRUE, caption = TCap, digits = 2,
        col.names = c("Contrast", "Ratio", "SE", "df", "LCL", "UCL")) %>%
  kable_styling() %>%
  add_header_above(c(" " = 4, "95% CI" = 2)) %>%
  group_rows(group_label = "Model 1: Ratios of means per offender",
             start_row = 1, end_row = 3, italic = TRUE) %>%
  group_rows(group_label = "Model 2: Ratios of means per offender-year",
             start_row = 4, end_row = 6, italic = TRUE) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
## Joining, by = c("contrast", "ratio", "SE", "df", "asymp.LCL", "asymp.UCL")
```

Contrast	Ratio	SE	df	95% CI LCL	95% CI UCL
Model 1: Ratios of means per offender					
After / Before	3.07	0.72	Inf	1.77	5.33
After / During	10.48	3.72	Inf	4.56	24.08
Before / During	3.41	1.40	Inf	1.30	8.93
Model 2: Ratios of means per offender-year					
After / Before	1.99	0.46	Inf	1.16	3.42
After / During	0.19	0.07	Inf	0.08	0.43
Before / During	0.09	0.04	Inf	0.04	0.24

Note:
Model 1 did not use an offset, so means are counts. Model 2 used log(Years) as an offset term, so means are annual incidence rates. These contrasts estimate ratios of those quantities and used Tukey's method to adjust for multiple comparisons. Confidence intervals are based on z-scores.

Table 54: Contrasts Estimating Ratios of Marginal Means for Homicide Incident Counts and Incidence Rates By When Incidents Occurred

5.7.8 Larceny/Theft/Fraud

Among these 1,082 suspected sexual offenders, there were 627 (58%) offenders who had criminal histories containing at least one incident associated with an arrest, charge, or conviction for larceny, theft, or fraud. Table 55 summarizes the counts of unique larceny, theft, or fraud incidents overall and by when the incident occurred, while Table 56 shows the frequency distributions underlying those summaries.

```
# Table caption.
TCap <- paste("Offender-Level Larceny, Theft, or Fraud Incidents (Overall and",
              "By When Incident Occurred)")

# Summarize crime category counts (both overall and broken down by IWhen).
IDNEW %>%
  select(HXCat12_8, HXCat12_8_Before, HXCat12_8_During, HXCat12_8_After) %>%
  describe(., quant=c(.25, .50, .75)) %>%
  bind_cols(data.frame(Variable = TRowLabs), .) %>%
  select(Variable, n, mean, sd, min, max, skew, kurtosis, Q0.25, Q0.5, Q0.75) %>%
  kable(., format = "latex", booktabs = TRUE, digits = 2, row.names = FALSE,
        col.names = CNames, caption = TCap,
        linesep = c('', '', '', '\\addlinespace')) %>%
  column_spec(column = 1, width = "6 cm") %>%
  footnote(general = FN1, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
# Table caption.
TCap <- paste("Frequency Distributions for Number of Larceny, Theft, or Fraud",
              "Incidents for Which Offender Was Arrested, Charged, or",
              "Convicted (Overall and By When Incident Occurred)")

# Get frequency distributions.
IDNEW %>%
  select(OID, HXCat12_8, HXCat12_8_Before, HXCat12_8_During, HXCat12_8_After) %>%
  rename(All = HXCat12_8, Before = HXCat12_8_Before,
         During = HXCat12_8_During, After = HXCat12_8_After) %>%
  pivot_longer(cols = -OID, names_to = "When", values_to = "Value") %>%
  group_by(When, Value) %>%
  count() %>%
  arrange(Value) %>%
  pivot_wider(names_from = When, values_from = n, values_fill = 0) %>%
  mutate(All_p = 100*All/N_All,
```

Variable	N	Mean	SD	Min	Max	Skew	Kurtosis	Q25	Q50	Q75
Incident count (arrested, charged, or convicted)	1082	1.75	2.91	0	38	4.34	33.05	0	1	2
...Before testing window	1041	0.74	1.39	0	10	2.97	11.28	0	0	1
...During testing window	1052	0.04	0.22	0	2	5.18	28.61	0	0	0
...After testing window	1082	1.00	2.25	0	37	6.51	74.51	0	0	1

Note: Only offenders with valid testing window start dates are included. Only incidents where the offender was arrested for, charged with, or convicted of at least one offense from one of the 12 main crime categories were counted. Before, during, and after refer to when the incident associated with the record occurred, not when the arrest date, charge date, or conviction date occurred relative to the testing window.

Table 55: Offender-Level Larceny, Theft, or Fraud Incidents (Overall and By When Incident Occurred)

```
       All_v = if_else(is.na(Value),
                       true = as.numeric(NA),
                       false = 100*All/N_All),
       Before_p = 100*Before/N_All,
       Before_v = if_else(is.na(Value),
                          true = as.numeric(NA),
                          false = 100*Before/N_Before),
       During_p = 100*During/N_All,
       During_v = if_else(is.na(Value),
                          true = as.numeric(NA),
                          false = 100*During/N_During),
       After_p = 100*After/N_All,
       After_v = if_else(is.na(Value),
                         true = as.numeric(NA),
                         false = 100*After/N_After)) %>%
select(Value, All, All_p, All_v, Before, Before_p, Before_v, During,
       During_p, During_v, After, After_p, After_v) %>%
kable(format = "latex", booktabs = TRUE, digits = 2,
      col.names = c("Value", rep(c("N", "%", "Valid %"), 4)),
      caption = TCap) %>%
kable_styling() %>%
add_header_above(c(" " = 1, "Overall" = 3, "Before TW" = 3, "During TW" = 3,
                   "After TW" = 3)) %>%
footnote(general = FN5, general_title = "Note: ", footnote_as_chunk = TRUE,
       threeparttable = TRUE)
```

	Overall			Before TW			During TW			After TW		
Value	N	%	Valid %	N	%	Valid %	N	%	Valid %	N	%	Valid %
0	455	42.05	42.05	662	61.18	63.59	1008	93.16	95.82	662	61.18	61.18
1	237	21.90	21.90	203	18.76	19.50	41	3.79	3.90	199	18.39	18.39
2	151	13.96	13.96	86	7.95	8.26	3	0.28	0.29	97	8.96	8.96
3	78	7.21	7.21	42	3.88	4.03	0	0.00	0.00	43	3.97	3.97
4	45	4.16	4.16	19	1.76	1.83	0	0.00	0.00	26	2.40	2.40
5	39	3.60	3.60	7	0.65	0.67	0	0.00	0.00	16	1.48	1.48
6	16	1.48	1.48	10	0.92	0.96	0	0.00	0.00	9	0.83	0.83
7	17	1.57	1.57	5	0.46	0.48	0	0.00	0.00	9	0.83	0.83
8	11	1.02	1.02	2	0.18	0.19	0	0.00	0.00	5	0.46	0.46
9	6	0.55	0.55	3	0.28	0.29	0	0.00	0.00	5	0.46	0.46
10	5	0.46	0.46	2	0.18	0.19	0	0.00	0.00	3	0.28	0.28
11	8	0.74	0.74	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
12	2	0.18	0.18	0	0.00	0.00	0	0.00	0.00	2	0.18	0.18
13	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
14	3	0.28	0.28	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
16	3	0.28	0.28	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
17	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
20	3	0.28	0.28	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
23	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
26	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
37	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
38	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
NA	0	0.00	NA	41	3.79	NA	30	2.77	NA	0	0.00	NA

Note: N = 1,082. TW, earliest testing window. Only offenders with valid testing window start dates are included. Before, during, and after refer to when the incident associated with the record occurred, not when the arrest, charge, or conviction date occurred relative to the testing window. For example, an offender could be arrested during the testing window for an incident that occurred before it began. That would show up here in the before column. An incident is counted if there are any arrest, charge, or conviction records (any one type, or any combination of them will suffice) for the specified crime category associated with it.

Table 56: Frequency Distributions for Number of Larceny, Theft, or Fraud Incidents for Which Offender Was Arrested, Charged, or Convicted (Overall and By When Incident Occurred)

5.7.8.1 GEE Models To more closely examine the larceny, theft or fraud incident counts, we fitted two generalized estimating equation (GEE) models to the long-format *IDNEWL* data, after dropping rows for unobserved periods (*Years* = 0) because the counts are missing for those observations.

Model 1 used *When* as a predictor of the period-specific counts, but did not adjust for the period durations. Model 2 supplemented *When* as a predictor with an offset term for log(*Years*) to adjust for the period durations. Both models used Poisson distributions, an exchangeable correlation structure, a log link function, and robust standard errors based on the sandwich estimator.

```
HXCat12_8_gee1 <- IDNEWL %>%
  filter(Years > 0) %>%
  geeglm(HXCat12_8 ~ When,
         family = poisson(link = "log"), data = .,
         id = OID, corstr = "exchangeable", std.err = "san.se")

HXCat12_8_gee2 <- IDNEWL %>%
  filter(Years > 0) %>%
  geeglm(HXCat12_8 ~ When + offset(log(Years)),
         family = poisson(link = "log"), data = .,
         id = OID, corstr = "exchangeable", std.err = "san.se")
```

Table 57 below shows the parameter estimates for both GEE models.

```
TCap <- paste("Generalized Estimating Equation Models for Larceny, Theft, or",
              "Fraud Incident Counts",
              "Predicted by When Incidents Occurred")
FN <- paste ("\\item * p < .05, null hypothesis value outside 95\\% confidence",
             "interval based on z-score.",
             "\\\\\n\\item Note.",
             "The data contain up to 3 longitudinal observations (one per",
             "period) for each offender (i.e., cluster).",
             "The reference priod for When was after the testing window.",
             "Both models used a log link function, exchangeable correlation",
             "structure, a Poisson distribution, and robust standard errors",
             "(sandwich estimator).",
             "Model 1 did not use an offset.",
             "Model 2 used log(Years) as an offset term.")

texreg(list(HXCat12_8_gee1, HXCat12_8_gee2), booktabs = TRUE, dcolumn = TRUE,
       threeparttable = TRUE, fontsize = "normalsize", table = TRUE,
       use.packages = FALSE, ci.force = TRUE, label = "tab:HXCat12_8_texreg",
       stars = 0.05, caption = TCap, custom.note = FN)
```

	Model 1	Model 2
(Intercept)	−0.00	−2.90*
	[−0.14; 0.13]	[−3.03; −2.76]
WhenBefore	−0.29*	0.13
	[−0.45; −0.14]	[−0.03; 0.28]
WhenDuring	−3.09*	0.90*
	[−3.40; −2.79]	[0.60; 1.21]
Scale parameter: gamma	2.94	2.78
Scale parameter: SE	0.41	0.35
Correlation parameter: alpha	0.15	0.16
Correlation parameter: SE	0.03	0.03
Num. obs.	3175	3175
Num. clust.	1082	1082

* p < .05, null hypothesis value outside 95% confidence interval based on z-score.

Note. The data contain up to 3 longitudinal observations (one per period) for each offender (i.e., cluster). The reference priod for When was after the testing window. Both models used a log link function, exchangeable correlation structure, a Poisson distribution, and robust standard errors (sandwich estimator). Model 1 did not use an offset. Model 2 used log(Years) as an offset term.

Table 57: Generalized Estimating Equation Models for Larceny, Theft, or Fraud Incident Counts Predicted by When Incidents Occurred

5.7.8.2 Estimated Marginal Means Table 58 provides the estimated marginal means obtained from Models 1 and 2, along with corresponding confidence intervals. The units of measurement for those means are counts per offender for Model 1 (which ignore period duration) and counts per offender-year for Model 2 (which adjust for duration and are thus considered annual incidence rates).

```
# Table caption.
TCap <- paste("Marginal Means of Larceny, Theft, or Fraud Incident Counts and",
              "Incidence Rates By When Incidents Occurred")

# Table footnote.
FN <- paste("Model 1 did not use an offset, so means are counts.",
```

```
            "Model 2 used log(Years) as an offset term, so means are annual",
            "incidence rates.",
            "Confidence intervals are based on z-scores.")

HXCat12_8_emmeans1 <- emmeans(HXCat12_8_gee1, specs = pairwise ~ When,
                              type = "response", ref = "During")
HXCat12_8_emmeans2 <- emmeans(HXCat12_8_gee2, specs = pairwise ~ When,
                              type = "response", ref = "During")

as_tibble(HXCat12_8_emmeans1$emmeans) %>%
  full_join(x = ., y = as_tibble(HXCat12_8_emmeans2$emmeans)) %>%
  mutate(Model = c(1, 1, 1, 2, 2, 2),
         When = factor(When, levels = c("Before", "During", "After"))) %>%
  arrange(Model, When) %>%
  select(When, rate, SE, df, asymp.LCL, asymp.UCL) %>%
  kable(format = "latex", booktabs = TRUE, caption = TCap, digits = 2,
        col.names = c("When", "Rate", "SE", "df", "LCL", "UCL")) %>%
  kable_styling() %>%
  add_header_above(c(" " = 4, "95% CI" = 2)) %>%
  group_rows(group_label = "Model 1: Mean per offender", start_row = 1,
             end_row = 3, italic = TRUE) %>%
  group_rows(group_label = "Model 2: Mean per offender-year",
             start_row = 4, end_row = 6, italic = TRUE) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
## Joining, by = c("When", "rate", "SE", "df", "asymp.LCL", "asymp.UCL")
```

				95% CI	
When	Rate	SE	df	LCL	UCL
Model 1: Mean per offender					
Before	0.74	0.04	Inf	0.66	0.83
During	0.05	0.01	Inf	0.03	0.06
After	1.00	0.07	Inf	0.87	1.14
Model 2: Mean per offender-year					
Before	0.63	0.03	Inf	0.57	0.70
During	1.38	0.21	Inf	1.02	1.86
After	0.56	0.04	Inf	0.49	0.64

Note: Model 1 did not use an offset, so means are counts. Model 2 used log(Years) as an offset term, so means are annual incidence rates. Confidence intervals are based on z-scores.

Table 58: Marginal Means of Larceny, Theft, or Fraud Incident Counts and Incidence Rates By When Incidents Occurred

5.7.8.3 Contrasts for Ratios of Estimated Marginal Means Finally, we also estimated a set of contrasts to compare the marginal means from Models 1 and 2 (Table 59). Each of these contrasts estimates the ratios of a pair of means rather than a raw difference between means. This is because of the nature of the underlying Poisson GEE model.

```
TCap <- paste("Contrasts Estimating Ratios of Marginal Means for Larceny,",
              "Theft, or Fraud Incident Counts and Incidence Rates By When",
              "Incidents Occurred")

# Table footnote.
FN <- paste("Model 1 did not use an offset, so means are counts.",
            "Model 2 used log(Years) as an offset term, so means are annual",
            "incidence rates.",
            "These contrasts estimate ratios of those quantities and used",
            "Tukey's method to adjust for multiple comparisons.",
```

```
           "Confidence intervals are based on z-scores.")

as_tibble(confint(HXCat12_8_emmeans1$contrasts)) %>%
  full_join(x = ., y = as_tibble(confint(HXCat12_8_emmeans2$contrasts))) %>%
  kable(format = "latex", booktabs = TRUE, caption = TCap, digits = 2,
        col.names = c("Contrast", "Ratio", "SE", "df", "LCL", "UCL")) %>%
  kable_styling() %>%
  add_header_above(c(" " = 4, "95% CI" = 2)) %>%
  group_rows(group_label = "Model 1: Ratios of means per offender",
             start_row = 1, end_row = 3, italic = TRUE) %>%
  group_rows(group_label = "Model 2: Ratios of means per offender-year",
             start_row = 4, end_row = 6, italic = TRUE) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
## Joining, by = c("contrast", "ratio", "SE", "df", "asymp.LCL", "asymp.UCL")
```

				95% CI	
Contrast	Ratio	SE	df	LCL	UCL
Model 1: Ratios of means per offender					
After / Before	1.34	0.11	Inf	1.11	1.62
After / During	22.06	3.40	Inf	15.38	31.65
Before / During	16.43	2.54	Inf	11.44	23.59
Model 2: Ratios of means per offender-year					
After / Before	0.88	0.07	Inf	0.73	1.06
After / During	0.41	0.06	Inf	0.28	0.59
Before / During	0.46	0.07	Inf	0.32	0.66

Note:
Model 1 did not use an offset, so means are counts. Model 2 used log(Years) as an offset term, so means are annual incidence rates. These contrasts estimate ratios of those quantities and used Tukey's method to adjust for multiple comparisons. Confidence intervals are based on z-scores.

Table 59: Contrasts Estimating Ratios of Marginal Means for Larceny, Theft, or Fraud Incident Counts and Incidence Rates By When Incidents Occurred

5.7.9 Robbery

Among these 1,082 suspected sexual offenders, there were 399 (37%) offenders who had criminal histories containing at least one incident associated with an arrest, charge, or conviction for robbery. Table 60 summarizes the counts of unique robbery incidents overall and by when the incident occurred, while Table 61 shows the frequency distributions underlying those summaries.

```
# Table caption.
TCap <- paste("Offender-Level Robbery Incident Counts (Overall and By When",
              "Incident Occurred)")

# Summarize crime category counts (both overall and broken down by IWhen).
IDNEW %>%
  select(HXCat12_9, HXCat12_9_Before, HXCat12_9_During, HXCat12_9_After) %>%
  describe(., quant=c(.25, .50, .75)) %>%
  bind_cols(data.frame(Variable = TRowLabs), .) %>%
  select(Variable, n, mean, sd, min, max, skew, kurtosis, Q0.25, Q0.5, Q0.75) %>%
  kable(., format = "latex", booktabs = TRUE, digits = 2, row.names = FALSE,
        col.names = CNames, caption = TCap,
        linesep = c('', '', '', '\\addlinespace')) %>%
  column_spec(column = 1, width = "6 cm") %>%
```

```
  footnote(general = FN1, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

Variable	N	Mean	SD	Min	Max	Skew	Kurtosis	Q25	Q50	Q75
Incident count (arrested, charged, or convicted)	1082	0.59	0.99	0	6	2.30	6.57	0	0	1
...Before testing window	1041	0.24	0.56	0	5	2.93	11.83	0	0	0
...During testing window	1052	0.05	0.25	0	2	5.10	27.91	0	0	0
...After testing window	1082	0.31	0.72	0	6	3.37	15.00	0	0	0

Note: Only offenders with valid testing window start dates are included. Only incidents where the offender was arrested for, charged with, or convicted of at least one offense from one of the 12 main crime categories were counted. Before, during, and after refer to when the incident associated with the record occurred, not when the arrest date, charge date, or conviction date occurred relative to the testing window.

Table 60: Offender-Level Robbery Incident Counts (Overall and By When Incident Occurred)

```
# Table caption.
TCap <- paste("Frequency Distributions for Number of Robbery Incidents for",
              "Which Offender Was Arrested, Charged, or Convicted (Overall and",
              "By When Incident Occurred)")

# Get frequency distributions.
IDNEW %>%
  select(OID, HXCat12_9, HXCat12_9_Before, HXCat12_9_During, HXCat12_9_After) %>%
  rename(All = HXCat12_9, Before = HXCat12_9_Before, During = HXCat12_9_During,
         After = HXCat12_9_After) %>%
  pivot_longer(cols = -OID, names_to = "When", values_to = "Value") %>%
  group_by(When, Value) %>%
  count() %>%
  arrange(Value) %>%
  pivot_wider(names_from = When, values_from = n, values_fill = 0) %>%
  mutate(All_p = 100*All/N_All,
         All_v = if_else(is.na(Value),
                         true = as.numeric(NA),
                         false = 100*All/N_All),
         Before_p = 100*Before/N_All,
         Before_v = if_else(is.na(Value),
                            true = as.numeric(NA),
                            false = 100*Before/N_Before),
         During_p = 100*During/N_All,
         During_v = if_else(is.na(Value),
                            true = as.numeric(NA),
                            false = 100*During/N_During),
         After_p = 100*After/N_All,
         After_v = if_else(is.na(Value),
                           true = as.numeric(NA),
                           false = 100*After/N_After)) %>%
  select(Value, All, All_p, All_v, Before, Before_p, Before_v, During,
         During_p, During_v, After, After_p, After_v) %>%
  kable(format = "latex", booktabs = TRUE, digits = 2,
        col.names = c("Value", rep(c("N", "%", "Valid %"), 4)),
        caption = TCap) %>%
  kable_styling() %>%
  add_header_above(c(" " = 1, "Overall" = 3, "Before TW" = 3, "During TW" = 3,
                     "After TW" = 3)) %>%
  footnote(general = FN5, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

	Overall			Before TW			During TW			After TW		
Value	N	%	Valid %	N	%	Valid %	N	%	Valid %	N	%	Valid %
0	683	63.12	63.12	839	77.54	80.60	1003	92.70	95.34	854	78.93	78.93
1	251	23.20	23.20	162	14.97	15.56	43	3.97	4.09	165	15.25	15.25
2	99	9.15	9.15	33	3.05	3.17	6	0.55	0.57	39	3.60	3.60
3	23	2.13	2.13	4	0.37	0.38	0	0.00	0.00	14	1.29	1.29
4	13	1.20	1.20	2	0.18	0.19	0	0.00	0.00	6	0.55	0.55
5	9	0.83	0.83	1	0.09	0.10	0	0.00	0.00	2	0.18	0.18
6	4	0.37	0.37	0	0.00	0.00	0	0.00	0.00	2	0.18	0.18
NA	0	0.00	NA	41	3.79	NA	30	2.77	NA	0	0.00	NA

Note: N = 1,082. TW, earliest testing window. Only offenders with valid testing window start dates are included. Before, during, and after refer to when the incident associated with the record occurred, not when the arrest, charge, or conviction date occurred relative to the testing window. For example, an offender could be arrested during the testing window for an incident that occurred before it began. That would show up here in the before column. An incident is counted if there are any arrest, charge, or conviction records (any one type, or any combination of them will suffice) for the specified crime category associated with it.

Table 61: Frequency Distributions for Number of Robbery Incidents for Which Offender Was Arrested, Charged, or Convicted (Overall and By When Incident Occurred)

5.7.9.1 GEE Models To more closely examine the robbery incident counts, we fitted two generalized estimating equation (GEE) models to the long-format *IDNEWL* data, after dropping rows for unobserved periods (*Years* = 0) because the counts are missing for those observations.

Model 1 used *When* as a predictor of the period-specific counts, but did not adjust for the period durations. Model 2 supplemented *When* as a predictor with an offset term for log(*Years*) to adjust for the period durations. Both models used Poisson distributions, an exchangeable correlation structure, a log link function, and robust standard errors based on the sandwich estimator.

```
HXCat12_9_gee1 <- IDNEWL %>%
  filter(Years > 0) %>%
  geeglm(HXCat12_9 ~ When,
         family = poisson(link = "log"), data = .,
         id = OID, corstr = "exchangeable", std.err = "san.se")

HXCat12_9_gee2 <- IDNEWL %>%
  filter(Years > 0) %>%
  geeglm(HXCat12_9 ~ When + offset(log(Years)),
         family = poisson(link = "log"), data = .,
         id = OID, corstr = "exchangeable", std.err = "san.se")
```

Table 62 below shows the parameter estimates for both GEE models.

```
TCap <- paste("Generalized Estimating Equation Models for Robbery Incident Counts",
              "Predicted by When Incidents Occurred")
FN <- paste ("\\item * p < .05, null hypothesis value outside 95\\% confidence",
             "interval based on z-score.",
             "\\\\\\n\\item Note.",
             "The data contain up to 3 longitudinal observations (one per",
             "period) for each offender (i.e., cluster).",
             "The reference priod for When was after the testing window.",
             "Both models used a log link function, exchangeable correlation",
             "structure, a Poisson distribution, and robust standard errors",
             "(sandwich estimator).",
             "Model 1 did not use an offset.",
             "Model 2 used log(Years) as an offset term.")
```

```
texreg(list(HXCat12_9_gee1, HXCat12_9_gee2), booktabs = TRUE, dcolumn = TRUE,
       threeparttable = TRUE, fontsize = "normalsize", table = TRUE,
       use.packages = FALSE, ci.force = TRUE, label = "tab:HXCat12_9_texreg",
       stars = 0.05, caption = TCap, custom.note = FN)
```

	Model 1	Model 2
(Intercept)	−1.18*	−4.08*
	[−1.33; −1.04]	[−4.22; −3.94]
WhenBefore	−0.23*	0.20*
	[−0.42; −0.04]	[0.02; 0.39]
WhenDuring	−1.76*	2.25*
	[−2.08; −1.45]	[1.93; 2.57]
Scale parameter: gamma	1.39	1.38
Scale parameter: SE	0.20	0.21
Correlation parameter: alpha	0.06	0.08
Correlation parameter: SE	0.02	0.02
Num. obs.	3175	3175
Num. clust.	1082	1082

* p < .05, null hypothesis value outside 95% confidence interval based on z-score.

Note. The data contain up to 3 longitudinal observations (one per period) for each offender (i.e., cluster). The reference priod for When was after the testing window. Both models used a log link function, exchangeable correlation structure, a Poisson distribution, and robust standard errors (sandwich estimator). Model 1 did not use an offset. Model 2 used log(Years) as an offset term.

Table 62: Generalized Estimating Equation Models for Robbery Incident Counts Predicted by When Incidents Occurred

5.7.9.2 Estimated Marginal Means Table 63 provides the estimated marginal means obtained from Models 1 and 2, along with corresponding confidence intervals. The units of measurement for those means are counts per offender for Model 1 (which ignore period duration) and counts per offender-year for Model 2 (which adjust for duration and are thus considered annual incidence rates).

```
# Table caption.
TCap <- paste("Marginal Means of Robbery Incident Counts and Incidence Rates",
              "By When Incidents Occurred")

# Table footnote.
FN <- paste("Model 1 did not use an offset, so means are counts.",
            "Model 2 used log(Years) as an offset term, so means are annual",
            "incidence rates.",
            "Confidence intervals are based on z-scores.")

HXCat12_9_emmeans1 <- emmeans(HXCat12_9_gee1, specs = pairwise ~ When,
                              type = "response", ref = "During")
HXCat12_9_emmeans2 <- emmeans(HXCat12_9_gee2, specs = pairwise ~ When,
                              type = "response", ref = "During")

as_tibble(HXCat12_9_emmeans1$emmeans) %>%
  full_join(x = ., y = as_tibble(HXCat12_9_emmeans2$emmeans)) %>%
  mutate(Model = c(1, 1, 1, 2, 2, 2),
         When = factor(When, levels = c("Before", "During", "After"))) %>%
  arrange(Model, When) %>%
  select(When, rate, SE, df, asymp.LCL, asymp.UCL) %>%
  kable(format = "latex", booktabs = TRUE, caption = TCap, digits = 2,
        col.names = c("When", "Rate", "SE", "df", "LCL", "UCL")) %>%
  kable_styling() %>%
  add_header_above(c(" " = 4, "95% CI" = 2)) %>%
  group_rows(group_label = "Model 1: Mean per offender", start_row = 1,
             end_row = 3, italic = TRUE) %>%
```

```
  group_rows(group_label = "Model 2: Mean per offender-year",
             start_row = 4, end_row = 6, italic = TRUE) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
## Joining, by = c("When", "rate", "SE", "df", "asymp.LCL", "asymp.UCL")
```

				95% CI	
When	Rate	SE	df	LCL	UCL
Model 1: Mean per offender					
Before	0.24	0.02	Inf	0.21	0.28
During	0.05	0.01	Inf	0.04	0.07
After	0.31	0.02	Inf	0.27	0.35
Model 2: Mean per offender-year					
Before	0.21	0.01	Inf	0.18	0.24
During	1.63	0.24	Inf	1.22	2.16
After	0.17	0.01	Inf	0.15	0.20

Note: Model 1 did not use an offset, so means are counts. Model 2 used log(Years) as an offset term, so means are annual incidence rates. Confidence intervals are based on z-scores.

Table 63: Marginal Means of Robbery Incident Counts and Incidence Rates By When Incidents Occurred

5.7.9.3 Contrasts for Ratios of Estimated Marginal Means Finally, we also estimated a set of contrasts to compare the marginal means from Models 1 and 2 (Table 64). Each of these contrasts estimates the ratios of a pair of means rather than a raw difference between means. This is because of the nature of the underlying Poisson GEE model.

```
TCap <- paste("Contrasts Estimating Ratios of Marginal Means for Robbery",
              "Incident Counts and Incidence Rates By When Incidents",
              "Occurred")

# Table footnote.
FN <- paste("Model 1 did not use an offset, so means are counts.",
            "Model 2 used log(Years) as an offset term, so means are annual",
            "incidence rates.",
            "These contrasts estimate ratios of those quantities and used",
            "Tukey's method to adjust for multiple comparisons.",
            "Confidence intervals are based on z-scores.")

as_tibble(confint(HXCat12_9_emmeans1$contrasts)) %>%
  full_join(x = ., y = as_tibble(confint(HXCat12_9_emmeans2$contrasts))) %>%
  kable(format = "latex", booktabs = TRUE, caption = TCap, digits = 2,
        col.names = c("Contrast", "Ratio", "SE", "df", "LCL", "UCL")) %>%
  kable_styling() %>%
  add_header_above(c(" " = 4, "95% CI" = 2)) %>%
  group_rows(group_label = "Model 1: Ratios of means per offender",
             start_row = 1, end_row = 3, italic = TRUE) %>%
  group_rows(group_label = "Model 2: Ratios of means per offender-year",
             start_row = 4, end_row = 6, italic = TRUE) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
## Joining, by = c("contrast", "ratio", "SE", "df", "asymp.LCL", "asymp.UCL")
```

				95% CI	
Contrast	Ratio	SE	df	LCL	UCL
Model 1: Ratios of means per offender					
After / Before	1.26	0.12	Inf	1.00	1.58
After / During	5.84	0.94	Inf	4.00	8.52
Before / During	4.64	0.72	Inf	3.23	6.67
Model 2: Ratios of means per offender-year					
After / Before	0.81	0.08	Inf	0.65	1.02
After / During	0.11	0.02	Inf	0.07	0.15
Before / During	0.13	0.02	Inf	0.09	0.19

Note:
Model 1 did not use an offset, so means are counts. Model 2 used log(Years) as an offset term, so means are annual incidence rates. These contrasts estimate ratios of those quantities and used Tukey's method to adjust for multiple comparisons. Confidence intervals are based on z-scores.

Table 64: Contrasts Estimating Ratios of Marginal Means for Robbery Incident Counts and Incidence Rates By When Incidents Occurred

5.7.10 Sex Crimes

Among these 1,082 suspected sexual offenders, there were 91 (8%) offenders who had criminal histories containing at least one incident associated with an arrest, charge, or conviction for sex crimes. Table 65 summarizes the counts of unique sex crime incidents overall and by when the incident occurred, while Table 66 shows the frequency distributions underlying those summaries.

```
# Table caption.
TCap <- paste("Offender-Level Sex Crime Incident Counts (Overall and By When",
              "Incident Occurred)")

# Summarize crime category counts (both overall and broken down by IWhen).
IDNEW %>%
  select(HXCat12_10, HXCat12_10_Before, HXCat12_10_During, HXCat12_10_After) %>%
  describe(., quant=c(.25, .50, .75)) %>%
  bind_cols(data.frame(Variable = TRowLabs), .) %>%
  select(Variable, n, mean, sd, min, max, skew, kurtosis, Q0.25, Q0.5, Q0.75) %>%
  kable(., format = "latex", booktabs = TRUE, digits = 2, row.names = FALSE,
        col.names = CNames, caption = TCap,
        linesep = c('', '', '', '\\addlinespace')) %>%
  column_spec(column = 1, width = "6 cm") %>%
  footnote(general = FN1, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
# Table caption.
TCap <- paste("Frequency Distributions for Number of Sex Crime Incidents for",
              "Which Offender Was Arrested, Charged, or Convicted (Overall and",
              "By When Incident Occurred)")

# Get frequency distributions.
IDNEW %>%
  select(OID, HXCat12_10, HXCat12_10_Before, HXCat12_10_During, HXCat12_10_After) %>%
  rename(All = HXCat12_10, Before = HXCat12_10_Before, During = HXCat12_10_During,
         After = HXCat12_10_After) %>%
  pivot_longer(cols = -OID, names_to = "When", values_to = "Value") %>%
  group_by(When, Value) %>%
  count() %>%
  arrange(Value) %>%
  pivot_wider(names_from = When, values_from = n, values_fill = 0) %>%
  mutate(All_p = 100*All/N_All,
```

Variable	N	Mean	SD	Min	Max	Skew	Kurtosis	Q25	Q50	Q75
Incident count (arrested, charged, or convicted)	1082	0.11	0.40	0	5	5.07	35.24	0	0	0
...Before testing window	1041	0.01	0.12	0	2	10.40	120.01	0	0	0
...During testing window	1052	0.01	0.10	0	1	9.61	90.47	0	0	0
...After testing window	1082	0.09	0.35	0	5	5.88	49.72	0	0	0

Note: Only offenders with valid testing window start dates are included. Only incidents where the offender was arrested for, charged with, or convicted of at least one offense from one of the 12 main crime categories were counted. Before, during, and after refer to when the incident associated with the record occurred, not when the arrest date, charge date, or conviction date occurred relative to the testing window.

Table 65: Offender-Level Sex Crime Incident Counts (Overall and By When Incident Occurred)

```
       All_v = if_else(is.na(Value),
                       true = as.numeric(NA),
                       false = 100*All/N_All),
       Before_p = 100*Before/N_All,
       Before_v = if_else(is.na(Value),
                          true = as.numeric(NA),
                          false = 100*Before/N_Before),
       During_p = 100*During/N_All,
       During_v = if_else(is.na(Value),
                          true = as.numeric(NA),
                          false = 100*During/N_During),
       After_p = 100*After/N_All,
       After_v = if_else(is.na(Value),
                         true = as.numeric(NA),
                         false = 100*After/N_After)) %>%
select(Value, All, All_p, All_v, Before, Before_p, Before_v, During,
       During_p, During_v, After, After_p, After_v) %>%
kable(format = "latex", booktabs = TRUE, digits = 2,
      col.names = c("Value", rep(c("N", "%", "Valid %"), 4)),
      caption = TCap) %>%
kable_styling() %>%
add_header_above(c(" " = 1, "Overall" = 3, "Before TW" = 3, "During TW" = 3,
                   "After TW" = 3)) %>%
footnote(general = FN5, general_title = "Note: ", footnote_as_chunk = TRUE,
       threeparttable = TRUE)
```

	Overall			Before TW			During TW			After TW		
Value	N	%	Valid %	N	%	Valid %	N	%	Valid %	N	%	Valid %
0	991	91.59	91.59	1029	95.10	98.85	1041	96.21	98.95	1008	93.16	93.16
1	73	6.75	6.75	11	1.02	1.06	11	1.02	1.05	61	5.64	5.64
2	13	1.20	1.20	1	0.09	0.10	0	0.00	0.00	10	0.92	0.92
3	4	0.37	0.37	0	0.00	0.00	0	0.00	0.00	2	0.18	0.18
5	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
NA	0	0.00	NA	41	3.79	NA	30	2.77	NA	0	0.00	NA

Note: N = 1,082. TW, earliest testing window. Only offenders with valid testing window start dates are included. Before, during, and after refer to when the incident associated with the record occurred, not when the arrest, charge, or conviction date occurred relative to the testing window. For example, an offender could be arrested during the testing window for an incident that occurred before it began. That would show up here in the before column. An incident is counted if there are any arrest, charge, or conviction records (any one type, or any combination of them will suffice) for the specified crime category associated with it.

Table 66: Frequency Distributions for Number of Sex Crime Incidents for Which Offender Was Arrested, Charged, or Convicted (Overall and By When Incident Occurred)

5.7.10.1 GEE Models To more closely examine the sex crime incident counts, we fitted two generalized estimating equation (GEE) models to the long-format *IDNEWL* data, after dropping rows for unobserved periods ($Years = 0$) because the counts are missing for those observations.

Model 1 used *When* as a predictor of the period-specific counts, but did not adjust for the period durations. Model 2 supplemented *When* as a predictor with an offset term for log(*Years*) to adjust for the period durations. Both models used Poisson distributions, an exchangeable correlation structure, a log link function, and robust standard errors based on the sandwich estimator.

```
HXCat12_10_gee1 <- IDNEWL %>%
  filter(Years > 0) %>%
  geeglm(HXCat12_10 ~ When,
         family = poisson(link = "log"), data = .,
         id = OID, corstr = "exchangeable", std.err = "san.se")

HXCat12_10_gee2 <- IDNEWL %>%
  filter(Years > 0) %>%
  geeglm(HXCat12_10 ~ When + offset(log(Years)),
         family = poisson(link = "log"), data = .,
         id = OID, corstr = "exchangeable", std.err = "san.se")
```

Table 67 below shows the parameter estimates for both GEE models.

```
TCap <- paste("Generalized Estimating Equation Models for Sex Crime Incident",
              "Counts Predicted by When Incidents Occurred")
FN <- paste ("\\item * p < .05, null hypothesis value outside 95\\% confidence",
             "interval based on z-score.",
             "\\\\\n\\item Note.",
             "The data contain up to 3 longitudinal observations (one per",
             "period) for each offender (i.e., cluster).",
             "The reference priod for When was after the testing window.",
             "Both models used a log link function, exchangeable correlation",
             "structure, a Poisson distribution, and robust standard errors",
             "(sandwich estimator).",
             "Model 1 did not use an offset.",
             "Model 2 used log(Years) as an offset term.")

texreg(list(HXCat12_10_gee1, HXCat12_10_gee2), booktabs = TRUE, dcolumn = TRUE,
       threeparttable = TRUE, fontsize = "normalsize", table = TRUE,
       use.packages = FALSE, ci.force = TRUE, label = "tab:HXCat12_10_texreg",
       stars = 0.05, caption = TCap, custom.note = FN)
```

	Model 1	Model 2
(Intercept)	−2.46*	−5.36*
	[−2.71; −2.22]	[−5.60; −5.11]
WhenBefore	−1.92*	−1.49*
	[−2.54; −1.30]	[−2.11; −0.88]
WhenDuring	−2.10*	1.92*
	[−2.72; −1.48]	[1.30; 2.54]
Scale parameter: gamma	1.21	1.15
Scale parameter: SE	1.25	1.24
Correlation parameter: alpha	0.05	0.05
Correlation parameter: SE	0.06	0.06
Num. obs.	3175	3175
Num. clust.	1082	1082

* p < .05, null hypothesis value outside 95% confidence interval based on z-score.

Note. The data contain up to 3 longitudinal observations (one per period) for each offender (i.e., cluster). The reference priod for When was after the testing window. Both models used a log link function, exchangeable correlation structure, a Poisson distribution, and robust standard errors (sandwich estimator). Model 1 did not use an offset. Model 2 used log(Years) as an offset term.

Table 67: Generalized Estimating Equation Models for Sex Crime Incident Counts Predicted by When Incidents Occurred

5.7.10.2 Estimated Marginal Means Table 68 provides the estimated marginal means obtained from Models 1 and 2, along with corresponding confidence intervals. The units of measurement for those means are counts per offender for Model 1 (which ignore period duration) and counts per offender-year for Model 2 (which adjust for duration and are thus considered annual incidence rates).

```
# Table caption.
TCap <- paste("Marginal Means of Sex Crime Incident Counts and Incidence Rates",
              "By When Incidents Occurred")

# Table footnote.
FN <- paste("Model 1 did not use an offset, so means are counts.",
            "Model 2 used log(Years) as an offset term, so means are annual",
            "incidence rates.",
            "Confidence intervals are based on z-scores.")

HXCat12_10_emmeans1 <- emmeans(HXCat12_10_gee1, specs = pairwise ~ When,
                           type = "response", ref = "During")
HXCat12_10_emmeans2 <- emmeans(HXCat12_10_gee2, specs = pairwise ~ When,
                           type = "response", ref = "During")

as_tibble(HXCat12_10_emmeans1$emmeans) %>%
  full_join(x = ., y = as_tibble(HXCat12_10_emmeans2$emmeans)) %>%
  mutate(Model = c(1, 1, 1, 2, 2, 2),
         When = factor(When, levels = c("Before", "During", "After"))) %>%
  arrange(Model, When) %>%
  select(When, rate, SE, df, asymp.LCL, asymp.UCL) %>%
  kable(format = "latex", booktabs = TRUE, caption = TCap, digits = 2,
        col.names = c("When", "Rate", "SE", "df", "LCL", "UCL")) %>%
  kable_styling() %>%
  add_header_above(c(" " = 4, "95% CI" = 2)) %>%
  group_rows(group_label = "Model 1: Mean per offender", start_row = 1,
             end_row = 3, italic = TRUE) %>%
  group_rows(group_label = "Model 2: Mean per offender-year",
             start_row = 4, end_row = 6, italic = TRUE) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
## Joining, by = c("When", "rate", "SE", "df", "asymp.LCL", "asymp.UCL")
```

				95% CI	
When	Rate	SE	df	LCL	UCL
Model 1: Mean per offender					
Before	0.01	0.00	Inf	0.01	0.02
During	0.01	0.00	Inf	0.01	0.02
After	0.09	0.01	Inf	0.07	0.11
Model 2: Mean per offender-year					
Before	0.01	0.00	Inf	0.01	0.02
During	0.33	0.10	Inf	0.18	0.59
After	0.05	0.01	Inf	0.04	0.06

Note: Model 1 did not use an offset, so means are counts. Model 2 used log(Years) as an offset term, so means are annual incidence rates. Confidence intervals are based on z-scores.

Table 68: Marginal Means of Sex Crime Incident Counts and Incidence Rates By When Incidents Occurred

5.7.10.3 Contrasts for Ratios of Estimated Marginal Means Finally, we also estimated a set of contrasts to compare the marginal means from Models 1 and 2 (Table 69). Each of these contrasts estimates the ratios of a pair of means rather than a raw difference between means. This is because of the nature of the underlying Poisson GEE model.

```
TCap <- paste("Contrasts Estimating Ratios of Marginal Means for Sex Crime",
              "Incident Counts and Incidence Rates By When Incidents",
              "Occurred")

# Table footnote.
FN <- paste("Model 1 did not use an offset, so means are counts.",
            "Model 2 used log(Years) as an offset term, so means are annual",
            "incidence rates.",
            "These contrasts estimate ratios of those quantities and used",
            "Tukey's method to adjust for multiple comparisons.",
            "Confidence intervals are based on z-scores.")

as_tibble(confint(HXCat12_10_emmeans1$contrasts)) %>%
  full_join(x = ., y = as_tibble(confint(HXCat12_10_emmeans2$contrasts))) %>%
  kable(format = "latex", booktabs = TRUE, caption = TCap, digits = 2,
        col.names = c("Contrast", "Ratio", "SE", "df", "LCL", "UCL")) %>%
  kable_styling() %>%
  add_header_above(c(" " = 4, "95% CI" = 2)) %>%
  group_rows(group_label = "Model 1: Ratios of means per offender",
             start_row = 1, end_row = 3, italic = TRUE) %>%
  group_rows(group_label = "Model 2: Ratios of means per offender-year",
             start_row = 4, end_row = 6, italic = TRUE) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
## Joining, by = c("contrast", "ratio", "SE", "df", "asymp.LCL", "asymp.UCL")
```

Contrast	Ratio	SE	df	95% CI LCL	95% CI UCL
Model 1: Ratios of means per offender					
After / Before	6.83	2.15	Inf	3.26	14.30
After / During	8.15	2.58	Inf	3.88	17.13
Before / During	1.19	0.51	Inf	0.44	3.23
Model 2: Ratios of means per offender-year					
After / Before	4.45	1.39	Inf	2.14	9.27
After / During	0.15	0.05	Inf	0.07	0.31
Before / During	0.03	0.01	Inf	0.01	0.09

Note:
Model 1 did not use an offset, so means are counts. Model 2 used log(Years) as an offset term, so means are annual incidence rates. These contrasts estimate ratios of those quantities and used Tukey's method to adjust for multiple comparisons. Confidence intervals are based on z-scores.

Table 69: Contrasts Estimating Ratios of Marginal Means for Sex Crime Incident Counts and Incidence Rates By When Incidents Occurred

5.7.11 Traffic and Ordinances

Among these 1,082 suspected sexual offenders, there were 757 (70%) offenders who had criminal histories containing at least one incident associated with an arrest, charge, or conviction for traffic and ordinance violations. Table 70 summarizes the counts of unique traffic and ordinance incidents overall and by when the incident occurred, while Table 71 shows the frequency distributions underlying those summaries.

```
# Table caption.
TCap <- paste("Offender-Level Traffic and Ordinance Incidents (Overall and",
              "By When Incident Occurred)")

# Summarize crime category counts (both overall and broken down by IWhen).
IDNEW %>%
  select(HXCat12_11, HXCat12_11_Before, HXCat12_11_During, HXCat12_11_After) %>%
  describe(., quant=c(.25, .50, .75)) %>%
  bind_cols(data.frame(Variable = TRowLabs), .) %>%
  select(Variable, n, mean, sd, min, max, skew, kurtosis, Q0.25, Q0.5, Q0.75) %>%
  kable(., format = "latex", booktabs = TRUE, digits = 2, row.names = FALSE,
        col.names = CNames, caption = TCap,
        linesep = c('', '', '', '\\addlinespace')) %>%
  column_spec(column = 1, width = "6 cm") %>%
  footnote(general = FN1, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
# Table caption.
TCap <- paste("Frequency Distributions for Number of Traffic and Ordinance",
              "Incidents for Which Offender Was Arrested, Charged, or",
              "Convicted (Overall and By When Incident Occurred)")

# Get frequency distributions.
IDNEW %>%
  select(OID, HXCat12_11, HXCat12_11_Before, HXCat12_11_During, HXCat12_11_After) %>%
  rename(All = HXCat12_11, Before = HXCat12_11_Before,
         During = HXCat12_11_During, After = HXCat12_11_After) %>%
  pivot_longer(cols = -OID, names_to = "When", values_to = "Value") %>%
  group_by(When, Value) %>%
  count() %>%
  arrange(Value) %>%
  pivot_wider(names_from = When, values_from = n, values_fill = 0) %>%
  mutate(All_p = 100*All/N_All,
```

Variable	N	Mean	SD	Min	Max	Skew	Kurtosis	Q25	Q50	Q75
Incident count (arrested, charged, or convicted)	1082	2.28	2.64	0	28	2.37	11.62	0	2	3
...Before testing window	1041	0.39	0.86	0	8	3.30	15.08	0	0	0
...During testing window	1052	0.03	0.20	0	2	6.19	41.79	0	0	0
...After testing window	1082	1.87	2.46	0	28	2.65	14.74	0	1	3

Note: Only offenders with valid testing window start dates are included. Only incidents where the offender was arrested for, charged with, or convicted of at least one offense from one of the 12 main crime categories were counted. Before, during, and after refer to when the incident associated with the record occurred, not when the arrest date, charge date, or conviction date occurred relative to the testing window.

Table 70: Offender-Level Traffic and Ordinance Incidents (Overall and By When Incident Occurred)

```
       All_v = if_else(is.na(Value),
                       true = as.numeric(NA),
                       false = 100*All/N_All),
       Before_p = 100*Before/N_All,
       Before_v = if_else(is.na(Value),
                          true = as.numeric(NA),
                          false = 100*Before/N_Before),
       During_p = 100*During/N_All,
       During_v = if_else(is.na(Value),
                          true = as.numeric(NA),
                          false = 100*During/N_During),
       After_p = 100*After/N_All,
       After_v = if_else(is.na(Value),
                         true = as.numeric(NA),
                         false = 100*After/N_After)) %>%
select(Value, All, All_p, All_v, Before, Before_p, Before_v, During,
       During_p, During_v, After, After_p, After_v) %>%
kable(format = "latex", booktabs = TRUE, digits = 2,
      col.names = c("Value", rep(c("N", "%", "Valid %"), 4)),
      caption = TCap) %>%
kable_styling() %>%
add_header_above(c(" " = 1, "Overall" = 3, "Before TW" = 3, "During TW" = 3,
                   "After TW" = 3)) %>%
footnote(general = FN5, general_title = "Note: ", footnote_as_chunk = TRUE,
       threeparttable = TRUE)
```

	Overall			Before TW			During TW			After TW		
Value	N	%	Valid %	N	%	Valid %	N	%	Valid %	N	%	Valid %
0	325	30.04	30.04	789	72.92	75.79	1019	94.18	96.86	412	38.08	38.08
1	199	18.39	18.39	161	14.88	15.47	30	2.77	2.85	215	19.87	19.87
2	169	15.62	15.62	59	5.45	5.67	3	0.28	0.29	148	13.68	13.68
3	144	13.31	13.31	17	1.57	1.63	0	0.00	0.00	105	9.70	9.70
4	82	7.58	7.58	8	0.74	0.77	0	0.00	0.00	75	6.93	6.93
5	45	4.16	4.16	3	0.28	0.29	0	0.00	0.00	40	3.70	3.70
6	41	3.79	3.79	3	0.28	0.29	0	0.00	0.00	27	2.50	2.50
7	28	2.59	2.59	0	0.00	0.00	0	0.00	0.00	21	1.94	1.94
8	18	1.66	1.66	1	0.09	0.10	0	0.00	0.00	14	1.29	1.29
9	8	0.74	0.74	0	0.00	0.00	0	0.00	0.00	10	0.92	0.92
10	10	0.92	0.92	0	0.00	0.00	0	0.00	0.00	6	0.55	0.55
11	6	0.55	0.55	0	0.00	0.00	0	0.00	0.00	5	0.46	0.46
12	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
13	2	0.18	0.18	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
15	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
17	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
19	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
28	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
NA	0	0.00	NA	41	3.79	NA	30	2.77	NA	0	0.00	NA

Note: N = 1,082. TW, earliest testing window. Only offenders with valid testing window start dates are included. Before, during, and after refer to when the incident associated with the record occurred, not when the arrest, charge, or conviction date occurred relative to the testing window. For example, an offender could be arrested during the testing window for an incident that occurred before it began. That would show up here in the before column. An incident is counted if there are any arrest, charge, or conviction records (any one type, or any combination of them will suffice) for the specified crime category associated with it.

Table 71: Frequency Distributions for Number of Traffic and Ordinance Incidents for Which Offender Was Arrested, Charged, or Convicted (Overall and By When Incident Occurred)

5.7.11.1 GEE Models To more closely examine the traffic and ordinance incident counts, we fitted two generalized estimating equation (GEE) models to the long-format *IDNEWL* data, after dropping rows for unobserved periods (*Years* = 0) because the counts are missing for those observations.

Model 1 used *When* as a predictor of the period-specific counts, but did not adjust for the period durations. Model 2 supplemented *When* as a predictor with an offset term for log(*Years*) to adjust for the period durations. Both models used Poisson distributions, an exchangeable correlation structure, a log link function, and robust standard errors based on the sandwich estimator.

```
HXCat12_11_gee1 <- IDNEWL %>%
  filter(Years > 0) %>%
  geeglm(HXCat12_11 ~ When,
         family = poisson(link = "log"), data = .,
         id = OID, corstr = "exchangeable", std.err = "san.se")

HXCat12_11_gee2 <- IDNEWL %>%
  filter(Years > 0) %>%
  geeglm(HXCat12_11 ~ When + offset(log(Years)),
         family = poisson(link = "log"), data = .,
         id = OID, corstr = "exchangeable", std.err = "san.se")
```

Table 72 below shows the parameter estimates for both GEE models.

```
TCap <- paste("Generalized Estimating Equation Models for Traffic and",
              "Ordinance Incident Counts",
              "Predicted by When Incidents Occurred")
FN <- paste ("\\item * p < .05, null hypothesis value outside 95\\% confidence",
             "interval based on z-score.",
             "\\\\\n\\item Note.",
             "The data contain up to 3 longitudinal observations (one per",
             "period) for each offender (i.e., cluster).",
             "The reference priod for When was after the testing window.",
             "Both models used a log link function, exchangeable correlation",
             "structure, a Poisson distribution, and robust standard errors",
             "(sandwich estimator).",
             "Model 1 did not use an offset.",
             "Model 2 used log(Years) as an offset term.")

texreg(list(HXCat12_11_gee1, HXCat12_11_gee2), booktabs = TRUE, dcolumn = TRUE,
       threeparttable = TRUE, fontsize = "normalsize", table = TRUE,
       use.packages = FALSE, ci.force = TRUE, label = "tab:HXCat12_11_texreg",
       stars = 0.05, caption = TCap, custom.note = FN)
```

	Model 1	Model 2
(Intercept)	0.63*	−2.26*
	[0.55; 0.70]	[−2.34; −2.18]
WhenBefore	−1.57*	−1.15*
	[−1.73; −1.42]	[−1.29; −1.00]
WhenDuring	−4.00*	−0.02
	[−4.35; −3.65]	[−0.38; 0.34]
Scale parameter: gamma	2.10	2.19
Scale parameter: SE	0.31	0.34
Correlation parameter: alpha	0.05	0.10
Correlation parameter: SE	0.02	0.03
Num. obs.	3175	3175
Num. clust.	1082	1082

* p < .05, null hypothesis value outside 95% confidence interval based on z-score.

Note. The data contain up to 3 longitudinal observations (one per period) for each offender (i.e., cluster). The reference priod for When was after the testing window. Both models used a log link function, exchangeable correlation structure, a Poisson distribution, and robust standard errors (sandwich estimator). Model 1 did not use an offset. Model 2 used log(Years) as an offset term.

Table 72: Generalized Estimating Equation Models for Traffic and Ordinance Incident Counts Predicted by When Incidents Occurred

5.7.11.2 Estimated Marginal Means Table 73 provides the estimated marginal means obtained from Models 1 and 2, along with corresponding confidence intervals. The units of measurement for those means are counts per offender for Model 1 (which ignore period duration) and counts per offender-year for Model 2 (which adjust for duration and are thus considered annual incidence rates).

```
# Table caption.
TCap <- paste("Marginal Means of Traffic and Ordinance Incident Counts and",
              "Incidence Rates By When Incidents Occurred")

# Table footnote.
FN <- paste("Model 1 did not use an offset, so means are counts.",
            "Model 2 used log(Years) as an offset term, so means are annual",
            "incidence rates.",
```

```
           "Confidence intervals are based on z-scores.")

HXCat12_11_emmeans1 <- emmeans(HXCat12_11_gee1, specs = pairwise ~ When,
                           type = "response", ref = "During")
HXCat12_11_emmeans2 <- emmeans(HXCat12_11_gee2, specs = pairwise ~ When,
                           type = "response", ref = "During")

as_tibble(HXCat12_11_emmeans1$emmeans) %>%
  full_join(x = ., y = as_tibble(HXCat12_11_emmeans2$emmeans)) %>%
  mutate(Model = c(1, 1, 1, 2, 2, 2),
         When = factor(When, levels = c("Before", "During", "After"))) %>%
  arrange(Model, When) %>%
  select(When, rate, SE, df, asymp.LCL, asymp.UCL) %>%
  kable(format = "latex", booktabs = TRUE, caption = TCap, digits = 2,
        col.names = c("When", "Rate", "SE", "df", "LCL", "UCL")) %>%
  kable_styling() %>%
  add_header_above(c(" " = 4, "95% CI" = 2)) %>%
  group_rows(group_label = "Model 1: Mean per offender", start_row = 1,
             end_row = 3, italic = TRUE) %>%
  group_rows(group_label = "Model 2: Mean per offender-year",
             start_row = 4, end_row = 6, italic = TRUE) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
## Joining, by = c("When", "rate", "SE", "df", "asymp.LCL", "asymp.UCL")
```

				95% CI	
When	Rate	SE	df	LCL	UCL
Model 1: Mean per offender					
Before	0.39	0.03	Inf	0.34	0.44
During	0.03	0.01	Inf	0.02	0.05
After	1.87	0.07	Inf	1.73	2.02
Model 2: Mean per offender-year					
Before	0.33	0.02	Inf	0.29	0.38
During	1.03	0.19	Inf	0.72	1.48
After	1.05	0.04	Inf	0.97	1.14

Note: Model 1 did not use an offset, so means are counts. Model 2 used log(Years) as an offset term, so means are annual incidence rates. Confidence intervals are based on z-scores.

Table 73: Marginal Means of Traffic and Ordinance Incident Counts and Incidence Rates By When Incidents Occurred

5.7.11.3 Contrasts for Ratios of Estimated Marginal Means Finally, we also estimated a set of contrasts to compare the marginal means from Models 1 and 2 (Table 74). Each of these contrasts estimates the ratios of a pair of means rather than a raw difference between means. This is because of the nature of the underlying Poisson GEE model.

```
TCap <- paste("Contrasts Estimating Ratios of Marginal Means for Traffic and",
              "Ordinance Incident Counts and Incidence Rates By When",
              "Incidents Occurred")

# Table footnote.
FN <- paste("Model 1 did not use an offset, so means are counts.",
            "Model 2 used log(Years) as an offset term, so means are annual",
            "incidence rates.",
            "These contrasts estimate ratios of those quantities and used",
            "Tukey's method to adjust for multiple comparisons.",
            "Confidence intervals are based on z-scores.")
```

```
as_tibble(confint(HXCat12_11_emmeans1$contrasts)) %>%
  full_join(x = ., y = as_tibble(confint(HXCat12_11_emmeans2$contrasts))) %>%
  kable(format = "latex", booktabs = TRUE, caption = TCap, digits = 2,
        col.names = c("Contrast", "Ratio", "SE", "df", "LCL", "UCL")) %>%
  kable_styling() %>%
  add_header_above(c(" " = 4, "95% CI" = 2)) %>%
  group_rows(group_label = "Model 1: Ratios of means per offender",
             start_row = 1, end_row = 3, italic = TRUE) %>%
  group_rows(group_label = "Model 2: Ratios of means per offender-year",
             start_row = 4, end_row = 6, italic = TRUE) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
## Joining, by = c("contrast", "ratio", "SE", "df", "asymp.LCL", "asymp.UCL")
```

				95% CI	
Contrast	Ratio	SE	df	LCL	UCL
Model 1: Ratios of means per offender					
After / Before	4.82	0.38	Inf	4.01	5.80
After / During	54.51	9.73	Inf	35.87	82.83
Before / During	11.30	2.09	Inf	7.32	17.45
Model 2: Ratios of means per offender-year					
After / Before	3.15	0.23	Inf	2.65	3.75
After / During	1.02	0.19	Inf	0.66	1.57
Before / During	0.32	0.06	Inf	0.21	0.51

Note:
Model 1 did not use an offset, so means are counts. Model 2 used log(Years) as an offset term, so means are annual incidence rates. These contrasts estimate ratios of those quantities and used Tukey's method to adjust for multiple comparisons. Confidence intervals are based on z-scores.

Table 74: Contrasts Estimating Ratios of Marginal Means for Traffic and Ordinance Incident Counts and Incidence Rates By When Incidents Occurred

5.7.12 Weapons

Among these 1,082 suspected sexual offenders, there were 560 (52%) offenders who had criminal histories containing at least one incident associated with an arrest, charge, or conviction for weapons crimes. Table 75 summarizes the counts of unique weapons crime incidents overall and by when the incident occurred, while Table 76 shows the frequency distributions underlying those summaries.

```
# Table caption.
TCap <- paste("Offender-Level Weapons Crime Incident Counts (Overall and By",
              "When Incident Occurred)")

# Summarize crime category counts (both overall and broken down by IWhen).
IDNEW %>%
  select(HXCat12_12, HXCat12_12_Before, HXCat12_12_During, HXCat12_12_After) %>%
  describe(., quant=c(.25, .50, .75)) %>%
  bind_cols(data.frame(Variable = TRowLabs), .) %>%
  select(Variable, n, mean, sd, min, max, skew, kurtosis, Q0.25, Q0.5, Q0.75) %>%
  kable(., format = "latex", booktabs = TRUE, digits = 2, row.names = FALSE,
        col.names = CNames, caption = TCap,
        linesep = c('', '', '', '\\addlinespace')) %>%
  column_spec(column = 1, width = "6 cm") %>%
  footnote(general = FN1, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

Variable	N	Mean	SD	Min	Max	Skew	Kurtosis	Q25	Q50	Q75
Incident count (arrested, charged, or convicted)	1082	0.85	1.07	0	8	1.55	3.31	0	1	1
...Before testing window	1041	0.30	0.61	0	4	2.26	5.61	0	0	0
...During testing window	1052	0.04	0.21	0	2	4.70	21.87	0	0	0
...After testing window	1082	0.52	0.86	0	6	2.14	6.08	0	0	1

Note: Only offenders with valid testing window start dates are included. Only incidents where the offender was arrested for, charged with, or convicted of at least one offense from one of the 12 main crime categories were counted. Before, during, and after refer to when the incident associated with the record occurred, not when the arrest date, charge date, or conviction date occurred relative to the testing window.

Table 75: Offender-Level Weapons Crime Incident Counts (Overall and By When Incident Occurred)

```
# Table caption.
TCap <- paste("Frequency Distributions for Number of Weapons Crime Incidents for",
              "Which Offender Was Arrested, Charged, or Convicted (Overall and",
              "By When Incident Occurred)")

# Get frequency distributions.
IDNEW %>%
  select(OID, HXCat12_12, HXCat12_12_Before, HXCat12_12_During, HXCat12_12_After) %>%
  rename(All = HXCat12_12, Before = HXCat12_12_Before, During = HXCat12_12_During,
         After = HXCat12_12_After) %>%
  pivot_longer(cols = -OID, names_to = "When", values_to = "Value") %>%
  group_by(When, Value) %>%
  count() %>%
  arrange(Value) %>%
  pivot_wider(names_from = When, values_from = n, values_fill = 0) %>%
  mutate(All_p = 100*All/N_All,
         All_v = if_else(is.na(Value),
                         true = as.numeric(NA),
                         false = 100*All/N_All),
         Before_p = 100*Before/N_All,
         Before_v = if_else(is.na(Value),
                            true = as.numeric(NA),
                            false = 100*Before/N_Before),
         During_p = 100*During/N_All,
         During_v = if_else(is.na(Value),
                            true = as.numeric(NA),
                            false = 100*During/N_During),
         After_p = 100*After/N_All,
         After_v = if_else(is.na(Value),
                           true = as.numeric(NA),
                           false = 100*After/N_After)) %>%
  select(Value, All, All_p, All_v, Before, Before_p, Before_v, During,
         During_p, During_v, After, After_p, After_v) %>%
  kable(format = "latex", booktabs = TRUE, digits = 2,
        col.names = c("Value", rep(c("N", "%", "Valid %"), 4)),
        caption = TCap) %>%
  kable_styling() %>%
  add_header_above(c(" " = 1, "Overall" = 3, "Before TW" = 3, "During TW" = 3,
                     "After TW" = 3)) %>%
  footnote(general = FN5, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

	Overall			Before TW			During TW			After TW		
Value	N	%	Valid %	N	%	Valid %	N	%	Valid %	N	%	Valid %
0	522	48.24	48.24	799	73.84	76.75	1006	92.98	95.63	696	64.33	64.33
1	322	29.76	29.76	184	17.01	17.68	45	4.16	4.28	262	24.21	24.21
2	154	14.23	14.23	49	4.53	4.71	1	0.09	0.10	88	8.13	8.13
3	55	5.08	5.08	7	0.65	0.67	0	0.00	0.00	23	2.13	2.13
4	21	1.94	1.94	2	0.18	0.19	0	0.00	0.00	9	0.83	0.83
5	6	0.55	0.55	0	0.00	0.00	0	0.00	0.00	2	0.18	0.18
6	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	2	0.18	0.18
8	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
NA	0	0.00	NA	41	3.79	NA	30	2.77	NA	0	0.00	NA

Note: N = 1,082. TW, earliest testing window. Only offenders with valid testing window start dates are included. Before, during, and after refer to when the incident associated with the record occurred, not when the arrest, charge, or conviction date occurred relative to the testing window. For example, an offender could be arrested during the testing window for an incident that occurred before it began. That would show up here in the before column. An incident is counted if there are any arrest, charge, or conviction records (any one type, or any combination of them will suffice) for the specified crime category associated with it.

Table 76: Frequency Distributions for Number of Weapons Crime Incidents for Which Offender Was Arrested, Charged, or Convicted (Overall and By When Incident Occurred)

5.7.12.1 GEE Models To more closely examine the weapon incident counts, we fitted two generalized estimating equation (GEE) models to the long-format *IDNEWL* data, after dropping rows for unobserved periods (*Years* = 0) because the counts are missing for those observations.

Model 1 used *When* as a predictor of the period-specific counts, but did not adjust for the period durations. Model 2 supplemented *When* as a predictor with an offset term for log(*Years*) to adjust for the period durations. Both models used Poisson distributions, an exchangeable correlation structure, a log link function, and robust standard errors based on the sandwich estimator.

```
HXCat12_12_gee1 <- IDNEWL %>%
  filter(Years > 0) %>%
  geeglm(HXCat12_12 ~ When,
         family = poisson(link = "log"), data = .,
         id = OID, corstr = "exchangeable", std.err = "san.se")

HXCat12_12_gee2 <- IDNEWL %>%
  filter(Years > 0) %>%
  geeglm(HXCat12_12 ~ When + offset(log(Years)),
         family = poisson(link = "log"), data = .,
         id = OID, corstr = "exchangeable", std.err = "san.se")
```

Table 77 below shows the parameter estimates for both GEE models.

```
TCap <- paste("Generalized Estimating Equation Models for Weapon Incident Counts",
              "Predicted by When Incidents Occurred")
FN <- paste ("\\item * p < .05, null hypothesis value outside 95\\% confidence",
             "interval based on z-score.",
             "\\\\\n\\item Note.",
             "The data contain up to 3 longitudinal observations (one per",
             "period) for each offender (i.e., cluster).",
             "The reference priod for When was after the testing window.",
             "Both models used a log link function, exchangeable correlation",
             "structure, a Poisson distribution, and robust standard errors",
             "(sandwich estimator).",
             "Model 1 did not use an offset.",
```

```
             "Model 2 used log(Years) as an offset term.")

texreg(list(HXCat12_12_gee1, HXCat12_12_gee2), booktabs = TRUE, dcolumn = TRUE,
       threeparttable = TRUE, fontsize = "normalsize", table = TRUE,
       use.packages = FALSE, ci.force = TRUE, label = "tab:HXCat12_12_texreg",
       stars = 0.05, caption = TCap, custom.note = FN)
```

	Model 1	Model 2
(Intercept)	−0.65*	−3.54*
	[−0.75; −0.55]	[−3.64; −3.44]
WhenBefore	−0.56*	−0.13
	[−0.71; −0.40]	[−0.28; 0.02]
WhenDuring	−2.46*	1.55*
	[−2.76; −2.16]	[1.25; 1.86]
Scale parameter: gamma	1.21	1.24
Scale parameter: SE	0.18	0.19
Correlation parameter: alpha	0.01	0.05
Correlation parameter: SE	0.02	0.02
Num. obs.	3175	3175
Num. clust.	1082	1082

* p < .05, null hypothesis value outside 95% confidence interval based on z-score.

Note. The data contain up to 3 longitudinal observations (one per period) for each offender (i.e., cluster). The reference priod for When was after the testing window. Both models used a log link function, exchangeable correlation structure, a Poisson distribution, and robust standard errors (sandwich estimator). Model 1 did not use an offset. Model 2 used log(Years) as an offset term.

Table 77: Generalized Estimating Equation Models for Weapon Incident Counts Predicted by When Incidents Occurred

5.7.12.2 Estimated Marginal Means Table 78 provides the estimated marginal means obtained from Models 1 and 2, along with corresponding confidence intervals. The units of measurement for those means are counts per offender for Model 1 (which ignore period duration) and counts per offender-year for Model 2 (which adjust for duration and are thus considered annual incidence rates).

```
# Table caption.
TCap <- paste("Marginal Means of Weapon Incident Counts and Incidence Rates",
              "By When Incidents Occurred")

# Table footnote.
FN <- paste("Model 1 did not use an offset, so means are counts.",
            "Model 2 used log(Years) as an offset term, so means are annual",
            "incidence rates.",
            "Confidence intervals are based on z-scores.")

HXCat12_12_emmeans1 <- emmeans(HXCat12_12_gee1, specs = pairwise ~ When,
                           type = "response", ref = "During")
HXCat12_12_emmeans2 <- emmeans(HXCat12_12_gee2, specs = pairwise ~ When,
                           type = "response", ref = "During")

as_tibble(HXCat12_12_emmeans1$emmeans) %>%
  full_join(x = ., y = as_tibble(HXCat12_12_emmeans2$emmeans)) %>%
  mutate(Model = c(1, 1, 1, 2, 2, 2),
         When = factor(When, levels = c("Before", "During", "After"))) %>%
  arrange(Model, When) %>%
  select(When, rate, SE, df, asymp.LCL, asymp.UCL) %>%
  kable(format = "latex", booktabs = TRUE, caption = TCap, digits = 2,
        col.names = c("When", "Rate", "SE", "df", "LCL", "UCL")) %>%
  kable_styling() %>%
  add_header_above(c(" " = 4, "95% CI" = 2)) %>%
```

```
  group_rows(group_label = "Model 1: Mean per offender", start_row = 1,
             end_row = 3, italic = TRUE) %>%
  group_rows(group_label = "Model 2: Mean per offender-year",
             start_row = 4, end_row = 6, italic = TRUE) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
## Joining, by = c("When", "rate", "SE", "df", "asymp.LCL", "asymp.UCL")
```

				95% CI	
When	Rate	SE	df	LCL	UCL
Model 1: Mean per offender					
Before	0.30	0.02	Inf	0.26	0.34
During	0.04	0.01	Inf	0.03	0.06
After	0.52	0.03	Inf	0.47	0.58
Model 2: Mean per offender-year					
Before	0.26	0.02	Inf	0.23	0.29
During	1.38	0.20	Inf	1.04	1.84
After	0.29	0.01	Inf	0.27	0.32

Note: Model 1 did not use an offset, so means are counts. Model 2 used log(Years) as an offset term, so means are annual incidence rates. Confidence intervals are based on z-scores.

Table 78: Marginal Means of Weapon Incident Counts and Incidence Rates By When Incidents Occurred

5.7.12.3 Contrasts for Ratios of Estimated Marginal Means Finally, we also estimated a set of contrasts to compare the marginal means from Models 1 and 2 (Table 79). Each of these contrasts estimates the ratios of a pair of means rather than a raw difference between means. This is because of the nature of the underlying Poisson GEE model.

```
TCap <- paste("Contrasts Estimating Ratios of Marginal Means for Weapon",
              "Incident Counts and Incidence Rates By When Incidents",
              "Occurred")

# Table footnote.
FN <- paste("Model 1 did not use an offset, so means are counts.",
            "Model 2 used log(Years) as an offset term, so means are annual",
            "incidence rates.",
            "These contrasts estimate ratios of those quantities and used",
            "Tukey's method to adjust for multiple comparisons.",
            "Confidence intervals are based on z-scores.")

as_tibble(confint(HXCat12_12_emmeans1$contrasts)) %>%
  full_join(x = ., y = as_tibble(confint(HXCat12_12_emmeans2$contrasts))) %>%
  kable(format = "latex", booktabs = TRUE, caption = TCap, digits = 2,
        col.names = c("Contrast", "Ratio", "SE", "df", "LCL", "UCL")) %>%
  kable_styling() %>%
  add_header_above(c(" " = 4, "95% CI" = 2)) %>%
  group_rows(group_label = "Model 1: Ratios of means per offender",
             start_row = 1, end_row = 3, italic = TRUE) %>%
  group_rows(group_label = "Model 2: Ratios of means per offender-year",
             start_row = 4, end_row = 6, italic = TRUE) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
## Joining, by = c("contrast", "ratio", "SE", "df", "asymp.LCL", "asymp.UCL")
```

Contrast	Ratio	SE	df	95% CI LCL	95% CI UCL
Model 1: Ratios of means per offender					
After / Before	1.75	0.14	Inf	1.45	2.11
After / During	11.68	1.80	Inf	8.14	16.76
Before / During	6.68	1.04	Inf	4.64	9.63
Model 2: Ratios of means per offender-year					
After / Before	1.14	0.09	Inf	0.95	1.36
After / During	0.21	0.03	Inf	0.15	0.30
Before / During	0.19	0.03	Inf	0.13	0.27

Note:
Model 1 did not use an offset, so means are counts. Model 2 used log(Years) as an offset term, so means are annual incidence rates. These contrasts estimate ratios of those quantities and used Tukey's method to adjust for multiple comparisons. Confidence intervals are based on z-scores.

Table 79: Contrasts Estimating Ratios of Marginal Means for Weapon Incident Counts and Incidence Rates By When Incidents Occurred

5.8 Incident Counts by Broader Crime Category (4 levels) and Period

In this section, we analyze variables that contain the number of unique incidents in the offender's criminal history for a given broader category of crime. The numerical results for broader categories were omitted from the paper for the sake of brevity, but we retain them here for future reference.

For each broader crime category, we report include counts of incidents based on a combined criterion where the offender was *arrested for, charged with, or convicted of* (any of the three events, or any combination of them) any of the more specific types of crimes classified into the relevant category.

Note that these variables were created by first aggregating *ARR*, *CHG*, and *JUD* records to the incident level to flag incidents that fit the relevant category, then aggregated again to get to the offender level incident counts. It is possible to have multiple arrest offense records, charge records, or adjudication records for convictions on a single incident, but these variables ignore that and only count each incident once.

5.8.1 Sexual Crimes

Among these 1,082 suspected sexual offenders, there were 449 (41%) offenders who had criminal histories containing at least one incident associated with an arrest, charge, or conviction for a sexual crime (either a CSC or a non-CSC sex crime). Table 80 summarizes the counts of unique sexual crimes incidents overall and by when the incident occurred, while Table 81 shows the frequency distributions underlying those summaries.

```
# Table caption.
TCap <- paste("Offender-Level Sexual Crimes Incident Counts (Overall and By When",
              "Incident Occurred)")

# Summarize crime category counts (both overall and broken down by IWhen).
IDNEW %>%
  select(HXCat12_Sexual, HXCat12_Sexual_Before, HXCat12_Sexual_During,
         HXCat12_Sexual_After) %>%
  describe(., quant=c(.25, .50, .75)) %>%
  bind_cols(data.frame(Variable = TRowLabs), .) %>%
  select(Variable, n, mean, sd, min, max, skew, kurtosis, Q0.25, Q0.5, Q0.75) %>%
  kable(., format = "latex", booktabs = TRUE, digits = 2, row.names = FALSE,
        col.names = CNames, caption = TCap,
        linesep = c('', '', '', '\\addlinespace')) %>%
  column_spec(column = 1, width = "6 cm") %>%
```

```
  footnote(general = FN1, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

Variable	N	Mean	SD	Min	Max	Skew	Kurtosis	Q25	Q50	Q75
Incident count (arrested, charged, or convicted)	1082	0.70	1.14	0	8	2.61	9.21	0	0	1
...Before testing window	1041	0.13	0.38	0	3	3.33	12.60	0	0	0
...During testing window	1052	0.22	0.55	0	7	5.07	47.84	0	0	0
...After testing window	1082	0.37	0.87	0	7	3.59	16.44	0	0	0

Note: Only offenders with valid testing window start dates are included. Only incidents where the offender was arrested for, charged with, or convicted of at least one offense from one of the 12 main crime categories were counted. Before, during, and after refer to when the incident associated with the record occurred, not when the arrest date, charge date, or conviction date occurred relative to the testing window.

Table 80: Offender-Level Sexual Crimes Incident Counts (Overall and By When Incident Occurred)

```
# Table caption.
TCap <- paste("Frequency Distributions for Number of Sexual Crimes Incidents ",
              "for Which Offender Was Arrested, Charged, or Convicted (Overall",
              "and By When Incident Occurred)")

# Get frequency distributions.
IDNEW %>%
  select(OID, HXCat12_Sexual, HXCat12_Sexual_Before, HXCat12_Sexual_During,
         HXCat12_Sexual_After) %>%
  rename(All = HXCat12_Sexual, Before = HXCat12_Sexual_Before,
         During = HXCat12_Sexual_During, After = HXCat12_Sexual_After) %>%
  pivot_longer(cols = -OID, names_to = "When", values_to = "Value") %>%
  group_by(When, Value) %>%
  count() %>%
  arrange(Value) %>%
  pivot_wider(names_from = When, values_from = n, values_fill = 0) %>%
  mutate(All_p = 100*All/N_All,
         All_v = if_else(is.na(Value),
                         true = as.numeric(NA),
                         false = 100*All/N_All),
         Before_p = 100*Before/N_All,
         Before_v = if_else(is.na(Value),
                            true = as.numeric(NA),
                            false = 100*Before/N_Before),
         During_p = 100*During/N_All,
         During_v = if_else(is.na(Value),
                            true = as.numeric(NA),
                            false = 100*During/N_During),
         After_p = 100*After/N_All,
         After_v = if_else(is.na(Value),
                           true = as.numeric(NA),
                           false = 100*After/N_After)) %>%
  select(Value, All, All_p, All_v, Before, Before_p, Before_v, During,
         During_p, During_v, After, After_p, After_v) %>%
  kable(format = "latex", booktabs = TRUE, digits = 2,
        col.names = c("Value", rep(c("N", "%", "Valid %"), 4)),
        caption = TCap) %>%
  kable_styling() %>%
  add_header_above(c(" " = 1, "Overall" = 3, "Before TW" = 3, "During TW" = 3,
                     "After TW" = 3)) %>%
  footnote(general = FN5, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

	Overall			Before TW			During TW			After TW		
Value	N	%	Valid %	N	%	Valid %	N	%	Valid %	N	%	Valid %
0	633	58.50	58.50	925	85.49	88.86	851	78.65	80.89	836	77.26	77.26
1	278	25.69	25.69	102	9.43	9.80	184	17.01	17.49	165	15.25	15.25
2	104	9.61	9.61	12	1.11	1.15	12	1.11	1.14	49	4.53	4.53
3	35	3.23	3.23	2	0.18	0.19	2	0.18	0.19	14	1.29	1.29
4	13	1.20	1.20	0	0.00	0.00	1	0.09	0.10	8	0.74	0.74
5	6	0.55	0.55	0	0.00	0.00	0	0.00	0.00	3	0.28	0.28
6	7	0.65	0.65	0	0.00	0.00	0	0.00	0.00	6	0.55	0.55
7	4	0.37	0.37	0	0.00	0.00	2	0.18	0.19	1	0.09	0.09
8	2	0.18	0.18	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
NA	0	0.00	NA	41	3.79	NA	30	2.77	NA	0	0.00	NA

Note: N = 1,082. TW, earliest testing window. Only offenders with valid testing window start dates are included. Before, during, and after refer to when the incident associated with the record occurred, not when the arrest, charge, or conviction date occurred relative to the testing window. For example, an offender could be arrested during the testing window for an incident that occurred before it began. That would show up here in the before column. An incident is counted if there are any arrest, charge, or conviction records (any one type, or any combination of them will suffice) for the specified crime category associated with it.

Table 81: Frequency Distributions for Number of Sexual Crimes Incidents for Which Offender Was Arrested, Charged, or Convicted (Overall and By When Incident Occurred)

5.8.1.1 GEE Models To more closely examine the sexual crimes incident counts, we fitted two generalized estimating equation (GEE) models to the long-format *IDNEWL* data, after dropping rows for unobserved periods (*Years* = 0) because the counts are missing for those observations.

Model 1 used *When* as a predictor of the period-specific counts, but did not adjust for the period durations. Model 2 supplemented *When* as a predictor with an offset term for log(*Years*) to adjust for the period durations. Both models used Poisson distributions, an exchangeable correlation structure, a log link function, and robust standard errors based on the sandwich estimator.

```
HXCat12_Sexual_gee1 <- IDNEWL %>%
  filter(Years > 0) %>%
  geeglm(HXCat12_Sexual ~ When,
         family = poisson(link = "log"), data = .,
         id = OID, corstr = "exchangeable", std.err = "san.se")

HXCat12_Sexual_gee2 <- IDNEWL %>%
  filter(Years > 0) %>%
  geeglm(HXCat12_Sexual ~ When + offset(log(Years)),
         family = poisson(link = "log"), data = .,
         id = OID, corstr = "exchangeable", std.err = "san.se")
```

Table 82 below shows the parameter estimates for both GEE models.

```
TCap <- paste("Generalized Estimating Equation Models for Sexual Crimes Incident Counts",
              "Predicted by When Incidents Occurred")
FN <- paste ("\\item * p < .05, null hypothesis value outside 95\\% confidence",
             "interval based on z-score.",
             "\\\\\n\\item Note.",
             "The data contain up to 3 longitudinal observations (one per",
             "period) for each offender (i.e., cluster).",
             "The reference priod for When was after the testing window.",
             "Both models used a log link function, exchangeable correlation",
             "structure, a Poisson distribution, and robust standard errors",
             "(sandwich estimator).",
```

```
             "Model 1 did not use an offset.",
             "Model 2 used log(Years) as an offset term.")

texreg(list(HXCat12_Sexual_gee1, HXCat12_Sexual_gee2), booktabs = TRUE,
       dcolumn = TRUE, threeparttable = TRUE, fontsize = "normalsize",
       table = TRUE, use.packages = FALSE, ci.force = TRUE,
       label = "tab:HXCat12_Sexual_texreg",
       stars = 0.05, caption = TCap, custom.note = FN)
```

	Model 1	Model 2
(Intercept)	-1.01^*	-3.90^*
	$[-1.15; -0.87]$	$[-4.04; -3.76]$
WhenBefore	-1.06^*	-0.64^*
	$[-1.28; -0.84]$	$[-0.85; -0.42]$
WhenDuring	-0.51^*	3.51^*
	$[-0.71; -0.30]$	$[3.31; 3.72]$
Scale parameter: gamma	1.52	1.44
Scale parameter: SE	0.11	0.11
Correlation parameter: alpha	0.06	0.07
Correlation parameter: SE	0.02	0.02
Num. obs.	3175	3175
Num. clust.	1082	1082

* $p < .05$, null hypothesis value outside 95% confidence interval based on z-score.

Note. The data contain up to 3 longitudinal observations (one per period) for each offender (i.e., cluster). The reference priod for When was after the testing window. Both models used a log link function, exchangeable correlation structure, a Poisson distribution, and robust standard errors (sandwich estimator). Model 1 did not use an offset. Model 2 used log(Years) as an offset term.

Table 82: Generalized Estimating Equation Models for Sexual Crimes Incident Counts Predicted by When Incidents Occurred

5.8.1.2 Estimated Marginal Means Table 83 provides the estimated marginal means obtained from Models 1 and 2, along with corresponding confidence intervals. The units of measurement for those means are counts per offender for Model 1 (which ignore period duration) and counts per offender-year for Model 2 (which adjust for duration and are thus considered annual incidence rates).

```
# Table caption.
TCap <- paste("Marginal Means of Sexual Crimes Incident Counts and Incidence Rates",
              "By When Incidents Occurred")

# Table footnote.
FN <- paste("Model 1 did not use an offset, so means are counts.",
            "Model 2 used log(Years) as an offset term, so means are annual",
            "incidence rates.",
            "Confidence intervals are based on z-scores.")

HXCat12_Sexual_emmeans1 <- emmeans(HXCat12_Sexual_gee1, specs = pairwise ~ When,
                           type = "response", ref = "During")
HXCat12_Sexual_emmeans2 <- emmeans(HXCat12_Sexual_gee2, specs = pairwise ~ When,
                           type = "response", ref = "During")

as_tibble(HXCat12_Sexual_emmeans1$emmeans) %>%
  full_join(x = ., y = as_tibble(HXCat12_Sexual_emmeans2$emmeans)) %>%
  mutate(Model = c(1, 1, 1, 2, 2, 2),
         When = factor(When, levels = c("Before", "During", "After"))) %>%
  arrange(Model, When) %>%
  select(When, rate, SE, df, asymp.LCL, asymp.UCL) %>%
  kable(format = "latex", booktabs = TRUE, caption = TCap, digits = 2,
```

```
       col.names = c("When", "Rate", "SE", "df", "LCL", "UCL")) %>%
  kable_styling() %>%
  add_header_above(c(" " = 4, "95% CI" = 2)) %>%
  group_rows(group_label = "Model 1: Mean per offender", start_row = 1,
             end_row = 3, italic = TRUE) %>%
  group_rows(group_label = "Model 2: Mean per offender-year",
             start_row = 4, end_row = 6, italic = TRUE) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
## Joining, by = c("When", "rate", "SE", "df", "asymp.LCL", "asymp.UCL")
```

				95% CI	
When	Rate	SE	df	LCL	UCL
Model 1: Mean per offender					
Before	0.13	0.01	Inf	0.11	0.15
During	0.22	0.02	Inf	0.19	0.26
After	0.37	0.03	Inf	0.32	0.42
Model 2: Mean per offender-year					
Before	0.11	0.01	Inf	0.09	0.13
During	6.88	0.52	Inf	5.92	7.98
After	0.20	0.01	Inf	0.18	0.24

Note: Model 1 did not use an offset, so means are counts. Model 2 used log(Years) as an offset term, so means are annual incidence rates. Confidence intervals are based on z-scores.

Table 83: Marginal Means of Sexual Crimes Incident Counts and Incidence Rates By When Incidents Occurred

5.8.1.3 Contrasts for Ratios of Estimated Marginal Means Finally, we also estimated a set of contrasts to compare the marginal means from Models 1 and 2 (Table 84). Each of these contrasts estimates the ratios of a pair of means rather than a raw difference between means. This is because of the nature of the underlying Poisson GEE model.

```
TCap <- paste("Contrasts Estimating Ratios of Marginal Means for Sexual Crimes",
              "Incident Counts and Incidence Rates By When Incidents",
              "Occurred")

# Table footnote.
FN <- paste("Model 1 did not use an offset, so means are counts.",
            "Model 2 used log(Years) as an offset term, so means are annual",
            "incidence rates.",
            "These contrasts estimate ratios of those quantities and used",
            "Tukey's method to adjust for multiple comparisons.",
            "Confidence intervals are based on z-scores.")

as_tibble(confint(HXCat12_Sexual_emmeans1$contrasts)) %>%
  full_join(x = ., y = as_tibble(confint(HXCat12_Sexual_emmeans2$contrasts))) %>%
  kable(format = "latex", booktabs = TRUE, caption = TCap, digits = 2,
        col.names = c("Contrast", "Ratio", "SE", "df", "LCL", "UCL")) %>%
  kable_styling() %>%
  add_header_above(c(" " = 4, "95% CI" = 2)) %>%
  group_rows(group_label = "Model 1: Ratios of means per offender",
             start_row = 1, end_row = 3, italic = TRUE) %>%
  group_rows(group_label = "Model 2: Ratios of means per offender-year",
             start_row = 4, end_row = 6, italic = TRUE) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
## Joining, by = c("contrast", "ratio", "SE", "df", "asymp.LCL", "asymp.UCL")
```

Contrast	Ratio	SE	df	95% CI LCL	95% CI UCL
Model 1: Ratios of means per offender					
After / Before	2.89	0.32	Inf	2.23	3.74
After / During	1.66	0.17	Inf	1.30	2.11
Before / During	0.57	0.07	Inf	0.43	0.76
Model 2: Ratios of means per offender-year					
After / Before	1.89	0.21	Inf	1.46	2.44
After / During	0.03	0.00	Inf	0.02	0.04
Before / During	0.02	0.00	Inf	0.01	0.02

Note:
Model 1 did not use an offset, so means are counts. Model 2 used log(Years) as an offset term, so means are annual incidence rates. These contrasts estimate ratios of those quantities and used Tukey's method to adjust for multiple comparisons. Confidence intervals are based on z-scores.

Table 84: Contrasts Estimating Ratios of Marginal Means for Sexual Crimes Incident Counts and Incidence Rates By When Incidents Occurred

5.8.2 Violent Non-Sexual Crimes

Among these 1,082 suspected sexual offenders, there were 933 (86%) offenders who had criminal histories containing at least one incident associated with an arrest, charge, or conviction for a violent, non-sexual crime (homicide; assault involving domestic violence and/or stalking; non-sexual, non-domestic violence assault; robbery; or weapons). Table 85 summarizes the counts of unique sexual crimes incidents overall and by when the incident occurred, while Table 86 shows the frequency distributions underlying those summaries.

```
# Table caption.
TCap <- paste("Offender-Level Violent, Non-Sexual Crimes Incident Counts (Overall and By When",
              "Incident Occurred)")

# Summarize crime category counts (both overall and broken down by IWhen).
IDNEW %>%
  select(HXCat12_Violent, HXCat12_Violent_Before, HXCat12_Violent_During,
         HXCat12_Violent_After) %>%
  describe(., quant=c(.25, .50, .75)) %>%
  bind_cols(data.frame(Variable = TRowLabs), .) %>%
  select(Variable, n, mean, sd, min, max, skew, kurtosis, Q0.25, Q0.5, Q0.75) %>%
  kable(., format = "latex", booktabs = TRUE, digits = 2, row.names = FALSE,
        col.names = CNames, caption = TCap,
        linesep = c('', '', '', '\\addlinespace')) %>%
  column_spec(column = 1, width = "6 cm") %>%
  footnote(general = FN1, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
# Table caption.
TCap <- paste("Frequency Distributions for Number of Violent, Non-Sexual Crimes Incidents ",
              "for Which Offender Was Arrested, Charged, or Convicted (Overall",
              "and By When Incident Occurred)")

# Get frequency distributions.
IDNEW %>%
  select(OID, HXCat12_Violent, HXCat12_Violent_Before, HXCat12_Violent_During,
         HXCat12_Violent_After) %>%
  rename(All = HXCat12_Violent, Before = HXCat12_Violent_Before,
         During = HXCat12_Violent_During, After = HXCat12_Violent_After) %>%
  pivot_longer(cols = -OID, names_to = "When", values_to = "Value") %>%
  group_by(When, Value) %>%
  count() %>%
```

Variable	N	Mean	SD	Min	Max	Skew	Kurtosis	Q25	Q50	Q75
Incident count (arrested, charged, or convicted)	1082	2.61	2.13	0	15	1.21	2.30	1	2	4
...Before testing window	1041	0.78	1.10	0	8	1.86	4.53	0	0	1
...During testing window	1052	0.13	0.38	0	3	3.12	11.20	0	0	0
...After testing window	1082	1.73	1.84	0	13	1.67	4.15	0	1	3

Note: Only offenders with valid testing window start dates are included. Only incidents where the offender was arrested for, charged with, or convicted of at least one offense from one of the 12 main crime categories were counted. Before, during, and after refer to when the incident associated with the record occurred, not when the arrest date, charge date, or conviction date occurred relative to the testing window.

Table 85: Offender-Level Violent, Non-Sexual Crimes Incident Counts (Overall and By When Incident Occurred)

```
arrange(Value) %>%
pivot_wider(names_from = When, values_from = n, values_fill = 0) %>%
mutate(All_p = 100*All/N_All,
       All_v = if_else(is.na(Value),
                       true = as.numeric(NA),
                       false = 100*All/N_All),
       Before_p = 100*Before/N_All,
       Before_v = if_else(is.na(Value),
                          true = as.numeric(NA),
                          false = 100*Before/N_Before),
       During_p = 100*During/N_All,
       During_v = if_else(is.na(Value),
                          true = as.numeric(NA),
                          false = 100*During/N_During),
       After_p = 100*After/N_All,
       After_v = if_else(is.na(Value),
                         true = as.numeric(NA),
                         false = 100*After/N_After)) %>%
select(Value, All, All_p, All_v, Before, Before_p, Before_v, During,
       During_p, During_v, After, After_p, After_v) %>%
kable(format = "latex", booktabs = TRUE, digits = 2,
      col.names = c("Value", rep(c("N", "%", "Valid %"), 4)),
      caption = TCap) %>%
kable_styling() %>%
add_header_above(c(" " = 1, "Overall" = 3, "Before TW" = 3, "During TW" = 3,
                   "After TW" = 3)) %>%
footnote(general = FN5, general_title = "Note: ", footnote_as_chunk = TRUE,
       threeparttable = TRUE)
```

	Overall			Before TW			During TW			After TW		
Value	N	%	Valid %	N	%	Valid %	N	%	Valid %	N	%	Valid %
0	149	13.77	13.77	556	51.39	53.41	925	85.49	87.93	320	29.57	29.57
1	240	22.18	22.18	281	25.97	26.99	115	10.63	10.93	276	25.51	25.51
2	209	19.32	19.32	128	11.83	12.30	10	0.92	0.95	211	19.50	19.50
3	176	16.27	16.27	43	3.97	4.13	2	0.18	0.19	121	11.18	11.18
4	133	12.29	12.29	20	1.85	1.92	0	0.00	0.00	72	6.65	6.65
5	70	6.47	6.47	9	0.83	0.86	0	0.00	0.00	35	3.23	3.23
6	45	4.16	4.16	3	0.28	0.29	0	0.00	0.00	20	1.85	1.85
7	32	2.96	2.96	0	0.00	0.00	0	0.00	0.00	11	1.02	1.02
8	12	1.11	1.11	1	0.09	0.10	0	0.00	0.00	9	0.83	0.83
9	6	0.55	0.55	0	0.00	0.00	0	0.00	0.00	2	0.18	0.18
10	4	0.37	0.37	0	0.00	0.00	0	0.00	0.00	2	0.18	0.18
11	2	0.18	0.18	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
12	3	0.28	0.28	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
13	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
15	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
NA	0	0.00	NA	41	3.79	NA	30	2.77	NA	0	0.00	NA

Note: N = 1,082. TW, earliest testing window. Only offenders with valid testing window start dates are included. Before, during, and after refer to when the incident associated with the record occurred, not when the arrest, charge, or conviction date occurred relative to the testing window. For example, an offender could be arrested during the testing window for an incident that occurred before it began. That would show up here in the before column. An incident is counted if there are any arrest, charge, or conviction records (any one type, or any combination of them will suffice) for the specified crime category associated with it.

Table 86: Frequency Distributions for Number of Violent, Non-Sexual Crimes Incidents for Which Offender Was Arrested, Charged, or Convicted (Overall and By When Incident Occurred)

5.8.2.1 GEE Models To more closely examine the violent, non-sexual crimes incident counts, we fitted two generalized estimating equation (GEE) models to the long-format *IDNEWL* data, after dropping rows for unobserved periods (*Years* = 0) because the counts are missing for those observations.

Model 1 used *When* as a predictor of the period-specific counts, but did not adjust for the period durations. Model 2 supplemented *When* as a predictor with an offset term for log(*Years*) to adjust for the period durations. Both models used Poisson distributions, an exchangeable correlation structure, a log link function, and robust standard errors based on the sandwich estimator.

```
HXCat12_Violent_gee1 <- IDNEWL %>%
  filter(Years > 0) %>%
  geeglm(HXCat12_Violent ~ When,
         family = poisson(link = "log"), data = .,
         id = OID, corstr = "exchangeable", std.err = "san.se")

HXCat12_Violent_gee2 <- IDNEWL %>%
  filter(Years > 0) %>%
  geeglm(HXCat12_Violent ~ When + offset(log(Years)),
         family = poisson(link = "log"), data = .,
         id = OID, corstr = "exchangeable", std.err = "san.se")
```

Table 87 below shows the parameter estimates for both GEE models.

```
TCap <- paste("Generalized Estimating Equation Models for Violent, Non-Sexual Crimes Incident Counts",
              "Predicted by When Incidents Occurred")
FN <- paste ("\\item * p < .05, null hypothesis value outside 95\\% confidence",
             "interval based on z-score.",
             "\\\\\n\\item Note.",
```

```
             "The data contain up to 3 longitudinal observations (one per",
             "period) for each offender (i.e., cluster).",
             "The reference priod for When was after the testing window.",
             "Both models used a log link function, exchangeable correlation",
             "structure, a Poisson distribution, and robust standard errors",
             "(sandwich estimator).",
             "Model 1 did not use an offset.",
             "Model 2 used log(Years) as an offset term.")

texreg(list(HXCat12_Violent_gee1, HXCat12_Violent_gee2), booktabs = TRUE,
       dcolumn = TRUE, threeparttable = TRUE, fontsize = "normalsize",
       table = TRUE, use.packages = FALSE, ci.force = TRUE,
       label = "tab:HXCat12_Violent_texreg",
       stars = 0.05, caption = TCap, custom.note = FN)
```

	Model 1	Model 2
(Intercept)	0.55^*	-2.35^*
	$[0.48; 0.61]$	$[-2.41; -2.28]$
WhenBefore	-0.79^*	-0.36^*
	$[-0.90; -0.68]$	$[-0.46; -0.27]$
WhenDuring	-2.56^*	1.46^*
	$[-2.74; -2.37]$	$[1.27; 1.65]$
Scale parameter: gamma	1.53	1.53
Scale parameter: SE	0.08	0.08
Correlation parameter: alpha	−0.02	0.02
Correlation parameter: SE	0.02	0.02
Num. obs.	3175	3175
Num. clust.	1082	1082

* p < .05, null hypothesis value outside 95% confidence interval based on z-score.

Note. The data contain up to 3 longitudinal observations (one per period) for each offender (i.e., cluster). The reference priod for When was after the testing window. Both models used a log link function, exchangeable correlation structure, a Poisson distribution, and robust standard errors (sandwich estimator). Model 1 did not use an offset. Model 2 used log(Years) as an offset term.

Table 87: Generalized Estimating Equation Models for Violent, Non-Sexual Crimes Incident Counts Predicted by When Incidents Occurred

5.8.2.2 Estimated Marginal Means Table 88 provides the estimated marginal means obtained from Models 1 and 2, along with corresponding confidence intervals. The units of measurement for those means are counts per offender for Model 1 (which ignore period duration) and counts per offender-year for Model 2 (which adjust for duration and are thus considered annual incidence rates).

```
# Table caption.
TCap <- paste("Marginal Means of Violent, Non-Sexual Crimes Incident Counts and Incidence Rates",
              "By When Incidents Occurred")

# Table footnote.
FN <- paste("Model 1 did not use an offset, so means are counts.",
            "Model 2 used log(Years) as an offset term, so means are annual",
            "incidence rates.",
            "Confidence intervals are based on z-scores.")

HXCat12_Violent_emmeans1 <- emmeans(HXCat12_Violent_gee1, specs = pairwise ~ When,
                            type = "response", ref = "During")
HXCat12_Violent_emmeans2 <- emmeans(HXCat12_Violent_gee2, specs = pairwise ~ When,
                            type = "response", ref = "During")
```

```
as_tibble(HXCat12_Violent_emmeans1$emmeans) %>%
  full_join(x = ., y = as_tibble(HXCat12_Violent_emmeans2$emmeans)) %>%
  mutate(Model = c(1, 1, 1, 2, 2, 2),
         When = factor(When, levels = c("Before", "During", "After"))) %>%
  arrange(Model, When) %>%
  select(When, rate, SE, df, asymp.LCL, asymp.UCL) %>%
  kable(format = "latex", booktabs = TRUE, caption = TCap, digits = 2,
        col.names = c("When", "Rate", "SE", "df", "LCL", "UCL")) %>%
  kable_styling() %>%
  add_header_above(c(" " = 4, "95% CI" = 2)) %>%
  group_rows(group_label = "Model 1: Mean per offender", start_row = 1,
             end_row = 3, italic = TRUE) %>%
  group_rows(group_label = "Model 2: Mean per offender-year",
             start_row = 4, end_row = 6, italic = TRUE) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
## Joining, by = c("When", "rate", "SE", "df", "asymp.LCL", "asymp.UCL")
```

				95% CI	
When	Rate	SE	df	LCL	UCL
Model 1: Mean per offender					
Before	0.78	0.03	Inf	0.72	0.85
During	0.13	0.01	Inf	0.11	0.16
After	1.73	0.06	Inf	1.62	1.84
Model 2: Mean per offender-year					
Before	0.67	0.03	Inf	0.62	0.73
During	4.17	0.37	Inf	3.51	4.96
After	0.97	0.03	Inf	0.91	1.03

Note: Model 1 did not use an offset, so means are counts. Model 2 used log(Years) as an offset term, so means are annual incidence rates. Confidence intervals are based on z-scores.

Table 88: Marginal Means of Violent, Non-Sexual Crimes Incident Counts and Incidence Rates By When Incidents Occurred

5.8.2.3 Contrasts for Ratios of Estimated Marginal Means Finally, we also estimated a set of contrasts to compare the marginal means from Models 1 and 2 (Table 89). Each of these contrasts estimates the ratios of a pair of means rather than a raw difference between means. This is because of the nature of the underlying Poisson GEE model.

```
TCap <- paste("Contrasts Estimating Ratios of Marginal Means for Violent, Non-Sexual Crimes",
              "Incident Counts and Incidence Rates By When Incidents",
              "Occurred")

# Table footnote.
FN <- paste("Model 1 did not use an offset, so means are counts.",
            "Model 2 used log(Years) as an offset term, so means are annual",
            "incidence rates.",
            "These contrasts estimate ratios of those quantities and used",
            "Tukey's method to adjust for multiple comparisons.",
            "Confidence intervals are based on z-scores.")

as_tibble(confint(HXCat12_Violent_emmeans1$contrasts)) %>%
  full_join(x = ., y = as_tibble(confint(HXCat12_Violent_emmeans2$contrasts))) %>%
  kable(format = "latex", booktabs = TRUE, caption = TCap, digits = 2,
        col.names = c("Contrast", "Ratio", "SE", "df", "LCL", "UCL")) %>%
  kable_styling() %>%
  add_header_above(c(" " = 4, "95% CI" = 2)) %>%
```

```
  group_rows(group_label = "Model 1: Ratios of means per offender",
             start_row = 1, end_row = 3, italic = TRUE) %>%
  group_rows(group_label = "Model 2: Ratios of means per offender-year",
             start_row = 4, end_row = 6, italic = TRUE) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
## Joining, by = c("contrast", "ratio", "SE", "df", "asymp.LCL", "asymp.UCL")
```

				95% CI	
Contrast	Ratio	SE	df	LCL	UCL
Model 1: Ratios of means per offender					
After / Before	2.20	0.12	Inf	1.93	2.50
After / During	12.88	1.24	Inf	10.28	16.13
Before / During	5.86	0.56	Inf	4.68	7.34
Model 2: Ratios of means per offender-year					
After / Before	1.44	0.07	Inf	1.28	1.62
After / During	0.23	0.02	Inf	0.19	0.29
Before / During	0.16	0.02	Inf	0.13	0.20

Note:
Model 1 did not use an offset, so means are counts. Model 2 used log(Years) as an offset term, so means are annual incidence rates. These contrasts estimate ratios of those quantities and used Tukey's method to adjust for multiple comparisons. Confidence intervals are based on z-scores.

Table 89: Contrasts Estimating Ratios of Marginal Means for Violent, Non-Sexual Crimes Incident Counts and Incidence Rates By When Incidents Occurred

5.8.3 Property Crimes

Among these 1,082 suspected sexual offenders, there were 717 (66%) offenders who had criminal histories containing at least one incident associated with an arrest, charge, or conviction for a property crime (arson, burglary, or larceny/theft/fraud). Table 90 summarizes the counts of unique sexual crimes incidents overall and by when the incident occurred, while Table 91 shows the frequency distributions underlying those summaries.

```
# Table caption.
TCap <- paste("Offender-Level Property Crimes Incident Counts (Overall and By When",
              "Incident Occurred)")

# Summarize crime category counts (both overall and broken down by IWhen).
IDNEW %>%
  select(HXCat12_Property, HXCat12_Property_Before, HXCat12_Property_During,
         HXCat12_Property_After) %>%
  describe(., quant=c(.25, .50, .75)) %>%
  bind_cols(data.frame(Variable = TRowLabs), .) %>%
  select(Variable, n, mean, sd, min, max, skew, kurtosis, Q0.25, Q0.5, Q0.75) %>%
  kable(., format = "latex", booktabs = TRUE, digits = 2, row.names = FALSE,
        col.names = CNames, caption = TCap,
        linesep = c('', '', '', '\\addlinespace')) %>%
  column_spec(column = 1, width = "6 cm") %>%
  footnote(general = FN1, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
# Table caption.
TCap <- paste("Frequency Distributions for Number of Property Crimes Incidents ",
              "for Which Offender Was Arrested, Charged, or Convicted (Overall",
```

Variable	N	Mean	SD	Min	Max	Skew	Kurtosis	Q25	Q50	Q75
Incident count (arrested, charged, or convicted)	1082	2.35	3.31	0	38	3.32	19.93	0	1	3
...Before testing window	1041	0.99	1.69	0	12	2.57	8.34	0	0	1
...During testing window	1052	0.08	0.40	0	7	9.52	132.13	0	0	0
...After testing window	1082	1.31	2.51	0	37	5.05	47.37	0	0	2

Note: Only offenders with valid testing window start dates are included. Only incidents where the offender was arrested for, charged with, or convicted of at least one offense from one of the 12 main crime categories were counted. Before, during, and after refer to when the incident associated with the record occurred, not when the arrest date, charge date, or conviction date occurred relative to the testing window.

Table 90: Offender-Level Property Crimes Incident Counts (Overall and By When Incident Occurred)

```
              "and By When Incident Occurred)")

# Get frequency distributions.
IDNEW %>%
  select(OID, HXCat12_Property, HXCat12_Property_Before, HXCat12_Property_During,
         HXCat12_Property_After) %>%
  rename(All = HXCat12_Property, Before = HXCat12_Property_Before,
         During = HXCat12_Property_During, After = HXCat12_Property_After) %>%
  pivot_longer(cols = -OID, names_to = "When", values_to = "Value") %>%
  group_by(When, Value) %>%
  count() %>%
  arrange(Value) %>%
  pivot_wider(names_from = When, values_from = n, values_fill = 0) %>%
  mutate(All_p = 100*All/N_All,
         All_v = if_else(is.na(Value),
                         true = as.numeric(NA),
                         false = 100*All/N_All),
         Before_p = 100*Before/N_All,
         Before_v = if_else(is.na(Value),
                            true = as.numeric(NA),
                            false = 100*Before/N_Before),
         During_p = 100*During/N_All,
         During_v = if_else(is.na(Value),
                            true = as.numeric(NA),
                            false = 100*During/N_During),
         After_p = 100*After/N_All,
         After_v = if_else(is.na(Value),
                           true = as.numeric(NA),
                           false = 100*After/N_After)) %>%
  select(Value, All, All_p, All_v, Before, Before_p, Before_v, During,
         During_p, During_v, After, After_p, After_v) %>%
  kable(format = "latex", booktabs = TRUE, digits = 2,
        col.names = c("Value", rep(c("N", "%", "Valid %"), 4)),
        caption = TCap) %>%
  kable_styling() %>%
  add_header_above(c(" " = 1, "Overall" = 3, "Before TW" = 3, "During TW" = 3,
                     "After TW" = 3)) %>%
  footnote(general = FN5, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

	Overall			Before TW			During TW			After TW		
Value	N	%	Valid %	N	%	Valid %	N	%	Valid %	N	%	Valid %
0	365	33.73	33.73	603	55.73	57.93	984	90.94	93.54	577	53.33	53.33
1	222	20.52	20.52	195	18.02	18.73	59	5.45	5.61	219	20.24	20.24
2	149	13.77	13.77	100	9.24	9.61	6	0.55	0.57	100	9.24	9.24
3	96	8.87	8.87	60	5.55	5.76	1	0.09	0.10	61	5.64	5.64
4	76	7.02	7.02	34	3.14	3.27	0	0.00	0.00	52	4.81	4.81
5	42	3.88	3.88	19	1.76	1.83	0	0.00	0.00	17	1.57	1.57
6	34	3.14	3.14	11	1.02	1.06	1	0.09	0.10	12	1.11	1.11
7	27	2.50	2.50	7	0.65	0.67	1	0.09	0.10	14	1.29	1.29
8	17	1.57	1.57	3	0.28	0.29	0	0.00	0.00	6	0.55	0.55
9	10	0.92	0.92	3	0.28	0.29	0	0.00	0.00	4	0.37	0.37
10	10	0.92	0.92	4	0.37	0.38	0	0.00	0.00	7	0.65	0.65
11	9	0.83	0.83	1	0.09	0.10	0	0.00	0.00	3	0.28	0.28
12	9	0.83	0.83	1	0.09	0.10	0	0.00	0.00	3	0.28	0.28
13	4	0.37	0.37	0	0.00	0.00	0	0.00	0.00	3	0.28	0.28
14	2	0.18	0.18	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
15	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
16	3	0.28	0.28	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
17	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
18	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
22	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
23	2	0.18	0.18	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
27	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
37	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
38	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
NA	0	0.00	NA	41	3.79	NA	30	2.77	NA	0	0.00	NA

Note: N = 1,082. TW, earliest testing window. Only offenders with valid testing window start dates are included. Before, during, and after refer to when the incident associated with the record occurred, not when the arrest, charge, or conviction date occurred relative to the testing window. For example, an offender could be arrested during the testing window for an incident that occurred before it began. That would show up here in the before column. An incident is counted if there are any arrest, charge, or conviction records (any one type, or any combination of them will suffice) for the specified crime category associated with it.

Table 91: Frequency Distributions for Number of Property Crimes Incidents for Which Offender Was Arrested, Charged, or Convicted (Overall and By When Incident Occurred)

5.8.3.1 GEE Models To more closely examine the property crimes incident counts, we fitted two generalized estimating equation (GEE) models to the long-format *IDNEWL* data, after dropping rows for unobserved periods (*Years* = 0) because the counts are missing for those observations.

Model 1 used *When* as a predictor of the period-specific counts, but did not adjust for the period durations. Model 2 supplemented *When* as a predictor with an offset term for log(*Years*) to adjust for the period durations. Both models used Poisson distributions, an exchangeable correlation structure, a log link function, and robust standard errors based on the sandwich estimator.

```
HXCat12_Property_gee1 <- IDNEWL %>%
  filter(Years > 0) %>%
  geeglm(HXCat12_Property ~ When,
         family = poisson(link = "log"), data = .,
         id = OID, corstr = "exchangeable", std.err = "san.se")

HXCat12_Property_gee2 <- IDNEWL %>%
  filter(Years > 0) %>%
  geeglm(HXCat12_Property ~ When + offset(log(Years)),
```

```
        family = poisson(link = "log"), data = .,
        id = OID, corstr = "exchangeable", std.err = "san.se")
```

Table 92 below shows the parameter estimates for both GEE models.

```
TCap <- paste("Generalized Estimating Equation Models for Property Crimes Incident Counts",
              "Predicted by When Incidents Occurred")
FN <- paste ("\\item * p < .05, null hypothesis value outside 95\\% confidence",
             "interval based on z-score.",
             "\\\\\\n\\item Note.",
             "The data contain up to 3 longitudinal observations (one per",
             "period) for each offender (i.e., cluster).",
             "The reference priod for When was after the testing window.",
             "Both models used a log link function, exchangeable correlation",
             "structure, a Poisson distribution, and robust standard errors",
             "(sandwich estimator).",
             "Model 1 did not use an offset.",
             "Model 2 used log(Years) as an offset term.")

texreg(list(HXCat12_Property_gee1, HXCat12_Property_gee2), booktabs = TRUE,
       dcolumn = TRUE, threeparttable = TRUE, fontsize = "normalsize",
       table = TRUE, use.packages = FALSE, ci.force = TRUE,
       label = "tab:HXCat12_Property_texreg",
       stars = 0.05, caption = TCap, custom.note = FN)
```

	Model 1	Model 2
(Intercept)	0.27*	−2.62*
	[0.16; 0.39]	[−2.74; −2.51]
WhenBefore	−0.28*	0.15*
	[−0.42; −0.14]	[0.01; 0.28]
WhenDuring	−2.76*	1.25*
	[−3.07; −2.46]	[0.94; 1.55]
Scale parameter: gamma	3.22	3.01
Scale parameter: SE	0.34	0.30
Correlation parameter: alpha	0.12	0.14
Correlation parameter: SE	0.03	0.03
Num. obs.	3175	3175
Num. clust.	1082	1082

* p < .05, null hypothesis value outside 95% confidence interval based on z-score.

Note. The data contain up to 3 longitudinal observations (one per period) for each offender (i.e., cluster). The reference priod for When was after the testing window. Both models used a log link function, exchangeable correlation structure, a Poisson distribution, and robust standard errors (sandwich estimator). Model 1 did not use an offset. Model 2 used log(Years) as an offset term.

Table 92: Generalized Estimating Equation Models for Property Crimes Incident Counts Predicted by When Incidents Occurred

5.8.3.2 Estimated Marginal Means Table 93 provides the estimated marginal means obtained from Models 1 and 2, along with corresponding confidence intervals. The units of measurement for those means are counts per offender for Model 1 (which ignore period duration) and counts per offender-year for Model 2 (which adjust for duration and are thus considered annual incidence rates).

```
# Table caption.
TCap <- paste("Marginal Means of Property Crimes Incident Counts and Incidence Rates",
              "By When Incidents Occurred")

# Table footnote.
FN <- paste("Model 1 did not use an offset, so means are counts.",
            "Model 2 used log(Years) as an offset term, so means are annual",
            "incidence rates.",
            "Confidence intervals are based on z-scores.")

HXCat12_Property_emmeans1 <- emmeans(HXCat12_Property_gee1, specs = pairwise ~ When,
                             type = "response", ref = "During")
HXCat12_Property_emmeans2 <- emmeans(HXCat12_Property_gee2, specs = pairwise ~ When,
                             type = "response", ref = "During")

as_tibble(HXCat12_Property_emmeans1$emmeans) %>%
  full_join(x = ., y = as_tibble(HXCat12_Property_emmeans2$emmeans)) %>%
  mutate(Model = c(1, 1, 1, 2, 2, 2),
         When = factor(When, levels = c("Before", "During", "After"))) %>%
  arrange(Model, When) %>%
  select(When, rate, SE, df, asymp.LCL, asymp.UCL) %>%
  kable(format = "latex", booktabs = TRUE, caption = TCap, digits = 2,
        col.names = c("When", "Rate", "SE", "df", "LCL", "UCL")) %>%
  kable_styling() %>%
  add_header_above(c(" " = 4, "95% CI" = 2)) %>%
  group_rows(group_label = "Model 1: Mean per offender", start_row = 1,
             end_row = 3, italic = TRUE) %>%
  group_rows(group_label = "Model 2: Mean per offender-year",
             start_row = 4, end_row = 6, italic = TRUE) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
## Joining, by = c("When", "rate", "SE", "df", "asymp.LCL", "asymp.UCL")
```

				95% CI	
When	Rate	SE	df	LCL	UCL
Model 1: Mean per offender					
Before	0.99	0.05	Inf	0.90	1.10
During	0.08	0.01	Inf	0.06	0.11
After	1.31	0.08	Inf	1.17	1.47
Model 2: Mean per offender-year					
Before	0.85	0.04	Inf	0.77	0.94
During	2.55	0.38	Inf	1.90	3.43
After	0.74	0.04	Inf	0.66	0.82

Note: Model 1 did not use an offset, so means are counts. Model 2 used log(Years) as an offset term, so means are annual incidence rates. Confidence intervals are based on z-scores.

Table 93: Marginal Means of Property Crimes Incident Counts and Incidence Rates By When Incidents Occurred

5.8.3.3 Contrasts for Ratios of Estimated Marginal Means Finally, we also estimated a set of contrasts to compare the marginal means from Models 1 and 2 (Table 94). Each of these contrasts estimates the ratios of a pair of means rather than a raw difference between means. This is because of the nature of the underlying Poisson GEE model.

```
TCap <- paste("Contrasts Estimating Ratios of Marginal Means for Property Crimes",
              "Incident Counts and Incidence Rates By When Incidents",
              "Occurred")
```

```
# Table footnote.
FN <- paste("Model 1 did not use an offset, so means are counts.",
            "Model 2 used log(Years) as an offset term, so means are annual",
            "incidence rates.",
            "These contrasts estimate ratios of those quantities and used",
            "Tukey's method to adjust for multiple comparisons.",
            "Confidence intervals are based on z-scores.")

as_tibble(confint(HXCat12_Property_emmeans1$contrasts)) %>%
  full_join(x = ., y = as_tibble(confint(HXCat12_Property_emmeans2$contrasts))) %>%
  kable(format = "latex", booktabs = TRUE, caption = TCap, digits = 2,
        col.names = c("Contrast", "Ratio", "SE", "df", "LCL", "UCL")) %>%
  kable_styling() %>%
  add_header_above(c(" " = 4, "95% CI" = 2)) %>%
  group_rows(group_label = "Model 1: Ratios of means per offender",
             start_row = 1, end_row = 3, italic = TRUE) %>%
  group_rows(group_label = "Model 2: Ratios of means per offender-year",
             start_row = 4, end_row = 6, italic = TRUE) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
## Joining, by = c("contrast", "ratio", "SE", "df", "asymp.LCL", "asymp.UCL")
```

				95% CI	
Contrast	Ratio	SE	df	LCL	UCL
Model 1: Ratios of means per offender					
After / Before	1.32	0.09	Inf	1.12	1.56
After / During	15.81	2.47	Inf	10.97	22.79
Before / During	11.99	1.84	Inf	8.36	17.19
Model 2: Ratios of means per offender-year					
After / Before	0.86	0.06	Inf	0.73	1.01
After / During	0.29	0.05	Inf	0.20	0.42
Before / During	0.33	0.05	Inf	0.23	0.48

Note:
Model 1 did not use an offset, so means are counts. Model 2 used log(Years) as an offset term, so means are annual incidence rates. These contrasts estimate ratios of those quantities and used Tukey's method to adjust for multiple comparisons. Confidence intervals are based on z-scores.

Table 94: Contrasts Estimating Ratios of Marginal Means for Property Crimes Incident Counts and Incidence Rates By When Incidents Occurred

5.8.4 Other Crimes

Among these 1,082 suspected sexual offenders, there were 717 (66%) offenders who had criminal histories containing at least one incident associated with an arrest, charge, or conviction for other crimes (drug crimes or traffic and ordinances). Table 95 summarizes the counts of unique sexual crimes incidents overall and by when the incident occurred, while Table 96 shows the frequency distributions underlying those summaries.

```
# Table caption.
TCap <- paste("Offender-Level Other Crimes Incident Counts (Overall and By When",
              "Incident Occurred)")

# Summarize crime category counts (both overall and broken down by IWhen).
IDNEW %>%
  select(HXCat12_Other, HXCat12_Other_Before, HXCat12_Other_During,
         HXCat12_Other_After) %>%
```

```
  describe(., quant=c(.25, .50, .75)) %>%
  bind_cols(data.frame(Variable = TRowLabs), .) %>%
  select(Variable, n, mean, sd, min, max, skew, kurtosis, Q0.25, Q0.5, Q0.75) %>%
  kable(., format = "latex", booktabs = TRUE, digits = 2, row.names = FALSE,
        col.names = CNames, caption = TCap,
        linesep = c('', '', '', '\\addlinespace')) %>%
  column_spec(column = 1, width = "6 cm") %>%
  footnote(general = FN1, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

Variable	N	Mean	SD	Min	Max	Skew	Kurtosis	Q25	Q50	Q75
Incident count (arrested, charged, or convicted)	1082	3.51	3.40	0	40	2.16	13.03	1	3	5
...Before testing window	1041	0.86	1.29	0	10	2.10	5.97	0	0	1
...During testing window	1052	0.07	0.27	0	2	4.21	18.62	0	0	0
...After testing window	1082	2.62	3.11	0	40	2.78	20.39	0	2	4

Note: Only offenders with valid testing window start dates are included. Only incidents where the offender was arrested for, charged with, or convicted of at least one offense from one of the 12 main crime categories were counted. Before, during, and after refer to when the incident associated with the record occurred, not when the arrest date, charge date, or conviction date occurred relative to the testing window.

Table 95: Offender-Level Other Crimes Incident Counts (Overall and By When Incident Occurred)

```
# Table caption.
TCap <- paste("Frequency Distributions for Number of Other Crimes Incidents ",
              "for Which Offender Was Arrested, Charged, or Convicted (Overall",
              "and By When Incident Occurred)")

# Get frequency distributions.
IDNEW %>%
  select(OID, HXCat12_Other, HXCat12_Other_Before, HXCat12_Other_During,
         HXCat12_Other_After) %>%
  rename(All = HXCat12_Other, Before = HXCat12_Other_Before,
         During = HXCat12_Other_During, After = HXCat12_Other_After) %>%
  pivot_longer(cols = -OID, names_to = "When", values_to = "Value") %>%
  group_by(When, Value) %>%
  count() %>%
  arrange(Value) %>%
  pivot_wider(names_from = When, values_from = n, values_fill = 0) %>%
  mutate(All_p = 100*All/N_All,
         All_v = if_else(is.na(Value),
                         true = as.numeric(NA),
                         false = 100*All/N_All),
         Before_p = 100*Before/N_All,
         Before_v = if_else(is.na(Value),
                            true = as.numeric(NA),
                            false = 100*Before/N_Before),
         During_p = 100*During/N_All,
         During_v = if_else(is.na(Value),
                            true = as.numeric(NA),
                            false = 100*During/N_During),
         After_p = 100*After/N_All,
         After_v = if_else(is.na(Value),
                           true = as.numeric(NA),
                           false = 100*After/N_After)) %>%
  select(Value, All, All_p, All_v, Before, Before_p, Before_v, During,
         During_p, During_v, After, After_p, After_v) %>%
  kable(format = "latex", booktabs = TRUE, digits = 2,
        col.names = c("Value", rep(c("N", "%", "Valid %"), 4)),
        caption = TCap) %>%
  kable_styling() %>%
  add_header_above(c(" " = 1, "Overall" = 3, "Before TW" = 3, "During TW" = 3,
                     "After TW" = 3)) %>%
  footnote(general = FN5, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

	Overall			Before TW			During TW			After TW		
Value	N	%	Valid %	N	%	Valid %	N	%	Valid %	N	%	Valid %
0	196	18.11	18.11	579	53.51	55.62	986	91.13	93.73	322	29.76	29.76
1	159	14.70	14.70	236	21.81	22.67	60	5.55	5.70	188	17.38	17.38
2	147	13.59	13.59	112	10.35	10.76	6	0.55	0.57	149	13.77	13.77
3	142	13.12	13.12	66	6.10	6.34	0	0.00	0.00	108	9.98	9.98
4	112	10.35	10.35	27	2.50	2.59	0	0.00	0.00	88	8.13	8.13
5	88	8.13	8.13	9	0.83	0.86	0	0.00	0.00	65	6.01	6.01
6	64	5.91	5.91	7	0.65	0.67	0	0.00	0.00	50	4.62	4.62
7	39	3.60	3.60	3	0.28	0.29	0	0.00	0.00	30	2.77	2.77
8	46	4.25	4.25	1	0.09	0.10	0	0.00	0.00	24	2.22	2.22
9	23	2.13	2.13	0	0.00	0.00	0	0.00	0.00	15	1.39	1.39
10	22	2.03	2.03	1	0.09	0.10	0	0.00	0.00	21	1.94	1.94
11	14	1.29	1.29	0	0.00	0.00	0	0.00	0.00	11	1.02	1.02
12	15	1.39	1.39	0	0.00	0.00	0	0.00	0.00	3	0.28	0.28
13	6	0.55	0.55	0	0.00	0.00	0	0.00	0.00	2	0.18	0.18
14	2	0.18	0.18	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
15	2	0.18	0.18	0	0.00	0.00	0	0.00	0.00	2	0.18	0.18
16	2	0.18	0.18	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
17	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
18	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
21	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
40	1	0.09	0.09	0	0.00	0.00	0	0.00	0.00	1	0.09	0.09
NA	0	0.00	NA	41	3.79	NA	30	2.77	NA	0	0.00	NA

Note: N = 1,082. TW, earliest testing window. Only offenders with valid testing window start dates are included. Before, during, and after refer to when the incident associated with the record occurred, not when the arrest, charge, or conviction date occurred relative to the testing window. For example, an offender could be arrested during the testing window for an incident that occurred before it began. That would show up here in the before column. An incident is counted if there are any arrest, charge, or conviction records (any one type, or any combination of them will suffice) for the specified crime category associated with it.

Table 96: Frequency Distributions for Number of Other Crimes Incidents for Which Offender Was Arrested, Charged, or Convicted (Overall and By When Incident Occurred)

5.8.4.1 GEE Models To more closely examine the other crimes incident counts, we fitted two generalized estimating equation (GEE) models to the long-format *IDNEWL* data, after dropping rows for unobserved periods (*Years* = 0) because the counts are missing for those observations.

Model 1 used *When* as a predictor of the period-specific counts, but did not adjust for the period durations. Model 2 supplemented *When* as a predictor with an offset term for log(*Years*) to adjust for the period durations. Both models used Poisson distributions, an exchangeable correlation structure, a log link function, and robust standard errors based on the sandwich estimator.

```
HXCat12_Other_gee1 <- IDNEWL %>%
  filter(Years > 0) %>%
  geeglm(HXCat12_Other ~ When,
         family = poisson(link = "log"), data = .,
         id = OID, corstr = "exchangeable", std.err = "san.se")

HXCat12_Other_gee2 <- IDNEWL %>%
  filter(Years > 0) %>%
  geeglm(HXCat12_Other ~ When + offset(log(Years)),
         family = poisson(link = "log"), data = .,
         id = OID, corstr = "exchangeable", std.err = "san.se")
```

Table 97 below shows the parameter estimates for both GEE models.

```
TCap <- paste("Generalized Estimating Equation Models for Other Crimes Incident Counts",
              "Predicted by When Incidents Occurred")
FN <- paste ("\\item * p < .05, null hypothesis value outside 95\\% confidence",
             "interval based on z-score.",
             "\\\\\n\\item Note.",
             "The data contain up to 3 longitudinal observations (one per",
             "period) for each offender (i.e., cluster).",
             "The reference priod for When was after the testing window.",
             "Both models used a log link function, exchangeable correlation",
             "structure, a Poisson distribution, and robust standard errors",
             "(sandwich estimator).",
             "Model 1 did not use an offset.",
             "Model 2 used log(Years) as an offset term.")

texreg(list(HXCat12_Other_gee1, HXCat12_Other_gee2), booktabs = TRUE,
       dcolumn = TRUE, threeparttable = TRUE, fontsize = "normalsize",
       table = TRUE, use.packages = FALSE, ci.force = TRUE,
       label = "tab:HXCat12_Other_texreg",
       stars = 0.05, caption = TCap, custom.note = FN)
```

	Model 1	Model 2
(Intercept)	0.96*	−1.93*
	[0.89; 1.03]	[−2.00; −1.86]
WhenBefore	−1.12*	−0.69*
	[−1.23; −1.00]	[−0.79; −0.58]
WhenDuring	−3.64*	0.33*
	[−3.89; −3.39]	[0.08; 0.59]
Scale parameter: gamma	2.25	2.40
Scale parameter: SE	0.21	0.21
Correlation parameter: alpha	0.04	0.10
Correlation parameter: SE	0.02	0.02
Num. obs.	3175	3175
Num. clust.	1082	1082

* p < .05, null hypothesis value outside 95% confidence interval based on z-score.

Note. The data contain up to 3 longitudinal observations (one per period) for each offender (i.e., cluster). The reference priod for When was after the testing window. Both models used a log link function, exchangeable correlation structure, a Poisson distribution, and robust standard errors (sandwich estimator). Model 1 did not use an offset. Model 2 used log(Years) as an offset term.

Table 97: Generalized Estimating Equation Models for Other Crimes Incident Counts Predicted by When Incidents Occurred

5.8.4.2 Estimated Marginal Means Table 98 provides the estimated marginal means obtained from Models 1 and 2, along with corresponding confidence intervals. The units of measurement for those means are counts per offender for Model 1 (which ignore period duration) and counts per offender-year for Model 2 (which adjust for duration and are thus considered annual incidence rates).

```
# Table caption.
TCap <- paste("Marginal Means of Other Crimes Incident Counts and Incidence Rates",
              "By When Incidents Occurred")

# Table footnote.
FN <- paste("Model 1 did not use an offset, so means are counts.",
            "Model 2 used log(Years) as an offset term, so means are annual",
            "incidence rates.",
```

```
           "Confidence intervals are based on z-scores.")

HXCat12_Other_emmeans1 <- emmeans(HXCat12_Other_gee1, specs = pairwise ~ When,
                          type = "response", ref = "During")
HXCat12_Other_emmeans2 <- emmeans(HXCat12_Other_gee2, specs = pairwise ~ When,
                          type = "response", ref = "During")

as_tibble(HXCat12_Other_emmeans1$emmeans) %>%
  full_join(x = ., y = as_tibble(HXCat12_Other_emmeans2$emmeans)) %>%
  mutate(Model = c(1, 1, 1, 2, 2, 2),
         When = factor(When, levels = c("Before", "During", "After"))) %>%
  arrange(Model, When) %>%
  select(When, rate, SE, df, asymp.LCL, asymp.UCL) %>%
  kable(format = "latex", booktabs = TRUE, caption = TCap, digits = 2,
        col.names = c("When", "Rate", "SE", "df", "LCL", "UCL")) %>%
  kable_styling() %>%
  add_header_above(c(" " = 4, "95% CI" = 2)) %>%
  group_rows(group_label = "Model 1: Mean per offender", start_row = 1,
             end_row = 3, italic = TRUE) %>%
  group_rows(group_label = "Model 2: Mean per offender-year",
             start_row = 4, end_row = 6, italic = TRUE) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
## Joining, by = c("When", "rate", "SE", "df", "asymp.LCL", "asymp.UCL")
```

				95% CI	
When	Rate	SE	df	LCL	UCL
Model 1: Mean per offender					
Before	0.86	0.04	Inf	0.78	0.94
During	0.07	0.01	Inf	0.05	0.09
After	2.62	0.09	Inf	2.44	2.81
Model 2: Mean per offender-year					
Before	0.74	0.03	Inf	0.68	0.81
During	2.05	0.26	Inf	1.60	2.64
After	1.47	0.05	Inf	1.37	1.58

Note: Model 1 did not use an offset, so means are counts. Model 2 used log(Years) as an offset term, so means are annual incidence rates. Confidence intervals are based on z-scores.

Table 98: Marginal Means of Other Crimes Incident Counts and Incidence Rates By When Incidents Occurred

5.8.4.3 Contrasts for Ratios of Estimated Marginal Means Finally, we also estimated a set of contrasts to compare the marginal means from Models 1 and 2 (Table 99). Each of these contrasts estimates the ratios of a pair of means rather than a raw difference between means. This is because of the nature of the underlying Poisson GEE model.

```
TCap <- paste("Contrasts Estimating Ratios of Marginal Means for Other Crimes",
              "Incident Counts and Incidence Rates By When Incidents",
              "Occurred")

# Table footnote.
FN <- paste("Model 1 did not use an offset, so means are counts.",
            "Model 2 used log(Years) as an offset term, so means are annual",
            "incidence rates.",
            "These contrasts estimate ratios of those quantities and used",
            "Tukey's method to adjust for multiple comparisons.",
            "Confidence intervals are based on z-scores.")
```

```
as_tibble(confint(HXCat12_Other_emmeans1$contrasts)) %>%
  full_join(x = ., y = as_tibble(confint(HXCat12_Other_emmeans2$contrasts))) %>%
  kable(format = "latex", booktabs = TRUE, caption = TCap, digits = 2,
        col.names = c("Contrast", "Ratio", "SE", "df", "LCL", "UCL")) %>%
  kable_styling() %>%
  add_header_above(c(" " = 4, "95% CI" = 2)) %>%
  group_rows(group_label = "Model 1: Ratios of means per offender",
             start_row = 1, end_row = 3, italic = TRUE) %>%
  group_rows(group_label = "Model 2: Ratios of means per offender-year",
             start_row = 4, end_row = 6, italic = TRUE) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

```
## Joining, by = c("contrast", "ratio", "SE", "df", "asymp.LCL", "asymp.UCL")
```

				95% CI	
Contrast	Ratio	SE	df	LCL	UCL
Model 1: Ratios of means per offender					
After / Before	3.06	0.18	Inf	2.66	3.50
After / During	38.20	4.85	Inf	28.38	51.44
Before / During	12.50	1.60	Inf	9.26	16.88
Model 2: Ratios of means per offender-year					
After / Before	1.99	0.11	Inf	1.75	2.26
After / During	0.72	0.09	Inf	0.53	0.97
Before / During	0.36	0.05	Inf	0.26	0.49

Note:
Model 1 did not use an offset, so means are counts. Model 2 used log(Years) as an offset term, so means are annual incidence rates. These contrasts estimate ratios of those quantities and used Tukey's method to adjust for multiple comparisons. Confidence intervals are based on z-scores.

Table 99: Contrasts Estimating Ratios of Marginal Means for Other Crimes Incident Counts and Incidence Rates By When Incidents Occurred

5.9 Before Period Summary

The content of this section was not reported in the manuscript for the sake of brevity. Table 100 remains here to document an analysis performed and provide it in a form consistent with the one used for some of the after period results.

```
TCap <- paste("Amount of Crime That Occurred Before the Testing Window")
FN <- paste0("Separately for each crime category, we counted offenders ",
             "arrested, charged, or convicted for at least one incident ",
             "occurring before the testing window. Total incidents are summed ",
             "across those offenders. The bottom four italicized broader ",
             "categories include offenders who were arrested, charged or ",
             "convicted for any of the more specific crime categories ",
             "nested below them, while the Any Crimes category includes ",
             "incidents associated with at least one of the 12 specific ",
             "crime categories. ",
             "The numerator for each percentage is the listed value for n; ",
             "the denominator is the sample size for the before period (N = ",
             format(N_Before, big.mark = ","),
             "). ",
             "Incidents associated with more than one kind of crime are ",
             "counted in all applicable specific categories, but only once ",
             "per applicable broader category. For example, an incident with ",
             "both a homicide and a robbery only counts once toward the total ",
```

```
            "incident count for the broader violent crimes category. TW, ",
            "testing window.")

IDNEWB %>%
  # Drops rows for crime category counts and individuals with no incidents
  filter(Variable != "HXCat12_Count") %>%
  # Drops rows for offenders with no incidents of a given type occuring after
  # the testing window.
  filter(Count > 0) %>%
  group_by(Variable, VLabel, VOrder) %>%
  summarise(N = n(),
            Pct = 100*N/N_Before,
            Incidents = sum(Count)) %>%
  ungroup() %>%
  arrange(VOrder) %>%
  select(VLabel, N, Pct, Incidents) %>%
  kable(., format = "latex", booktabs = TRUE, digits = 1, row.names = FALSE,
        col.names = c("Crime Category", "n", "%", "Total Incidents"),
        format.args = list(big.mark = ','), caption = TCap, linesep = "") %>%
  kable_styling() %>%
  add_header_above(c(" " = 1, "Offenders" = 2, " " = 1)) %>%
  row_spec(., row = c(1, 2, 5, 11, 15), italic = TRUE) %>%
  pack_rows(" ", start_row = 1, end_row = 1) %>%
  pack_rows(" ", start_row = 2, end_row = 4) %>%
  pack_rows(" ", start_row = 5, end_row = 10) %>%
  pack_rows(" ", start_row = 11, end_row = 14) %>%
  pack_rows(" ", start_row = 15, end_row = 17) %>%
  add_indent(., positions = c(3:4), level_of_indent = 1) %>%
  add_indent(., positions = c(6:10), level_of_indent = 1) %>%
  add_indent(., positions = c(12:14), level_of_indent = 1) %>%
  add_indent(., positions = c(16:17), level_of_indent = 1) %>%
  column_spec(., column = 2, width = "1.15cm") %>%
  column_spec(., column = 3, width = "1.15cm") %>%
  column_spec(., column = 4, width = "2cm") %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
           threeparttable = TRUE)
```

```
## `summarise()` has grouped output by 'Variable', 'VLabel'. You can override using the `.groups` argument.
```

Crime Category	Offenders		Total Incidents
	n	%	
Any Crimes (12 categories)	*765*	*73.5*	*2,590*
Sexual Crimes (2 categories)	*116*	*11.1*	*132*
Criminal Sexual Conduct (CSC)	112	10.8	127
Sex Crimes (non-CSC crimes)	12	1.2	13
Violent Non-Sexual Crimes (5 categories)	*485*	*46.6*	*817*
Assault, Domestic Violence	45	4.3	48
Assault, non-Domestic Violence	220	21.1	299
Homicide	25	2.4	27
Robbery	202	19.4	253
Weapons	242	23.2	311
Property Crimes (3 categories)	*438*	*42.1*	*1,035*
Arson	15	1.4	15
Burglary	184	17.7	307
Larceny, Theft, Fraud	379	36.4	770
Other Crimes (2 categories)	*462*	*44.4*	*892*
Drug Crimes	306	29.4	501
Traffic & Ordinances	252	24.2	403

Note: Separately for each crime category, we counted offenders arrested, charged, or convicted for at least one incident occurring before the testing window. Total incidents are summed across those offenders. The bottom four italicized broader categories include offenders who were arrested, charged or convicted for any of the more specific crime categories nested below them, while the Any Crimes category includes incidents associated with at least one of the 12 specific crime categories. The numerator for each percentage is the listed value for n; the denominator is the sample size for the before period (N = 1,041). Incidents associated with more than one kind of crime are counted in all applicable specific categories, but only once per applicable broader category. For example, an incident with both a homicide and a robbery only counts once toward the total incident count for the broader violent crimes category. TW, testing window.

Table 100: Amount of Crime That Occurred Before the Testing Window

5.10 During Period Summary

The content of this section was not reported in the manuscript for the sake of brevity. Table 101 remains here to document an analysis performed and provide it in a form consistent with the one used for some of the after period results.

```
TCap <- paste("Amount of Crime That Occurred During the Testing Window")
FN <- paste0("Separately for each crime category, we counted offenders ",
             "arrested, charged, or convicted for at least one incident ",
             "occurring during the testing window. Total incidents are summed ",
            "across those offenders. The bottom four italicized broader ",
             "categories include offenders who were arrested, charged or ",
             "convicted for any of the more specific crime categories ",
             "nested below them, while the Any Crimes category includes ",
             "incidents associated with at least one of the 12 specific ",
             "crime categories. ",
             "The numerator for each percentage is the listed value for n; ",
             "the denominator is the sample size for the during period (N = ",
```

```
            format(N_During, big.mark = ","), "). ",
            "Incidents associated with more than one kind of crime are ",
            "counted in all applicable specific categories, but only once ",
            "per applicable broader category. For example, an incident with ",
            "both a homicide and a robbery only counts once toward the total ",
            "incident count for the broader violent crimes category. TW, ",
            "testing window.")

IDNEWD %>%
  # Drops rows for crime category counts and individuals with no incidents
  filter(Variable != "HXCat12_Count") %>%
  # Drops rows for offenders with no incidents of a given type occuring after
  # the testing window.
  filter(Count > 0) %>%
  group_by(Variable, VLabel, VOrder) %>%
  summarise(N = n(),
            Pct = 100*N/N_During,
            Incidents = sum(Count)) %>%
  ungroup() %>%
  arrange(VOrder) %>%
  select(VLabel, N, Pct, Incidents) %>%
  kable(., format = "latex", booktabs = TRUE, digits = 1, row.names = FALSE,
        col.names = c("Crime Category", "n", "%", "Total Incidents"),
        format.args = list(big.mark = ','), caption = TCap, linesep = "") %>%
  kable_styling() %>%
  add_header_above(c(" " = 1, "Offenders" = 2, " " = 1)) %>%
 row_spec(., row = c(1, 2, 5, 11, 15), italic = TRUE) %>%
  pack_rows(" ", start_row = 1, end_row = 1) %>%
  pack_rows(" ", start_row = 2, end_row = 4) %>%
  pack_rows(" ", start_row = 5, end_row = 10) %>%
  pack_rows(" ", start_row = 11, end_row = 14) %>%
  pack_rows(" ", start_row = 15, end_row = 17) %>%
  add_indent(., positions = c(3:4), level_of_indent = 1) %>%
  add_indent(., positions = c(6:10), level_of_indent = 1) %>%
  add_indent(., positions = c(12:14), level_of_indent = 1) %>%
  add_indent(., positions = c(16:17), level_of_indent = 1) %>%
  column_spec(., column = 2, width = "1.15cm") %>%
  column_spec(., column = 3, width = "1.15cm") %>%
  column_spec(., column = 4, width = "2cm") %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
           threeparttable = TRUE)
```

```
## `summarise()` has grouped output by 'Variable', 'VLabel'. You can override using the `.groups` argument.
```

Crime Category	Offenders		Total Incidents
	n	%	
Any Crimes (12 categories)	*338*	*32.1*	*421*
Sexual Crimes (2 categories)	*201*	*19.1*	*232*
Criminal Sexual Conduct (CSC)	199	18.9	230
Sex Crimes (non-CSC crimes)	11	1.0	11
Violent Non-Sexual Crimes (5 categories)	*127*	*12.1*	*141*
Assault, Domestic Violence	18	1.7	20
Assault, non-Domestic Violence	48	4.6	51
Homicide	8	0.8	8
Robbery	49	4.7	55
Weapons	46	4.4	47
Property Crimes (3 categories)	*68*	*6.5*	*87*
Arson	1	0.1	2
Burglary	26	2.5	40
Larceny, Theft, Fraud	44	4.2	47
Other Crimes (2 categories)	*66*	*6.3*	*72*
Drug Crimes	35	3.3	38
Traffic & Ordinances	33	3.1	36

Note: Separately for each crime category, we counted offenders arrested, charged, or convicted for at least one incident occurring during the testing window. Total incidents are summed across those offenders. The bottom four italicized broader categories include offenders who were arrested, charged or convicted for any of the more specific crime categories nested below them, while the Any Crimes category includes incidents associated with at least one of the 12 specific crime categories. The numerator for each percentage is the listed value for n; the denominator is the sample size for the during period (N = 1,052). Incidents associated with more than one kind of crime are counted in all applicable specific categories, but only once per applicable broader category. For example, an incident with both a homicide and a robbery only counts once toward the total incident count for the broader violent crimes category. TW, testing window.

Table 101: Amount of Crime That Occurred During the Testing Window

5.11 After Period Summary

Next we summarize the amount of crime that occurred after the testing window. Separately for each crime category, Table 102 shows the number of offenders with one or more incidents from the given category. It also translates those offender counts into percentages relative to the total sample size and reports the total number of incidents associated with those offenders that occurred after the testing window ended.

```
TCap <- paste("Amount of Crime That Occurred After the Testing Window")
FN <- paste0("Separately for each crime category, we counted offenders ",
             "arrested, charged, or convicted for at least one incident " ,
             "occurring after the testing window. Total incidents are summed ",
             "across those offenders. The bottom four italicized broader ",
             "categories include offenders who were arrested, charged or ",
             "convicted for any of the more specific crime categories ",
             "nested below them, while the Any Crimes category includes ",
             "incidents associated with at least one of the 12 specific ",
```

```
            "crime categories. ",
            "The numerator for each percentage is the listed value for n; ",
            "the denominator is the sample size for the after period (N = ",
            format(N_After, big.mark = ","), "). ",
            "Incidents associated with more than one kind of crime are ",
            "counted in all applicable specific categories, but only once ",
            "per applicable broader category. For example, an incident with ",
            "both a homicide and a robbery only counts once toward the total ",
            "incident count for the broader violent crimes category. TW, ",
            "testing window.")

IDNEWA %>%
  # Drops rows for crime category counts and individuals with no incidents
  filter(Variable != "HXCat12_Count") %>%
  # Drops rows for offenders with no incidents of a given type occuring after
  # the testing window.
  filter(Count > 0) %>%
  group_by(Variable, VLabel, VOrder) %>%
  summarise(N = n(),
            Pct = 100*N/N_After,
            Incidents = sum(Count)) %>%
  ungroup() %>%
  arrange(VOrder) %>%
  select(VLabel, N, Pct, Incidents) %>%
  kable(., format = "latex", booktabs = TRUE, digits = 2, row.names = FALSE,
        col.names = c("Crime Category", "n", "%", "Total Incidents"),
        format.args = list(big.mark = ','), caption = TCap, linesep = "") %>%
  kable_styling() %>%
  add_header_above(c(" " = 1, "Offenders" = 2, " " = 1)) %>%
  row_spec(., row = c(1, 2, 5, 11, 15), italic = TRUE) %>%
  pack_rows(" ", start_row = 1, end_row = 1) %>%
  pack_rows(" ", start_row = 2, end_row = 4) %>%
  pack_rows(" ", start_row = 5, end_row = 10) %>%
  pack_rows(" ", start_row = 11, end_row = 14) %>%
  pack_rows(" ", start_row = 15, end_row = 17) %>%
  add_indent(., positions = c(3:4), level_of_indent = 1) %>%
  add_indent(., positions = c(6:10), level_of_indent = 1) %>%
  add_indent(., positions = c(12:14), level_of_indent = 1) %>%
  add_indent(., positions = c(16:17), level_of_indent = 1) %>%
  column_spec(., column = 2, width = "1.15cm") %>%
  column_spec(., column = 3, width = "1.15cm") %>%
  column_spec(., column = 4, width = "2cm") %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
           threeparttable = TRUE)
```

```
## `summarise()` has grouped output by 'Variable', 'VLabel'. You can override using the `.groups` argument.
```

Crime Category	Offenders		Total Incidents
	n	%	
Any Crimes (12 categories)	*992*	*91.68*	*5,577*
Sexual Crimes (2 categories)	*246*	*22.74*	*395*
Criminal Sexual Conduct (CSC)	207	19.13	320
Sex Crimes (non-CSC crimes)	74	6.84	92
Violent Non-Sexual Crimes (5 categories)	*762*	*70.43*	*1,867*
Assault, Domestic Violence	246	22.74	388
Assault, non-Domestic Violence	525	48.52	1,047
Homicide	74	6.84	86
Robbery	228	21.07	331
Weapons	386	35.67	565
Property Crimes (3 categories)	*505*	*46.67*	*1,419*
Arson	19	1.76	20
Burglary	210	19.41	379
Larceny, Theft, Fraud	420	38.82	1,077
Other Crimes (2 categories)	*760*	*70.24*	*2,836*
Drug Crimes	428	39.56	950
Traffic & Ordinances	670	61.92	2,024

Note: Separately for each crime category, we counted offenders arrested, charged, or convicted for at least one incident occurring after the testing window. Total incidents are summed across those offenders. The bottom four italicized broader categories include offenders who were arrested, charged or convicted for any of the more specific crime categories nested below them, while the Any Crimes category includes incidents associated with at least one of the 12 specific crime categories. The numerator for each percentage is the listed value for n; the denominator is the sample size for the after period (N = 1,082). Incidents associated with more than one kind of crime are counted in all applicable specific categories, but only once per applicable broader category. For example, an incident with both a homicide and a robbery only counts once toward the total incident count for the broader violent crimes category. TW, testing window.

Table 102: Amount of Crime That Occurred After the Testing Window

5.11.1 Incident Count Distributions By Crime Categeory

Figure 7 plots the distributions of the incident count variables for each of the 12 main crime categories during the period after the testing window. Because a few large outlier values make the details harder to see for the majority of the data (nearly all values are for 12 or fewer incidents), we also produced Figure 8, which zooms in on the 0-12 range of incident counts.

```
FCap <- paste("\\label{fig:plot_Crime_full_width}",
              "Distributions for Number of Incidents After the Testing Window",
              "for Which Offender Was Arrested, Charged, or Convicted,",
              "By Crime Category.",
              "Each panel shows a histogram above a strip plot with an",
              "overlaid boxplot. Histogram bar heights show the percentage of",
              "offenders who had a given number of incidents.",
              "The strip plots show a dot for each offender, with positions",
              "jittered to reduce overlap.")
```

```
# Graph after period frequency distributions.
IDNEWA %>%
  # Keep only the main 12 crime categories (drops the broader crime categories).
  filter(!is.na(Crime)) %>%
  # use y = -... to position boxplot and jitterplot below the histogram
  ggplot(data = ., aes(x = Count, y = -.4)) +
  # Add histogram with bar heights scaled to percentages via after_stat.
  geom_histogram(aes(y = after_stat(count/N_After)), binwidth = .5,
                 fill = "black") +
  # Add a jittered strip plot.
  geom_point(aes(), position = position_jitter(width = .25, height = .25),
             size = 1, color = "gray50") +
  # Overlay a boxplot on the strip plot from ggstance
  geom_boxplot(lwd = 1, width = .2, outlier.shape = NA) +
  # Merged those calls to one display
  guides(scale = "none") +
  # Split each crime category into a separate panel (facet)
  facet_wrap(~Crime, nrow = 6, strip.position = "top") +
  xlab("Incident Count After Testing Window") +
  # Remove y axis title label from the y axis
  theme(axis.title.y = element_blank()) +
  # Format the y axis ticks.
  scale_y_continuous(limits = c(-.65, 1.00),
                     breaks = c(0, .5, 1.0), labels = c("0%", "50%", "100%")) ->
  Figure_Crime_full_width

Figure_Crime_full_width
```

```
FCap <- paste("\\label{fig:plot_Crime_partial_width}",
              "Distributions for Number of Incidents After the Testing Window",
              "for Which Offender Was Arrested, Charged, or Convicted,",
              "By Crime Category.",
              "Each panel shows a histogram above a strip plot with an",
              "overlaid boxplot. Histogram bar heights show the percentage of",
              "offenders who had a given number of incidents.",
              "The strip plots show a dot for each offender, with positions",
              "jittered to reduce overlap. This plot has been zoomed",
              "in to show incident counts in the range 0-12, omitting a few",
              "outliers for (a) larceny, theft, fraud,",
              "(b) drug crimes, and (c) traffic and ordinances.")

Figure_Crime_full_width +
  scale_x_continuous(limits = c(-.6, 12.5)) ->
  Figure_Crime_partial_width

Figure_Crime_partial_width
```

```
## Warning: Removed 9 rows containing non-finite values (stat_bin).
```

```
## Warning: Removed 9 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 24 rows containing missing values (geom_bar).
```

```
## Warning: Removed 9 rows containing missing values (geom_point).
```

Figure 9 plots the distributions of the incident count variables for each of the four broader crime categories during the period after the testing window. Because a few large outlier values make the details harder to see for the majority of the data (nearly all values are for 15 or fewer incidents), we also produced Figure 10, which zooms in on the 0-15 range of incident counts.

```
FCap <- paste("\\label{fig:plot_BCrime_full_width}",
              "Distributions for Number of Incidents After the Testing Window",
              "for Which Offender Was Arrested, Charged, or Convicted,",
              "By Broader Crime Category.",
              "Each panel shows a histogram above a strip plot with an",
              "overlaid boxplot. Histogram bar heights show the percentage of",
              "offenders who had a given number of incidents.",
```

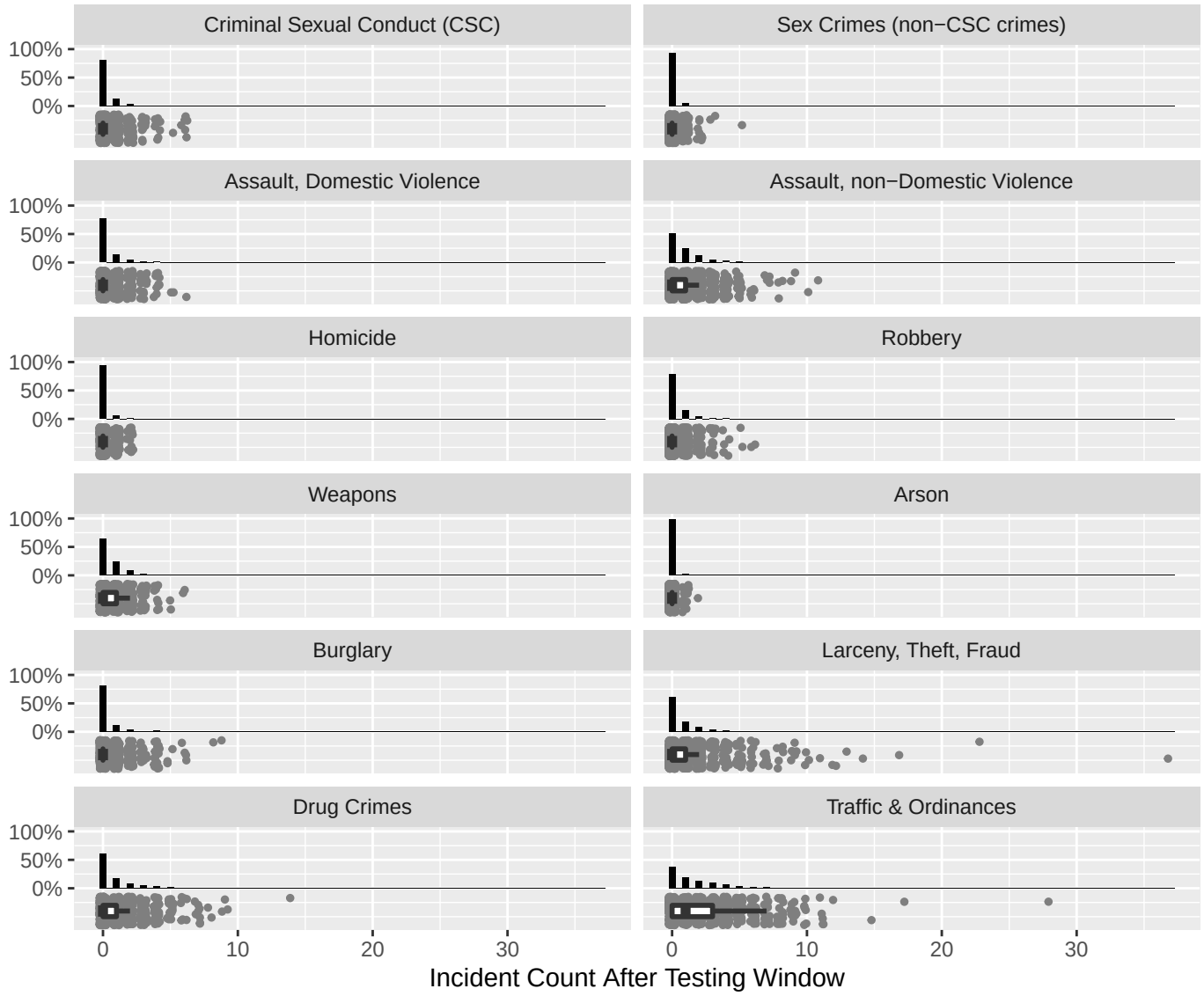

Figure 7: Distributions for Number of Incidents After the Testing Window for Which Offender Was Arrested, Charged, or Convicted, By Crime Category. Each panel shows a histogram above a strip plot with an overlaid boxplot. Histogram bar heights show the percentage of offenders who had a given number of incidents. The strip plots show a dot for each offender, with positions jittered to reduce overlap.

```
              "The strip plots show a dot for each offender, with positions",
              "jittered to reduce overlap.")

# Graph after period frequency distributions.
IDNEWA %>%
  # Keep only the 4 broader crime categories (drops 12 main crime categories).
  filter(!is.na(BCrime)) %>%
  # use y = -... to position boxplot and jitterplot below the histogram
  ggplot(data = ., aes(x = Count, y = -.4)) +
  # Add histogram with bar heights scaled to percentages via after_stat.
  geom_histogram(aes(y = after_stat(count/N_After)), binwidth = .5,
                 fill = "black") +
  # Add a jittered strip plot.
  geom_point(aes(), position = position_jitter(width = .25, height = .25),
             size = 1, color = "gray50") +
  # Overlay a boxplot on the strip plot from ggstance
  geom_boxplot(lwd = 1, width = .2, outlier.shape = NA) +
  # Merged those calls to one display
```

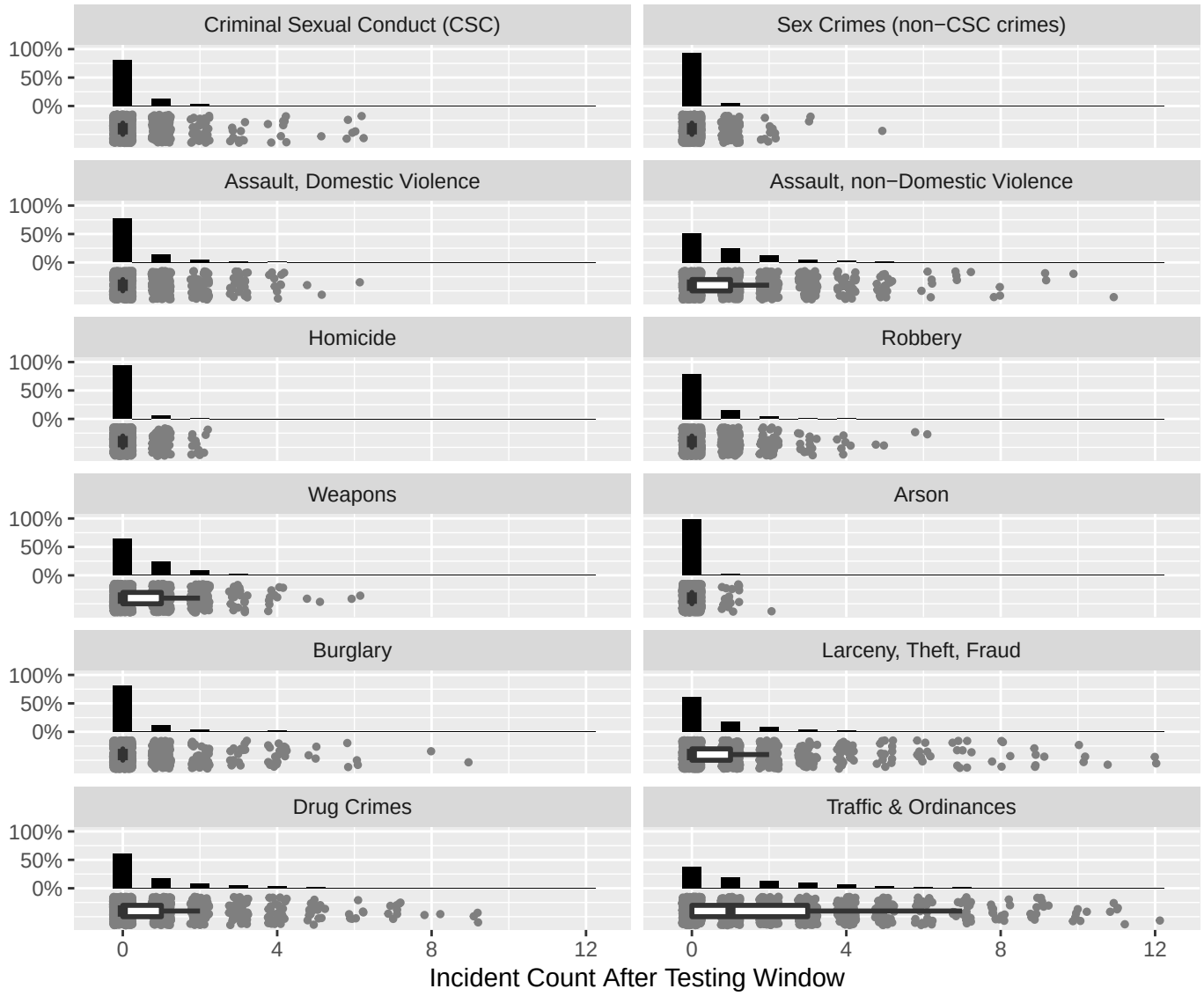

Figure 8: Distributions for Number of Incidents After the Testing Window for Which Offender Was Arrested, Charged, or Convicted, By Crime Category. Each panel shows a histogram above a strip plot with an overlaid boxplot. Histogram bar heights show the percentage of offenders who had a given number of incidents. The strip plots show a dot for each offender, with positions jittered to reduce overlap. This plot has been zoomed in to show incident counts in the range 0-12, omitting a few outliers for (a) larceny, theft, fraud, (b) drug crimes, and (c) traffic and ordinances.

```
  guides(scale = "none") +
  # Split each crime category into a separate panel (facet)
  facet_wrap(~BCrime, nrow = 4, strip.position = "top") +
  xlab("Incident Count After Testing Window") +
  # Remove y axis title label from the y axis
  theme(axis.title.y = element_blank()) +
  # Format the y axis ticks.
  scale_y_continuous(limits = c(-.65, 1.00),
                     breaks = c(0, .5, 1.0), labels = c("0%", "50%", "100%")) ->
  Figure_BCrime_full_width

Figure_BCrime_full_width
```

```
FCap <- paste("\\label{fig:plot_BCrime_partial_width}",
              "Distributions for Number of Incidents After the Testing Window",
              "for Which Offender Was Arrested, Charged, or Convicted,",
```

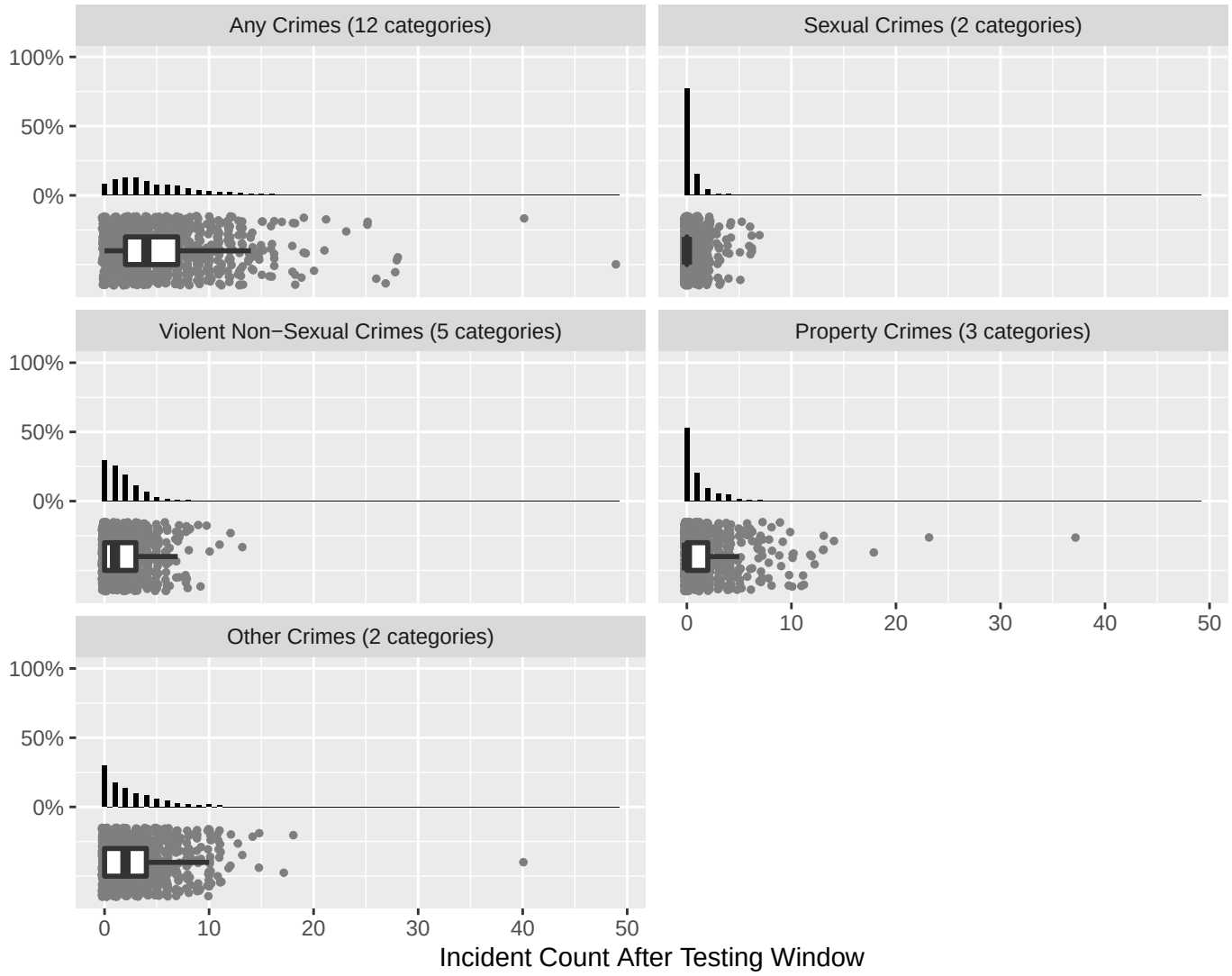

Figure 9: Distributions for Number of Incidents After the Testing Window for Which Offender Was Arrested, Charged, or Convicted, By Broader Crime Category. Each panel shows a histogram above a strip plot with an overlaid boxplot. Histogram bar heights show the percentage of offenders who had a given number of incidents. The strip plots show a dot for each offender, with positions jittered to reduce overlap.

```
              "By Broader Crime Category.",
              "Each panel shows a histogram above a strip plot with an",
              "overlaid boxplot. Histogram bar heights show the percentage of",
              "offenders who had a given number of incidents.",
              "The strip plots show a dot for each offender, with positions",
              "jittered to reduce overlap. This plot has been zoomed",
              "in to show incident counts in the range 0-15, omitting some",
              "larger values in the any crimes panel and some outliers for",
              "(a) property crimes and (b) other crimes.")

Figure_BCrime_full_width +
  scale_x_continuous(limits = c(-.6, 15.5)) ->
  Figure_BCrime_partial_width

Figure_BCrime_partial_width
```

```
## Warning: Removed 39 rows containing non-finite values (stat_bin).
```

```
## Warning: Removed 39 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 10 rows containing missing values (geom_bar).
```

```
## Warning: Removed 39 rows containing missing values (geom_point).
```

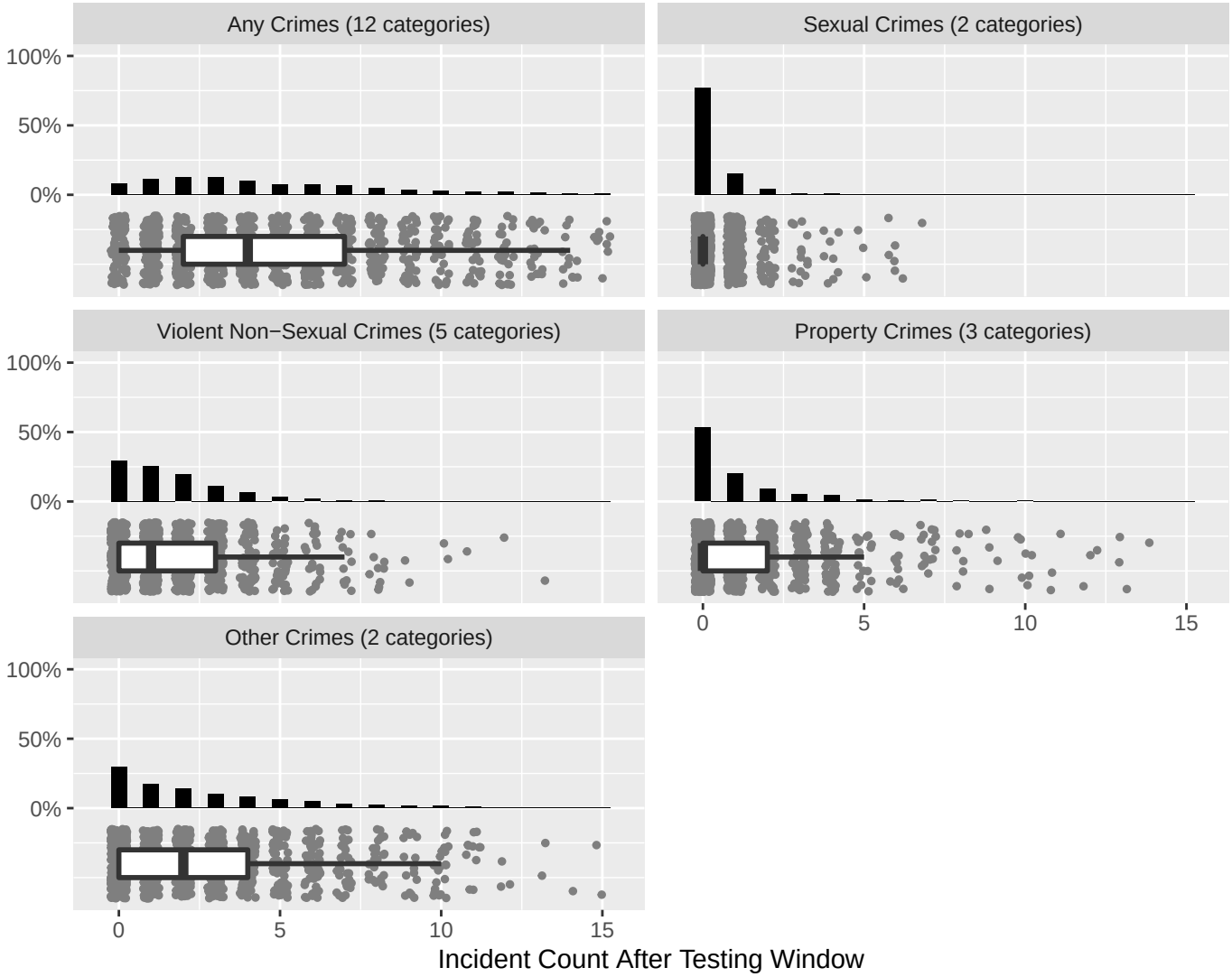

Figure 10: Distributions for Number of Incidents After the Testing Window for Which Offender Was Arrested, Charged, or Convicted, By Broader Crime Category. Each panel shows a histogram above a strip plot with an overlaid boxplot. Histogram bar heights show the percentage of offenders who had a given number of incidents. The strip plots show a dot for each offender, with positions jittered to reduce overlap. This plot has been zoomed in to show incident counts in the range 0-15, omitting some larger values in the any crimes panel and some outliers for (a) property crimes and (b) other crimes.

5.11.2 Years to Incident Date Distributions by Crime Category

Table 103 shows descriptive statistics about the distribution for *IYearsAfterTW*, which records how many years after the end of the testing window an incident actually occurred, broken down by crime category. This is particularly useful for identifying the median (Q50) and quartiles (Q25 and Q75), which clearly vary across types of crime.

```
# Table caption.
TCap <- paste("Years to Incident Date By Crime Category",
              "for Incidents After the Testing Window Where Offender Was",
              "Arrested, Charged, or Convicted.")
```

```
FN2 <- paste("Only offenders with valid testing window start dates are",
             "included.",
             "Only incidents that occurred after the testing window ended and",
             "where the offender was arrested for, charged with, or convicted",
             "of at least one offense from the relevant crime category were",
             "included.",
             "The bottom four italicized broader categories include offenders",
             "who were arrested, charged or convicted for any of the more",
             "specific crime categories nested below them, while the Any",
             "Crimes category includes incidents associated with at least one",
             "of the 12 specific crime categories.",
             "Incidents associated with more than one kind of crime are",
             "included in all applicable specific categories, but only once per",
             "applicable broader category. For example, an incident with both a",
             "homicide and a robbery only contributes one observation to the",
             "years summary for the broader violent crimes category.")

VLLevels <- c("Any Crimes (12 categories)",
              "Sexual Crimes (2 categories)",
              "Criminal Sexual Conduct (CSC)",
              "Sex Crimes (non-CSC crimes)",
              "Violent Non-Sexual Crimes (5 categories)",
              "Assault, Domestic Violence",
              "Assault, non-Domestic Violence",
              "Homicide",
              "Robbery",
              "Weapons",
              "Property Crimes (3 categories)",
              "Arson",
              "Burglary",
              "Larceny, Theft, Fraud",
              "Other Crimes (2 categories)",
              "Drug Crimes",
              "Traffic & Ordinances")

CNames2 <- c("Crime Category", "N", "Mean", "SD", "Min", "Max", "Skew",
             "Kurtosis", "Q25", "Q50", "Q75")

# Summarize crime category counts (both overall and broken down by IWhen).
INCEWA %>%
  filter(!is.na(VLabel)) %>%
  select(VLabel, IYearsAfterTW) %>%
  describeBy(., group = "VLabel", quant=c(.25, .50, .75), mat = TRUE) %>%
  filter(vars == 2) %>%
  rename(Category = group1) %>%
  mutate(Category = factor(Category, levels = VLLevels, labels = VLLevels)) %>%
  arrange(Category) %>%
  select(Category, n, mean, sd, min, max, skew, kurtosis, Q0.25, Q0.5, Q0.75) %>%
  kable(., format = "latex", booktabs = TRUE, digits = 2, row.names = FALSE,
        col.names = CNames2, caption = TCap, linesep = "") %>%
  row_spec(., row = c(1, 2, 5, 11, 15), italic = TRUE) %>%
  pack_rows(" ", start_row = 1, end_row = 1) %>%
  pack_rows(" ", start_row = 2, end_row = 4) %>%
  pack_rows(" ", start_row = 5, end_row = 10) %>%
  pack_rows(" ", start_row = 11, end_row = 14) %>%
  pack_rows(" ", start_row = 15, end_row = 17) %>%
  add_indent(., positions = c(3:4), level_of_indent = 1) %>%
  add_indent(., positions = c(6:10), level_of_indent = 1) %>%
  add_indent(., positions = c(12:14), level_of_indent = 1) %>%
  add_indent(., positions = c(16:17), level_of_indent = 1) %>%
  column_spec(column = 1, width = "6.8 cm") %>%
  footnote(general = FN2, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

Figure 11 shows the density plots (i.e., smoothed histograms) for the distribution of *IYearsAfterTW* for each crime category. The shapes of these plots help us better understand Table 103.

```
FCap <- paste("\\label{fig:plot_IYearsAfterTW_Crime}",
              "Distributions for Years to Incident Date",
              "for Incidents After the Testing Window Where Offender Was",
```

Crime Category	N	Mean	SD	Min	Max	Skew	Kurtosis	Q25	Q50	Q75
Any Crimes (12 categories)	*5577*	*8.70*	*5.96*	*0.01*	*28.37*	*0.45*	*-0.65*	*3.60*	*8.22*	*13.03*
Sexual Crimes (2 categories)	*395*	*7.47*	*6.44*	*0.01*	*27.44*	*0.61*	*-0.72*	*1.49*	*5.71*	*12.25*
Criminal Sexual Conduct (CSC)	320	6.30	5.98	0.01	23.81	0.88	-0.29	1.16	4.69	9.78
Sex Crimes (non-CSC crimes)	92	11.41	6.57	0.01	27.44	-0.12	-0.79	6.26	11.74	16.51
Violent Non-Sexual Crimes (5 categories)	*1867*	*8.44*	*5.97*	*0.01*	*25.65*	*0.49*	*-0.68*	*3.47*	*7.83*	*12.82*
Assault, Domestic Violence	388	10.81	5.91	0.01	24.33	0.04	-0.86	6.31	10.70	15.44
Assault, non-Domestic Violence	1047	9.27	6.04	0.03	25.65	0.34	-0.79	4.05	8.86	13.61
Homicide	86	5.35	4.59	0.12	19.57	1.15	1.02	1.62	4.23	7.44
Robbery	331	6.28	5.58	0.01	22.36	0.92	0.02	1.58	4.68	9.57
Weapons	565	7.32	5.67	0.01	24.07	0.73	-0.27	2.55	6.14	10.63
Property Crimes (3 categories)	*1419*	*8.16*	*5.82*	*0.01*	*28.37*	*0.54*	*-0.48*	*3.11*	*7.46*	*12.55*
Arson	20	6.88	5.96	0.20	22.36	0.91	0.02	1.99	5.33	9.77
Burglary	379	7.68	5.74	0.02	28.37	0.66	-0.26	2.88	6.38	12.04
Larceny, Theft, Fraud	1077	8.36	5.84	0.01	25.57	0.49	-0.55	3.38	7.81	12.71
Other Crimes (2 categories)	*2836*	*9.49*	*5.84*	*0.01*	*28.24*	*0.33*	*-0.67*	*4.60*	*9.26*	*13.65*
Drug Crimes	950	8.03	5.57	0.01	28.24	0.60	-0.26	3.30	7.55	11.81
Traffic & Ordinances	2024	10.23	5.82	0.05	26.36	0.21	-0.73	5.47	10.16	14.51

Note: Only offenders with valid testing window start dates are included. Only incidents that occurred after the testing window ended and where the offender was arrested for, charged with, or convicted of at least one offense from the relevant crime category were included. The bottom four italicized broader categories include offenders who were arrested, charged or convicted for any of the more specific crime categories nested below them, while the Any Crimes category includes incidents associated with at least one of the 12 specific crime categories. Incidents associated with more than one kind of crime are included in all applicable specific categories, but only once per applicable broader category. For example, an incident with both a homicide and a robbery only contributes one observation to the years summary for the broader violent crimes category.

Table 103: Years to Incident Date By Crime Category for Incidents After the Testing Window Where Offender Was Arrested, Charged, or Convicted.

```
            "Arrested, Charged, or Convicted, By Crime Category.",
            "Each panel shows a densityplot above a strip plot with an",
            "overlaid boxplot.",
            "The strip plots show a dot for each incident, with positions",
            "jittered to reduce overlap.")

# Graph after period frequency distributions.
INCEWA %>%
  # Keep only the main 12 crime categories (drops the broader crime categories).
  filter(!is.na(Crime)) %>%
  select(OID, IID, Crime, IYearsAfterTW) %>%
  # use y = -... to position boxplot and jitterplot below the histogram
  ggplot(data = ., aes(x = IYearsAfterTW, y = -.4)) +
  # Add density plot.
  stat_halfeye(slab_fill = "black", width = .6, justification = -.42,
               .width = 0, point_colour = "NA") +
  # Add a jittered strip plot.
  geom_point(aes(), position = position_jitter(width = 0, height = .25),
             size = 1, color = "gray50") +
  # Overlay a boxplot on the strip plot from ggstance
  geom_boxplot(lwd = 1, width = .20, outlier.shape = NA) +
  # Merged those calls to one display
  guides(scale = "none") +
  # Split each crime category into a separate panel (facet)
  facet_wrap(~Crime, nrow = 6, strip.position = "top") +
```

```
  xlab("Years to Incident Date") +
  ylab("Density") +
  # Remove y axis title label from the y axis
  theme(axis.title.y = element_blank(), axis.text.y = element_blank(),
        axis.ticks.y = element_blank()) +
  # Format the y axis ticks.
  scale_x_continuous(limits = c(-.5, 30.00),
                     breaks = c(0, 5, 10, 15, 20, 25, 30),
                     labels = c(0, 5, 10, 15, 20, 25, 30)) ->
  Figure_Crime_Years

Figure_Crime_Years
```

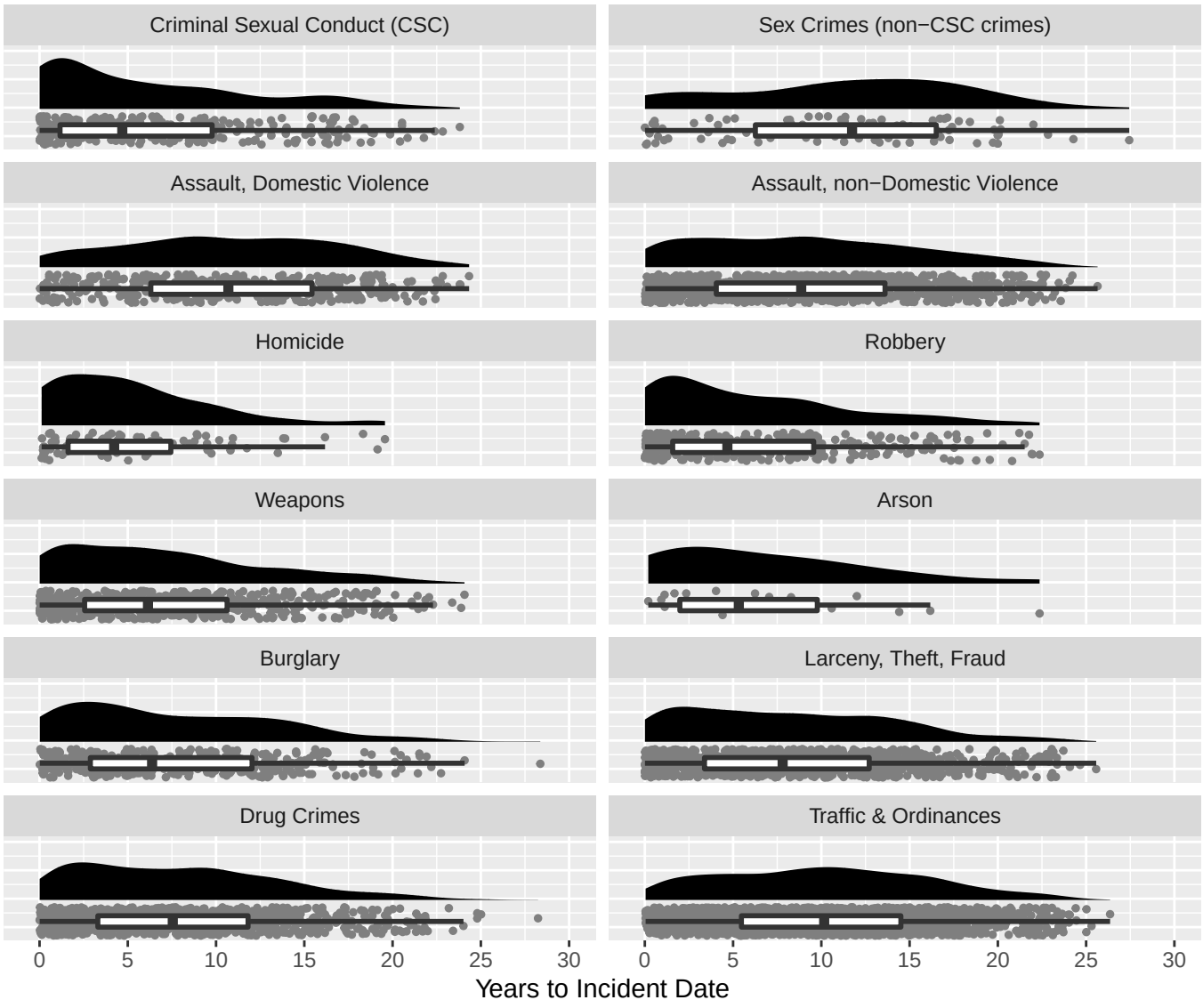

Figure 11: Distributions for Years to Incident Date for Incidents After the Testing Window Where Offender Was Arrested, Charged, or Convicted, By Crime Category. Each panel shows a densityplot above a strip plot with an overlaid boxplot. The strip plots show a dot for each incident, with positions jittered to reduce overlap.

Figure 12 shows the cumulative distribution functions for the number of years elapsed between the end of the testing window and each incident that occurred after the window ended, broken down by crime category. This is useful because the number of preventable incidents depends on when after the end of the testing window the offender reference sample gets uploaded to CODIS. This figure is an alternative way to visualize the same data shown in Figure 11 as a cumulative density up to a given number of years rather than the point density at each time point. The medians (Q50) shown in Table 103 are number

of years on the horizontal axis of 12 where the line rises to the 50% mark on the vertical axis. Similarly, the 25th and 75th quantiles (Q25 and Q75) are where the line rises to the 25% and 75% marks on the vertical axis.

```
FCap <- paste("\\label{fig:plot_IYearsAfterTW_Crime_ecdf}",
              "Cumulative Distribution Function for Years to Incident Date",
              "for Incidents After the Testing Window Where Offender Was",
              "Arrested, Charged, or Convicted, By Crime Category.",
              "Each panel shows a cumulative density function plot above a ",
              "strip plot with an overlaid boxplot.",
              "The strip plots show a dot for each incident, with positions",
              "jittered to reduce overlap.")

# Graph after period frequency distributions.
INCEWA %>%
  # Keep only the main 12 crime categories (drops the broader crime categories).
  filter(!is.na(Crime)) %>%
  select(OID, IID, Crime, IYearsAfterTW) %>%
  # use y = -... to position boxplot and jitterplot below the histogram
  ggplot(data = ., aes(x = IYearsAfterTW, y = -.4)) +
  # Add empirical cumulative distribution function plot.
  stat_ecdf(geom = "step", pad = FALSE) +
  # Add a jittered strip plot.
  geom_point(aes(), position = position_jitter(width = 0, height = .25),
             size = 1, color = "gray50") +
  # Overlay a boxplot on the strip plot from ggstance
  geom_boxplot(lwd = 1, width = .20, outlier.shape = NA) +
  # Merged those calls to one display
  guides(scale = "none") +
  # Split each crime category into a separate panel (facet)
  facet_wrap(~Crime, nrow = 6, strip.position = "top") +
  xlab("Years to Incident Date") +
  ylab("Cumulative Density") +
    # Format the y axis ticks.
  scale_y_continuous(limits = c(-.65, 1.00),
                     breaks = c(0, .5, 1.0), labels = c("0%", "50%", "100%")) +
  scale_x_continuous(limits = c(-.5, 30.00),
                     breaks = c(0, 5, 10, 15, 20, 25, 30),
                     labels = c(0, 5, 10, 15, 20, 25, 30))
```

Figure 13 shows the distributions for the number of years elapsed between the end of the testing window and each incident that occurred after the window ended broken down by broader crime category.

```
FCap <- paste("\\label{fig:plot_IYearsAfterTW_BCrime}",
              "Distributions for Years to Incident Date",
              "for Incidents After the Testing Window Where Offender Was",
              "Arrested, Charged, or Convicted, By Broader Crime Category.",
              "Each panel shows a densityplot above a strip plot with an",
              "overlaid boxplot.",
              "The strip plots show a dot for each incident, with positions",
              "jittered to reduce overlap.")

# Graph after period frequency distributions.
INCEWA %>%
  # Keep only the 4 broader crime categories (drops 12 main crime categories).
  filter(!is.na(BCrime)) %>%
  select(OID, IID, BCrime, IYearsAfterTW) %>%
  # use y = -... to position boxplot and jitterplot below the histogram
  ggplot(data = ., aes(x = IYearsAfterTW, y = -.4)) +
  # Add density plot.
  stat_halfeye(slab_fill = "black", width = .6, justification = -.42,
               .width = 0, point_colour = "NA") +
  # Add a jittered strip plot.
  geom_point(aes(), position = position_jitter(width = 0, height = .25),
             size = 1, color = "gray50") +
  # Overlay a boxplot on the strip plot from ggstance
  geom_boxplot(lwd = 1, width = .20, outlier.shape = NA) +
  # Merged those calls to one display
  guides(scale = "none") +
  # Split each crime category into a separate panel (facet)
  facet_wrap(~BCrime, nrow = 4, strip.position = "top") +
```

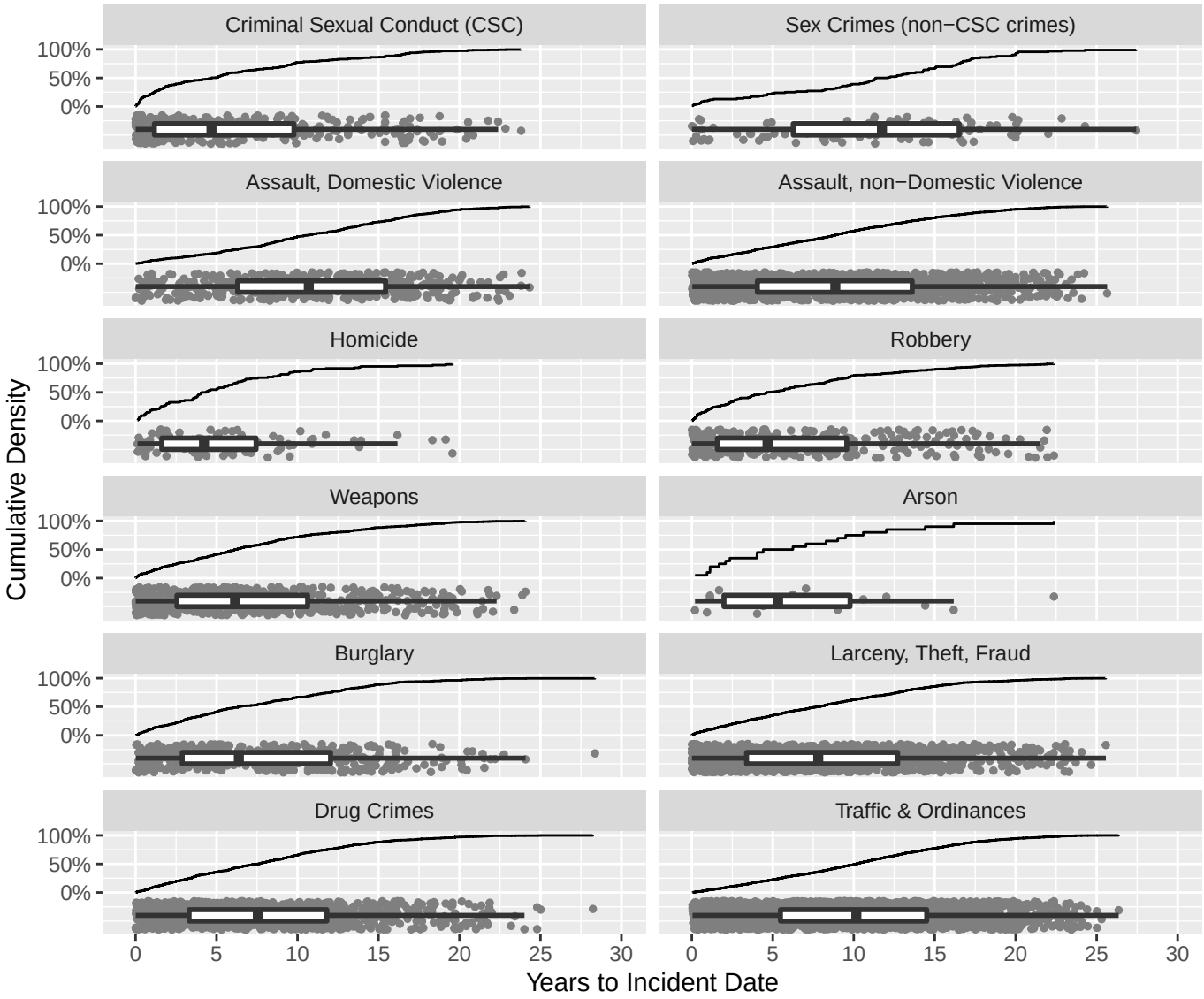

Figure 12: Cumulative Distribution Function for Years to Incident Date for Incidents After the Testing Window Where Offender Was Arrested, Charged, or Convicted, By Crime Category. Each panel shows a cumulative density function plot above a strip plot with an overlaid boxplot. The strip plots show a dot for each incident, with positions jittered to reduce overlap.

```
  xlab("Years to Incident Date") +
  ylab("Density") +
  # Remove y axis title label from the y axis
  theme(axis.title.y = element_blank(), axis.text.y = element_blank(),
        axis.ticks.y = element_blank()) +
  # Format the y axis ticks.
  scale_x_continuous(limits = c(-.5, 30.00),
                     breaks = c(0, 5, 10, 15, 20, 25, 30),
                     labels = c(0, 5, 10, 15, 20, 25, 30)) ->
  Figure_BCrime_Years

Figure_BCrime_Years
```

Figure 12 was the original version of a
figure we submitted in a revision to our manuscript. Reviewers found it confusing, so we created Figure 14 as a replacement when resubmitting the manuscript. The cumulative density at any given number of years after the end of the testing window is the proportion of incidents that had already occurred by that time. Switching to a focus on the proportion remaining

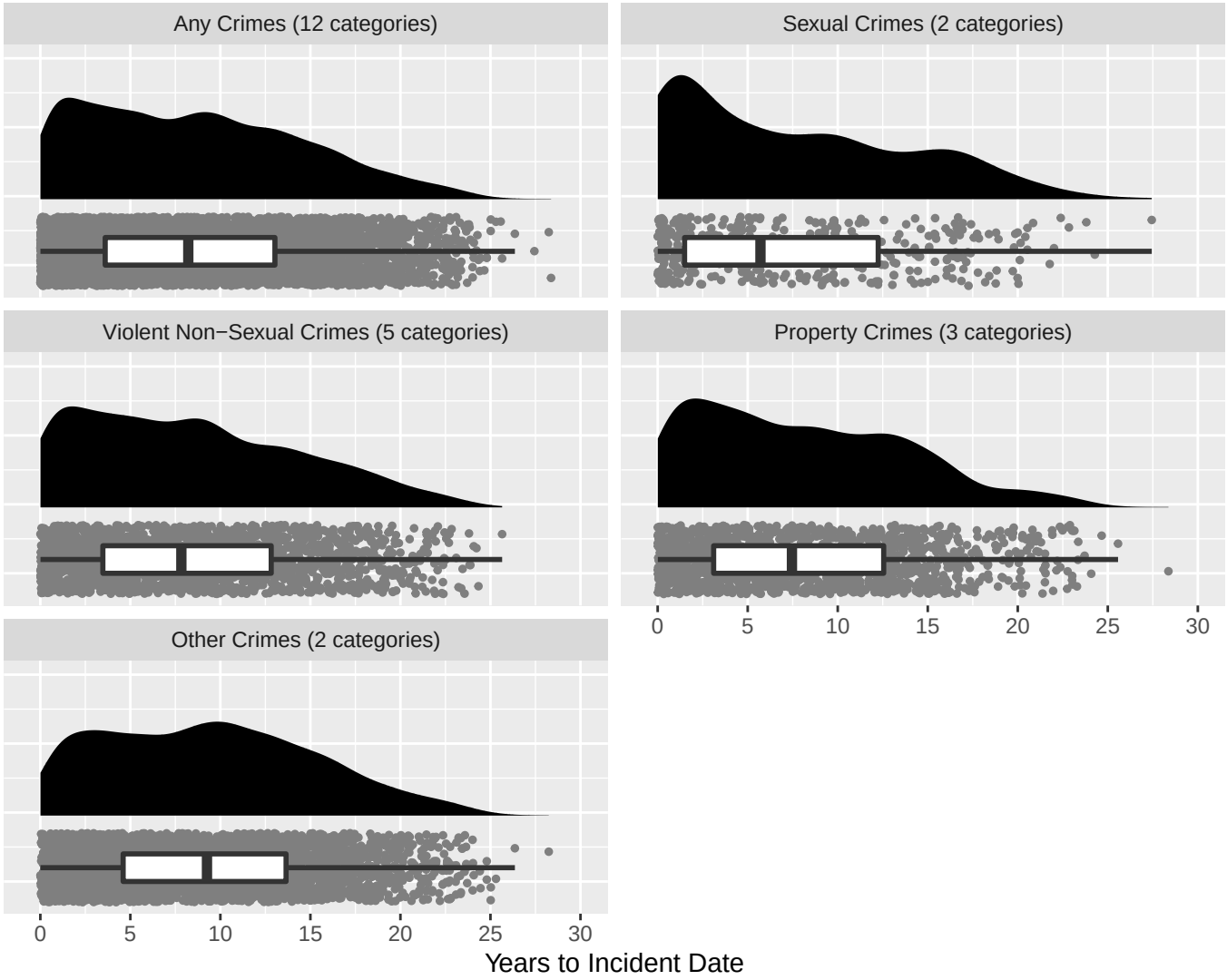

Figure 13: Distributions for Years to Incident Date for Incidents After the Testing Window Where Offender Was Arrested, Charged, or Convicted, By Broader Crime Category. Each panel shows a densityplot above a strip plot with an overlaid boxplot. The strip plots show a dot for each incident, with positions jittered to reduce overlap.

instead more closely aligns with the number of potentially preventable incidents. We also improved the labeling of the axes and the figure caption to address reviewer comments and make the figure easier to understand.

Subsequent peer review feedback caused us to eventually remove the figure even though we believe is is useful in communicating the results.

```
FCap <- paste("\\label{fig:plot_IYearsAfterTW_Crime_ecdf2}",
              "Percentage of Incidents Remaining As a Function of Years From",
              "End of Testing Window to Incident Date, By Crime Category.",
              "Only incidents occurring after the testing window where",
              "the offender was arrested for, charged with, or convicted of at",
              "least one offense from the relevant crime category were",
              "included.",
              "The lines show how the percentage of incidents remaining",
              "declines over time (they are the complements of the cumulative",
              "density functions).",
              "The strip plots below the lines show a dot for when each",
              "incident occurred: these are the data being summarized.",
              "Dot positions were jittered to reduce overlap.",
              "The overlaid boxplots provide an additional summary that ",
              "highlight when 75%, 50%, and 25% of the incidents remain.",
```

```
              "Bold x-axis values (at 1, 2, 5, and 10) are time points we",
              "examined to quantify implications of hypothetical delayed CODIS",
              "matches for how many crimes might still be preventable.")

x.ticks <- c(0, 1, 2, 5, 10, 15, 20, 25, 30)
x.face  <- c("0", expression(bold("1")), expression(bold("2")),
             expression(bold("5")), expression(bold("10")), "15", "20", "25",
             "30")

# Graph after period frequency distributions.
INCEWA %>%
  # Keep only the main 12 crime categories (drops the broader crime categories).
  filter(!is.na(Crime)) %>%
  select(OID, IID, Crime, IYearsAfterTW) %>%
  # use y = -... to position boxplot and jitterplot below the histogram
  ggplot(data = ., aes(x = IYearsAfterTW, y = -.4)) +
  # Add the complement of the empirical cumulative distribution function plot.
  stat_ecdf(mapping = aes(y = 1 - ..y..), geom = "step", pad = FALSE) +
  # Add a jittered strip plot.
  geom_point(aes(), position = position_jitter(width = 0, height = .25),
             size = 1, color = "gray50") +
  # Overlay a boxplot on the strip plot from ggstance
  geom_boxplot(lwd = 1, width = .20, outlier.shape = NA) +
  # Merged those calls to one display
  guides(scale = "none") +
  # Split each crime category into a separate panel (facet)
  facet_wrap(~Crime, nrow = 6, strip.position = "top") +
  xlab("Years From End of Testing Window to Incident Date") +
  ylab("Percentage of Incidents Remaining") +
    # Format the y axis ticks.
  scale_y_continuous(limits = c(-.65, 1.00),
                     breaks = c(0, .5, 1.0), labels = c("0%", "50%", "100%")) +
  scale_x_continuous(limits = c(-.5, 30.00),
                     breaks = x.ticks, labels = x.face) ->
  #theme(axis.text.x = element_text(face=x.face)) ->
  Figure_Crime_Years

Figure_Crime_Years
```

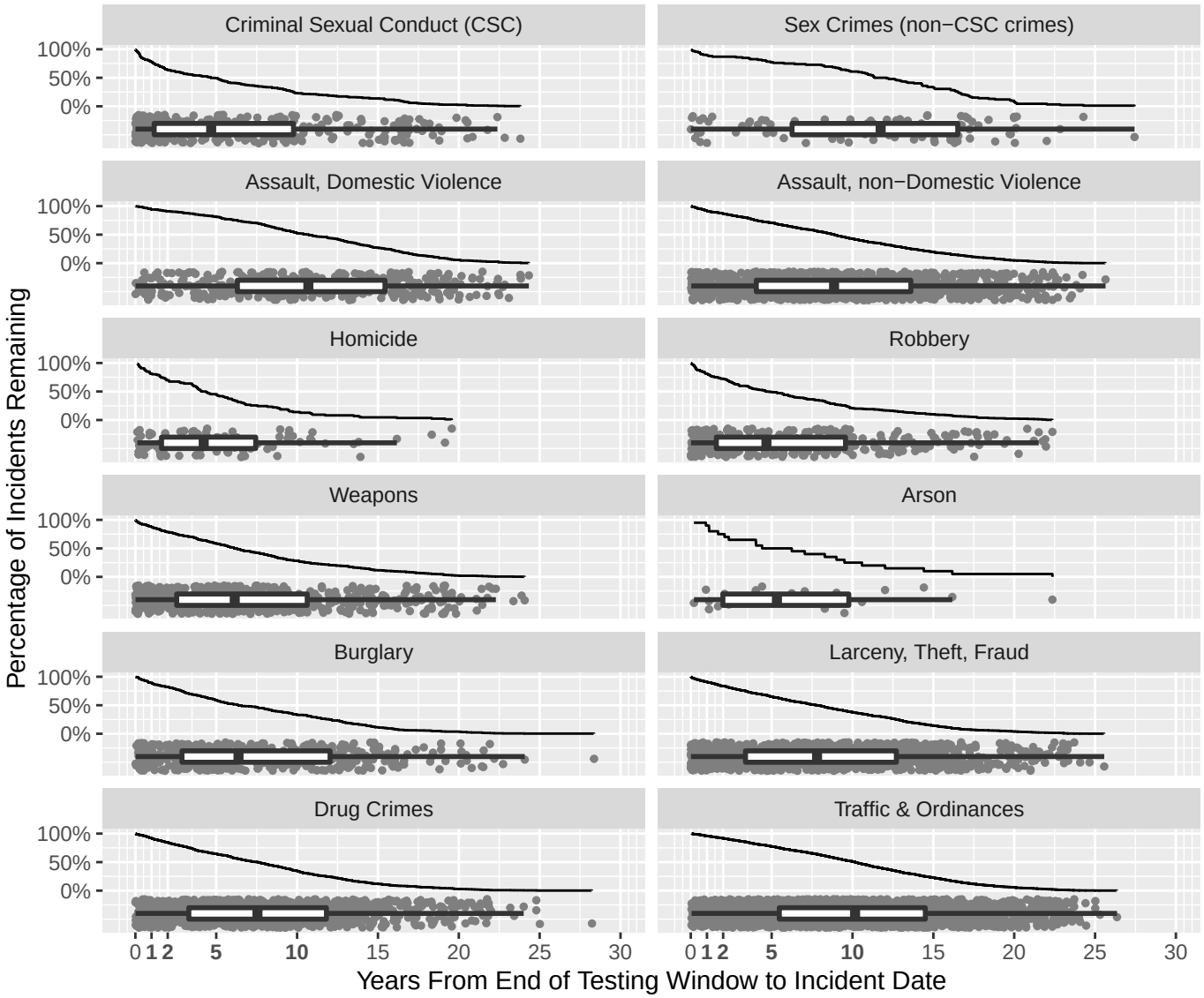

Figure 14: Percentage of Incidents Remaining As a Function of Years From End of Testing Window to Incident Date, By Crime Category. Only incidents occurring after the testing window where the offender was arrested for, charged with, or convicted of at least one offense from the relevant crime category were included. The lines show how the percentage of incidents remaining declines over time (they are the complements of the cumulative density functions). The strip plots below the lines show a dot for when each incident occurred: these are the data being summarized. Dot positions were jittered to reduce overlap. The overlaid boxplots provide an additional summary that highlight when 75%, 50%, and 25% of the incidents remain. Bold x-axis values (at 1, 2, 5, and 10) are time points we examined to quantify implications of hypothetical delayed CODIS matches for how many crimes might still be preventable.

5.11.3 Number of Potentially Preventable Incidents by Crime Category

Now we extract the empirical cumulative distribution functions data underlying Figure 12 and use them to prepare Table 104. A reviewer noted that uploading an offender DNA profile from a SAK to CODIS can only produce immediate identification of the offender if the corresponding offender reference sample was already present in CODIS beforehand. However, that may not always happen. The offender reference sample that allows identifying the offender associated with SAK may not arrive in CODIS until later. That means there could be a delay between uploading the SAK forensic testing results to CODIS as a forensic sample and getting a match to an offender. Incidents occurring between the end of the testing window and the arrival of a matching offender reference sample cannot be prevented through timely response to CODIS search results that are not yet available. Only crimes occurring after a match is available could be prevented that way.

It would be ideal if we had the CODIS upload date for the reference sample that matched to each SAK. That would allow us

to examine and describe the distribution of the delay between uploading the SAK sample and receiving an offender match. There would be no delay ($\leq$ 0 years) when the offender reference sample was already present in CODIS, but it could be anywhere from 1 day to many years before one is available. Knowing more about that distribution would allow us to find the subset of potentially preventable incidents that occurred both after the testing window ended and after the matching reference sample was uploaded to CODIS.

Unfortunately, we do not have access to those offender reference sample upload dates. Therefore, we examine the implications of assuming a fixed delay in when offender reference samples arrive in CODIS relative to the end of the testing window. We do that by using the cumulative distribution functions graphed in Figure 12 to compute how many incidents occurred 1, 2, 5, and 10 years after the end of the testing window. This makes the crude assumption that all SAKs experience the same delay.

```
# Table caption.
TCap <- paste("Number and Percentage of Incidents Remaining Given Hypothetical",
              "Delays Between End of Testing Window and Matching to an ",
              "Offender Reference Sample")
FN2 <- paste("Only offenders with valid testing window start dates are",
             "included.",
             "Only incidents that occurred after the testing window ended and",
             "where the offender was arrested for, charged with, or convicted",
             "of at least one offense from the relevant crime category were",
             "included.",
             "The bottom four italicized broader categories include offenders",
             "who were arrested, charged or convicted for any of the more",
             "specific crime categories nested below them, while the Any",
             "Crimes category includes incidents associated with at least one",
             "of the 12 specific crime categories. ",
             "Incidents associated with more than one kind of crime are",
             "included in all applicable specific categories, but only once per",
             "applicable broader category. For example, an incident with both a",
             "homicide and a robbery only contributes one observation to the",
             "years summary for the broader violent crimes category.",
             "N, total number of incidents after the end of testing window;",
             "n, number of incidents remaining after specified delay." )

INCEWA %>%
  filter(!is.na(VLabel)) %>%
  mutate(Category = factor(VLabel, levels = VLLevels, labels = VLLevels)) %>%
  select(Category, IYearsAfterTW) %>%
  arrange(Category, IYearsAfterTW) %>%
  group_by(Category) %>%
  summarize(N = n(),
            # If CODIS reference sample uploaded 1 year after SAK
            PRemain_1 = 1 - ecdf(IYearsAfterTW)(1),
            NRemain_1 = round(N*PRemain_1, digits = 0),
            # If CODIS reference sample uploaded 2 years after SAK
            PRemain_2 = 1 - ecdf(IYearsAfterTW)(2),
            NRemain_2 = round(N*PRemain_2, digits = 0),
            # If CODIS reference sample uploaded 5 years after SAK
            PRemain_5 = 1 - ecdf(IYearsAfterTW)(5),
            NRemain_5 = round(N*PRemain_5, digits = 0),
            # If CODIS reference sample uploaded 10 years after SAK
            PRemain_10 = 1 - ecdf(IYearsAfterTW)(10),
            NRemain_10 = round(N*PRemain_10, digits = 0)) %>%
  # Rescale proportions to percentages
  mutate(PRemain_1 = PRemain_1*100,
         PRemain_2 = PRemain_2*100,
         PRemain_5 = PRemain_5*100,
         PRemain_10 = PRemain_10*100) %>%
  select(Category, N, NRemain_1, PRemain_1, NRemain_2, PRemain_2,
         NRemain_5, PRemain_5, NRemain_10, PRemain_10) %>%
  kable(format = "latex", digits = 2, booktabs = TRUE, caption = TCap,
        format.args = list(big.mark = ','), linesep = "",
        col.names = c("Category", "N", rep(c("n", "%"), times = 4))) %>%
  kable_styling() %>%
  add_header_above(., c(" ", " ", "1 Year" = 2, "2 Years" = 2,
                        "5 Years" = 2, "10 Years" = 2)) %>%
  add_header_above(., c(" ", " ",
                        "Delay in Matching to Offender Reference Sample" = 8)) %>%
  row_spec(., row = c(1, 2, 5, 11, 15), italic = TRUE) %>%
```

```
pack_rows(" ", start_row = 1, end_row = 1) %>%
pack_rows(" ", start_row = 2, end_row = 4) %>%
pack_rows(" ", start_row = 5, end_row = 10) %>%
pack_rows(" ", start_row = 11, end_row = 14) %>%
pack_rows(" ", start_row = 15, end_row = 17) %>%
add_indent(., positions = c(3:4), level_of_indent = 1) %>%
add_indent(., positions = c(6:10), level_of_indent = 1) %>%
add_indent(., positions = c(12:14), level_of_indent = 1) %>%
add_indent(., positions = c(16:17), level_of_indent = 1) %>%
column_spec(column = 1, width = "6.8 cm") %>%
footnote(general = FN2, general_title = "Note: ", footnote_as_chunk = TRUE,
       threeparttable = TRUE)
```

		Delay in Matching to Offender Reference Sample							
		1 Year		2 Years		5 Years		10 Years	
Category	N	n	%	n	%	n	%	n	%
Any Crimes (12 categories)	*5,577*	*5,091*	*91.29*	*4,725*	*84.72*	*3,733*	*66.94*	*2,210*	*39.63*
Sexual Crimes (2 categories)	*395*	*320*	*81.01*	*274*	*69.37*	*221*	*55.95*	*125*	*31.65*
Criminal Sexual Conduct (CSC)	320	250	78.12	205	64.06	158	49.38	73	22.81
Sex Crimes (non-CSC crimes)	92	82	89.13	80	86.96	71	77.17	56	60.87
Violent Non-Sexual Crimes (5 categories)	*1,867*	*1,673*	*89.61*	*1,558*	*83.45*	*1,222*	*65.45*	*683*	*36.58*
Assault, Domestic Violence	388	366	94.33	353	90.98	315	81.19	204	52.58
Assault, non-Domestic Violence	1,047	961	91.79	910	86.91	739	70.58	449	42.88
Homicide	86	70	81.40	60	69.77	39	45.35	12	13.95
Robbery	331	268	80.97	239	72.21	163	49.24	69	20.85
Weapons	565	496	87.79	444	78.58	331	58.58	160	28.32
Property Crimes (3 categories)	*1,419*	*1,283*	*90.42*	*1,179*	*83.09*	*895*	*63.07*	*514*	*36.22*
Arson	20	18	90.00	15	75.00	10	50.00	5	25.00
Burglary	379	339	89.45	311	82.06	223	58.84	126	33.25
Larceny, Theft, Fraud	1,077	977	90.71	902	83.75	701	65.09	407	37.79
Other Crimes (2 categories)	*2,836*	*2,678*	*94.43*	*2,529*	*89.17*	*2,072*	*73.06*	*1,293*	*45.59*
Drug Crimes	950	873	91.89	801	84.32	614	64.63	328	34.53
Traffic & Ordinances	2,024	1,939	95.80	1,859	91.85	1,570	77.57	1,035	51.14

Note: Only offenders with valid testing window start dates are included. Only incidents that occurred after the testing window ended and where the offender was arrested for, charged with, or convicted of at least one offense from the relevant crime category were included. The bottom four italicized broader categories include offenders who were arrested, charged or convicted for any of the more specific crime categories nested below them, while the Any Crimes category includes incidents associated with at least one of the 12 specific crime categories. Incidents associated with more than one kind of crime are included in all applicable specific categories, but only once per applicable broader category. For example, an incident with both a homicide and a robbery only contributes one observation to the years summary for the broader violent crimes category. N, total number of incidents after the end of testing window; n, number of incidents remaining after specified delay.

Table 104: Number and Percentage of Incidents Remaining Given Hypothetical Delays Between End of Testing Window and Matching to an Offender Reference Sample

5.11.4 Export Figures for Manuscript

The American Psychological Association's general instructions regarding manuscript preparation refer authors to KnowledgeWorks Global Digital Art Support General Guidelines for guidance on preparing figures. We are complying with those

guidelines for preparing the following Figure files that we export.

We can choose between submitting TIFF versus EPS file. Because EPS files are orders of magnitude smaller, we choose those for submission. We also export PNG files for convenience (they are easier to view without specialized software).

We wrote the code below to export a version of Figure 14, but have subsequently disabled the code chunk below because the final manuscript will no longer contain the figure.

```
ggsave(filename = here::here("inst/Figure_Crime_Years.eps"),
       plot = Figure_Crime_Years, device = "eps", dpi = 600,
       height = 6.0, width = 7.2, units = "in")
ggsave(filename = here::here("inst/Figure_Crime_Years.png"),
       plot = Figure_Crime_Years, device = "png", dpi = 600,
       height = 6.0, width = 7.2, units = "in")
```

6 Arrest Offense Record Summaries

All summaries in this section directly summarize the ARR data without first aggregating to the incident or offender level. These outputs are not directly used in the manuscript.

6.1 Crime Category (12 levels)

```
FN <- paste("Only arrest records for offenders with valid testing window start",
            "dates and with offenses from the 12 main crime categories were",
            "included.")

kable(freq(as_factor(ARREW$ACat12), user.missing = "Excluded user-missing"),
      format = "latex", booktabs = TRUE, digits = 2,
      format.args = list(big.mark = ','),
      caption = "Number of Arrest Offense Records by Crime Category (12 levels)") %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
           threeparttable = TRUE)
```

	Frequency	Percent	Valid Percent
Arson	31	0.35	0.35
Assault - DV, stalking	144	1.61	1.61
Assault - non-sexual, non-DV	1,130	12.63	12.63
Burglary	579	6.47	6.47
Criminal sexual conduct	547	6.12	6.12
Drug crime	1,390	15.54	15.54
Homicide	97	1.08	1.08
Larceny/Theft/Fraud	1,735	19.40	19.40
Robbery	479	5.35	5.35
Sex crime, other (excluding CSC)	62	0.69	0.69
Traffic & Ordinances	2,134	23.86	23.86
Weapons	617	6.90	6.90
Excluded user-missing	0	0.00	NA
Total	8,945	100.00	100.00

Note: Only arrest records for offenders with valid testing window start dates and with offenses from the 12 main crime categories were included.

Table 105: Number of Arrest Offense Records by Crime Category (12 levels)

6.2 Age at Arrest for Criminal Sexual Conduct

```
# Table caption.
TCap <- paste("Offender Age at Arrest for Criminal Sexual Conduct (Overall)")
# Footnote text.
FN <- paste("This is an arrest offense record-level summary (offenders with",
            "multiple arrests contribute one age value per arrest record.",
            "Only arrest records for offenders with valid testing window start",
            "dates and where the arrest was for criminal sexual conduct are",
            "included.")

ARREW %>%
  filter(ACat12_5 == 1) %>%
  select(OAgeA, ALag) %>%
  describe(., quant=c(.25, .50, .75)) %>%
  filter(vars == 1) %>%
  bind_cols(data.frame(Variable = "Age at arrest"), .) %>%
  select(Variable, n, mean, sd, min, max, skew, kurtosis, Q0.25, Q0.5, Q0.75) %>%
  kable(., format = "latex", booktabs = TRUE, digits = 2, row.names = FALSE,
        col.names = CNames, caption = TCap) %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

Variable	N	Mean	SD	Min	Max	Skew	Kurtosis	Q25	Q50	Q75
Age at arrest	547	30.12	9.42	16	63	0.6	-0.1	22	29	37

Note: This is an arrest offense record-level summary (offenders with multiple arrests contribute one age value per arrest record. Only arrest records for offenders with valid testing window start dates and where the arrest was for criminal sexual conduct are included.

Table 106: Offender Age at Arrest for Criminal Sexual Conduct (Overall)

7 Prosecutor Charge Record Summaries

This section directly summarizes the CHG data without first aggregating to the incident or offender level. These outputs are not directly used in the manuscript.

```
FN <- paste("Only charge records for offenders with valid testing window start",
            "dates and with charges from the 12 main crime categories were",
            "included.")

kable(freq(as_factor(CHGEW$CCat12), user.missing = "Excluded user-missing"),
      format = "latex", booktabs = TRUE, digits = 2,
      format.args = list(big.mark = ','),
      caption = "Number of Prosecutor Charge Records by Crime Category (12 levels)") %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

	Frequency	Percent	Valid Percent
Arson	28	0.50	0.50
Assault - DV, stalking	288	5.11	5.11
Assault - non-sexual, non-DV	563	10.00	10.00
Burglary	387	6.87	6.87
Criminal sexual conduct	533	9.46	9.46
Drug crime	1,029	18.27	18.27
Homicide	83	1.47	1.47
Larceny/Theft/Fraud	888	15.77	15.77
Robbery	342	6.07	6.07
Sex crime, other (excluding CSC)	55	0.98	0.98
Traffic & Ordinances	743	13.19	13.19
Weapons	693	12.30	12.30
Excluded user-missing	0	0.00	NA
Total	5,632	100.00	100.00

Note: Only charge records for offenders with valid testing window start dates and with charges from the 12 main crime categories were included.

Table 107: Number of Prosecutor Charge Records by Crime Category (12 levels)

8 Judicial Charge Record Summaries

All summaries in this section directly summarize the JUD data without first aggregating to the incident or offender level. While some of these records are convictions, others will have other dispositions. These outputs are not directly used in the manuscript.

8.1 Crime Category (12 levels, all dispositions)

```
FN <- paste("Only records for offenders with valid testing window start dates",
            "and with charges from one of the 12 main crime categories were",
            "included.")

kable(freq(as_factor(JUDEW$JCat12), user.missing = "Excluded user-missing"),
      format = "latex", booktabs = TRUE, digits = 2,
      format.args = list(big.mark = ','),
      caption = "Number of Adjudicated Charge Records (All Dispositions) by Crime Category (12 levels)")
```

	Frequency	Percent	Valid Percent
Arson	36	0.33	0.33
Assault - DV, stalking	474	4.31	4.31
Assault - non-sexual, non-DV	1,161	10.56	10.56
Burglary	670	6.09	6.09
Criminal sexual conduct	1,483	13.49	13.49
Drug crime	1,742	15.84	15.84
Homicide	145	1.32	1.32
Larceny/Theft/Fraud	1,537	13.98	13.98
Robbery	725	6.59	6.59
Sex crime, other (excluding CSC)	87	0.79	0.79
Traffic & Ordinances	1,510	13.73	13.73
Weapons	1,425	12.96	12.96
Excluded user-missing	0	0.00	NA
Total	10,995	100.00	100.00

Table 108: Number of Adjudicated Charge Records (All Dispositions) by Crime Category (12 levels)

8.2 Crime Category (12 levels, convictions)

```
FN <- paste("Only conviction records for offenders with valid testing window",
            "start dates and with charges from one of the 12 main crime",
            "categories were included.")

kable(freq(as_factor(CONEW$JCat12), user.missing = "Excluded user-missing"),
      format = "latex", booktabs = TRUE, digits = 2,
      format.args = list(big.mark = ','),
      caption = "Number of Adjudicated Charge Records (Convictions Only) by Crime Category (12 levels)") %>%
  footnote(general = FN, general_title = "Note: ", footnote_as_chunk = TRUE,
         threeparttable = TRUE)
```

	Frequency	Percent	Valid Percent
Arson	19	0.32	0.32
Assault - DV, stalking	209	3.47	3.47
Assault - non-sexual, non-DV	512	8.50	8.50
Burglary	418	6.94	6.94
Criminal sexual conduct	632	10.50	10.50
Drug crime	955	15.86	15.86
Homicide	80	1.33	1.33
Larceny/Theft/Fraud	962	15.98	15.98
Robbery	442	7.34	7.34
Sex crime, other (excluding CSC)	48	0.80	0.80
Traffic & Ordinances	1,017	16.89	16.89
Weapons	727	12.07	12.07
Excluded user-missing	0	0.00	NA
Total	6,021	100.00	100.00

Note: Only conviction records for offenders with valid testing window start dates and with charges from one of the 12 main crime categories were included.

Table 109: Number of Adjudicated Charge Records (Convictions Only) by Crime Category (12 levels)

9 Wrap Up

9.1 Project Information

These materials are scholarly products based on research funded by the following grant.

Campbell, R., Pierce, S. J., & Sharma, D. (2015–2018). *Serial sexual assaults: A longitudinal examination of offending patterns using DNA evidence.* (NIJ Award # 2014-NE-BX-0006) [Grant]. National Institute of Justice.

9.2 References

Campbell, R. (2019). *Serial sexual assaults: A longitudinal examination of offending patterns using DNA evidence, Detroit, Michigan, 2009* [Data files, codebooks, computer programs, and statistical output]. ICPSR37134-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2019-02-28. Retrieved from: https://doi.org/10.3886/ICPSR37134.v1

9.3 Software Information

We use R Markdown to enhance reproducibility. Knitting the source R Markdown script *Step_02_Analysis.Rmd* generates this PDF file containing explanatory text, R code, plus R output (text and graphics).

- We used RStudio to work with R and R markdown files. The software chain looks like this: **Rmd file > RStudio > R > rmarkdown > knitr > md file > pandoc > tex file > TinyTeX > PDF file**.
- We recommend using TinyTeX to compile LaTeX files into PDF files. However, it should be viable to use MiKTeX or another LaTeX distribution instead.
- We used pandoc 2.14.0.3 for this document.

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

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

```
## [1] "is_tinytex = TRUE. We used TinyTeX."
```

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

```
## - Session info ---------------------------------------------------------------
##  setting  value
##  version  R version 4.1.2 (2021-11-01)
##  os       Windows 10 x64 (build 19042)
##  system   x86_64, mingw32
##  ui       RTerm
##  language (EN)
##  collate  English_United States.1252
##  ctype    English_United States.1252
##  tz       America/New_York
##  date     2022-01-15
##  pandoc   2.14.0.3 @ C:/Program Files/RStudio/bin/pandoc/ (via rmarkdown)
##
## - Packages -------------------------------------------------------------------
##  package        * version date (UTC) lib source
##  abind            1.4-5   2016-07-21 [1] CRAN (R 4.1.0)
##  assertthat       0.2.1   2019-03-21 [1] CRAN (R 4.1.0)
##  backports        1.4.1   2021-12-13 [1] CRAN (R 4.1.2)
##  broom            0.7.11  2022-01-03 [1] CRAN (R 4.1.2)
##  cachem           1.0.6   2021-08-19 [1] CRAN (R 4.1.1)
##  callr            3.7.0   2021-04-20 [1] CRAN (R 4.1.0)
##  car            * 3.0-12  2021-11-06 [1] CRAN (R 4.1.2)
##  carData        * 3.0-5   2022-01-06 [1] CRAN (R 4.1.2)
```

```
##  cli             3.1.0   2021-10-27 [1] CRAN (R 4.1.1)
##  coda            0.19-4  2020-09-30 [1] CRAN (R 4.1.0)
##  codetools       0.2-18  2020-11-04 [2] CRAN (R 4.1.2)
##  colorspace      2.0-2   2021-06-24 [1] CRAN (R 4.1.0)
##  crayon          1.4.2   2021-10-29 [1] CRAN (R 4.1.1)
##  DBI             1.1.2   2021-12-20 [1] CRAN (R 4.1.2)
##  desc            1.4.0   2021-09-28 [1] CRAN (R 4.1.1)
##  descr         * 1.1.5   2021-02-16 [1] CRAN (R 4.1.0)
##  devtools        2.4.3   2021-11-30 [1] CRAN (R 4.1.2)
##  digest          0.6.29  2021-12-01 [1] CRAN (R 4.1.2)
##  distributional  0.3.0   2022-01-05 [1] CRAN (R 4.1.2)
##  dplyr         * 1.0.7   2021-06-18 [1] CRAN (R 4.1.0)
##  ellipsis        0.3.2   2021-04-29 [1] CRAN (R 4.1.0)
##  emmeans       * 1.7.2   2022-01-04 [1] CRAN (R 4.1.2)
##  estimability    1.3     2018-02-11 [1] CRAN (R 4.1.0)
##  evaluate        0.14    2019-05-28 [1] CRAN (R 4.1.0)
##  fansi           0.5.0   2021-05-25 [1] CRAN (R 4.1.2)
##  farver          2.1.0   2021-02-28 [1] CRAN (R 4.1.0)
##  fastmap         1.1.0   2021-01-25 [1] CRAN (R 4.1.0)
##  forcats         0.5.1   2021-01-27 [1] CRAN (R 4.1.0)
##  fs              1.5.2   2021-12-08 [1] CRAN (R 4.1.2)
##  geepack       * 1.3.3   2022-01-09 [1] CRAN (R 4.1.2)
##  generics        0.1.1   2021-10-25 [1] CRAN (R 4.1.1)
##  ggdist        * 3.0.1   2021-11-30 [1] CRAN (R 4.1.2)
##  ggplot2       * 3.3.5   2021-06-25 [1] CRAN (R 4.1.0)
##  git2r           0.29.0  2021-11-22 [1] CRAN (R 4.1.2)
##  glue            1.6.0   2021-12-17 [1] CRAN (R 4.1.2)
##  gtable          0.3.0   2019-03-25 [1] CRAN (R 4.1.0)
##  haven         * 2.4.3   2021-08-04 [1] CRAN (R 4.1.0)
##  here          * 1.0.1   2020-12-13 [1] CRAN (R 4.1.0)
##  highr           0.9     2021-04-16 [1] CRAN (R 4.1.0)
##  hms             1.1.1   2021-09-26 [1] CRAN (R 4.1.1)
##  htmltools       0.5.2   2021-08-25 [1] CRAN (R 4.1.1)
##  httr            1.4.2   2020-07-20 [1] CRAN (R 4.1.0)
##  insight         0.15.0  2022-01-07 [1] CRAN (R 4.1.2)
##  jpeg            0.1-9   2021-07-24 [1] CRAN (R 4.1.0)
##  kableExtra    * 1.3.4   2021-02-20 [1] CRAN (R 4.1.0)
##  knitr         * 1.37    2021-12-16 [1] CRAN (R 4.1.2)
##  labeling        0.4.2   2020-10-20 [1] CRAN (R 4.1.0)
##  lattice       * 0.20-45 2021-09-22 [1] CRAN (R 4.1.1)
##  latticeExtra  * 0.6-29  2019-12-19 [1] CRAN (R 4.1.0)
##  lifecycle       1.0.1   2021-09-24 [1] CRAN (R 4.1.1)
##  lubridate     * 1.8.0   2021-10-07 [1] CRAN (R 4.1.1)
##  magrittr        2.0.1   2020-11-17 [1] CRAN (R 4.1.0)
##  MASS            7.3-54  2021-05-03 [2] CRAN (R 4.1.2)
##  Matrix          1.4-0   2021-12-08 [1] CRAN (R 4.1.2)
##  memoise         2.0.1   2021-11-26 [1] CRAN (R 4.1.2)
##  mnormt          2.0.2   2020-09-01 [1] CRAN (R 4.1.0)
##  multcomp        1.4-18  2022-01-04 [1] CRAN (R 4.1.2)
##  munsell         0.5.0   2018-06-12 [1] CRAN (R 4.1.0)
##  mvtnorm         1.1-3   2021-10-08 [1] CRAN (R 4.1.1)
##  nlme            3.1-153 2021-09-07 [1] CRAN (R 4.1.1)
##  pillar          1.6.4   2021-10-18 [1] CRAN (R 4.1.1)
##  pkgbuild        1.3.1   2021-12-20 [1] CRAN (R 4.1.2)
##  pkgconfig       2.0.3   2019-09-22 [1] CRAN (R 4.1.0)
##  pkgload         1.2.4   2021-11-30 [1] CRAN (R 4.1.2)
##  plyr          * 1.8.6   2020-03-03 [1] CRAN (R 4.1.0)
##  png             0.1-7   2013-12-03 [1] CRAN (R 4.1.0)
##  prettyunits     1.1.1   2020-01-24 [1] CRAN (R 4.1.0)
##  processx        3.5.2   2021-04-30 [1] CRAN (R 4.1.0)
##  ps              1.6.0   2021-02-28 [1] CRAN (R 4.1.0)
##  psych         * 2.1.9   2021-09-22 [1] CRAN (R 4.1.1)
##  purrr           0.3.4   2020-04-17 [1] CRAN (R 4.1.0)
##  R6              2.5.1   2021-08-19 [1] CRAN (R 4.1.1)
##  RColorBrewer    1.1-2   2014-12-07 [1] CRAN (R 4.1.0)
##  Rcpp            1.0.7   2021-07-07 [1] CRAN (R 4.1.0)
##  remotes         2.4.2   2021-11-30 [1] CRAN (R 4.1.2)
##  rlang           0.4.12  2021-10-18 [1] CRAN (R 4.1.1)
##  rmarkdown     * 2.11    2021-09-14 [1] CRAN (R 4.1.1)
##  rprojroot       2.0.2   2020-11-15 [1] CRAN (R 4.1.0)
##  rstudioapi      0.13    2020-11-12 [1] CRAN (R 4.1.0)
##  rvest           1.0.2   2021-10-16 [1] CRAN (R 4.1.1)
```

```
##  sandwich         3.0-1   2021-05-18 [1] CRAN (R 4.1.0)
##  scales           1.1.1   2020-05-11 [1] CRAN (R 4.1.0)
##  sessioninfo      1.2.2   2021-12-06 [1] CRAN (R 4.1.2)
##  sjlabelled     * 1.1.8   2021-05-11 [1] CRAN (R 4.1.0)
##  SSACHR         * 1.0.0   2022-01-15 [1] Github (sjpierce/SSACHR@7ecfb11)
##  stringi          1.7.6   2021-11-29 [1] CRAN (R 4.1.2)
##  stringr          1.4.0   2019-02-10 [1] CRAN (R 4.1.0)
##  survival         3.2-13  2021-08-24 [1] CRAN (R 4.1.1)
##  svglite          2.0.0   2021-02-20 [1] CRAN (R 4.1.0)
##  systemfonts      1.0.3   2021-10-13 [1] CRAN (R 4.1.1)
##  testthat         3.1.1   2021-12-03 [1] CRAN (R 4.1.2)
##  texreg         * 1.37.5  2020-06-18 [1] CRAN (R 4.1.0)
##  TH.data          1.1-0   2021-09-27 [1] CRAN (R 4.1.1)
##  tibble           3.1.6   2021-11-07 [1] CRAN (R 4.1.2)
##  tidyr          * 1.1.4   2021-09-27 [1] CRAN (R 4.1.1)
##  tidyselect       1.1.1   2021-04-30 [1] CRAN (R 4.1.0)
##  tinytex          0.36    2021-12-19 [1] CRAN (R 4.1.2)
##  tmvnsim          1.0-2   2016-12-15 [1] CRAN (R 4.1.0)
##  usethis          2.1.5   2021-12-09 [1] CRAN (R 4.1.2)
##  utf8             1.2.2   2021-07-24 [1] CRAN (R 4.1.0)
##  vctrs            0.3.8   2021-04-29 [1] CRAN (R 4.1.0)
##  viridisLite      0.4.0   2021-04-13 [1] CRAN (R 4.1.0)
##  webshot          0.5.2   2019-11-22 [1] CRAN (R 4.1.0)
##  withr            2.4.3   2021-11-30 [1] CRAN (R 4.1.2)
##  xfun             0.29    2021-12-14 [1] CRAN (R 4.1.2)
##  xml2             1.3.3   2021-11-30 [1] CRAN (R 4.1.2)
##  xtable           1.8-4   2019-04-21 [1] CRAN (R 4.1.0)
##  yaml             2.2.1   2020-02-01 [1] CRAN (R 4.1.0)
##  zoo              1.8-9   2021-03-09 [1] CRAN (R 4.1.0)
##
##  [1] C:/Users/pierces1/OneDrive - Michigan State University/CSTATRedirects/Documents/R/win-library/4.1
##  [2] C:/Program Files/R/R-4.1.2/library
##
## ------------------------------------------------------------------------------
```

The current Git commit details and status are:

```
git_report()
```

```
## Local:    main S:/14-286/Analyses/SSACHR
## Remote:   main @ origin (https://github.com/sjpierce/SSACHR.git)
## Head:     [7ecfb11] 2022-01-15: Updated version number, date, and news.
##


## Untracked files:
##  Untracked:  inst/Step_01_Data_Mgt_Published.pdf
##  Untracked:  inst/Step_02_Analysis_Published_files/
```