

# Reproduced analysis of FARS data

Ethan Walker

October 30, 2017

## Results

### Percentages of drivers testing positive by drug type, sex, and year group

```
table_one <- clean_fars %>%
  mutate(year_cat = cut(year, breaks = c(1999, 2002, 2006, 2010),
    labels = c("1999-2002", "2003-2006", "2007-2010"),
    include.lowest = TRUE, right = TRUE)) %>%
  filter(!is.na(sex)) %>%
  group_by(drug_type, sex, year_cat) %>%
  summarize(n_non_missing = sum(!is.na(positive_for_drug)),
    positive_test = sum(positive_for_drug, na.rm = TRUE),
    perc_positive = round(100 * positive_test / n_non_missing, 1)) %>%
  select(drug_type, sex, year_cat, perc_positive) %>%
  unite(sex_year_cat, sex, year_cat) %>%
  spread(sex_year_cat, perc_positive) %>%
  knitr::kable(col.names = c("Drug type", "F 1999-2002", "F 2003-2006",
    "F 2007-2010", "M 1999-2002", "M 2003-2006",
    "M 2007-2010"))

table_one
```

Drug type	F 1999-2002	F 2003-2006	F 2007-2010	M 1999-2002	M 2003-2006	M 2007-2010
Alcohol	26.4	24.3	27.1	43.2	42.9	43.3
Cannabinoid	2.8	5.7	7.3	5.8	10.3	11.8
Depressant	3.4	3.8	4.8	2.0	2.5	3.2
Narcotic	4.2	4.9	7.0	2.2	3.4	4.0
Other	5.6	6.6	7.2	4.3	4.5	4.2
Stimulant	7.2	9.1	8.7	10.5	11.9	9.2

Figure 1: Prevalence of nonalcohol drugs in fatally injured drivers by year and age group

```
fig_one <- clean_fars %>%
  filter(!is.na(agecat)) %>%
  filter(drug_type != "Alcohol") %>%
  group_by(unique_id, year, agecat) %>%
  summarize(positive_test = sum(positive_for_drug, na.rm = TRUE),
    positive = any(positive_test > 0)) %>%
  ungroup() %>%
  group_by(year, agecat) %>%
  summarize(total_tests = length(positive), positive = sum(positive),
    perc_positive = round(100 * positive / total_tests, 1))
head(fig_one)
```

```
## # A tibble: 6 x 5
## # Groups:   year [2]
##   year      agecat total_tests positive perc_positive
##   <int>    <fctr>      <int>    <int>      <dbl>
## 1  1999    < 25 years      501      72       14.4
## 2  1999  25--44 years      816     160       19.6
## 3  1999  45--64 years      411      49       11.9
## 4  1999  65 years +       246      15        6.1
## 5  2000    < 25 years      478     104       21.8
## 6  2000  25--44 years      825     180       21.8
```

```
fig_one.1 <- clean_fars %>%
  filter(!is.na(agecat)) %>%
  filter(drug_type != "Alcohol") %>%
  group_by(unique_id, year, agecat) %>%
  summarize(positive_for_drug = any(positive_for_drug)) %>%
  ungroup() %>%
  group_by(year, agecat) %>%
  summarize(total_tests = length(positive_for_drug),
            positive = sum(positive_for_drug, na.rm = TRUE),
            perc_positive = round(100 * positive / total_tests, 1))
head(fig_one.1)
```

```
## # A tibble: 6 x 5
## # Groups:   year [2]
##   year      agecat total_tests positive perc_positive
##   <int>    <fctr>      <int>    <int>      <dbl>
## 1  1999    < 25 years      501      72       14.4
## 2  1999  25--44 years      816     160       19.6
## 3  1999  45--64 years      411      49       11.9
## 4  1999  65 years +       246      15        6.1
## 5  2000    < 25 years      478     104       21.8
## 6  2000  25--44 years      825     180       21.8
```

```
fig_one_plot <- fig_one.1 %>%
  ggplot(aes(x = year, y = perc_positive, position = agecat)) +
  geom_line() +
  geom_point(aes(x = year, y = perc_positive, shape = agecat)) +
  scale_shape_manual("Age", values = c(17, 16, 15, 3)) +
  scale_y_continuous(limits = c(0, 35)) +
  theme_few() +
  labs(x = "Year", y = "Positive for Nonalcoholic Drugs, %")
fig_one_plot
```

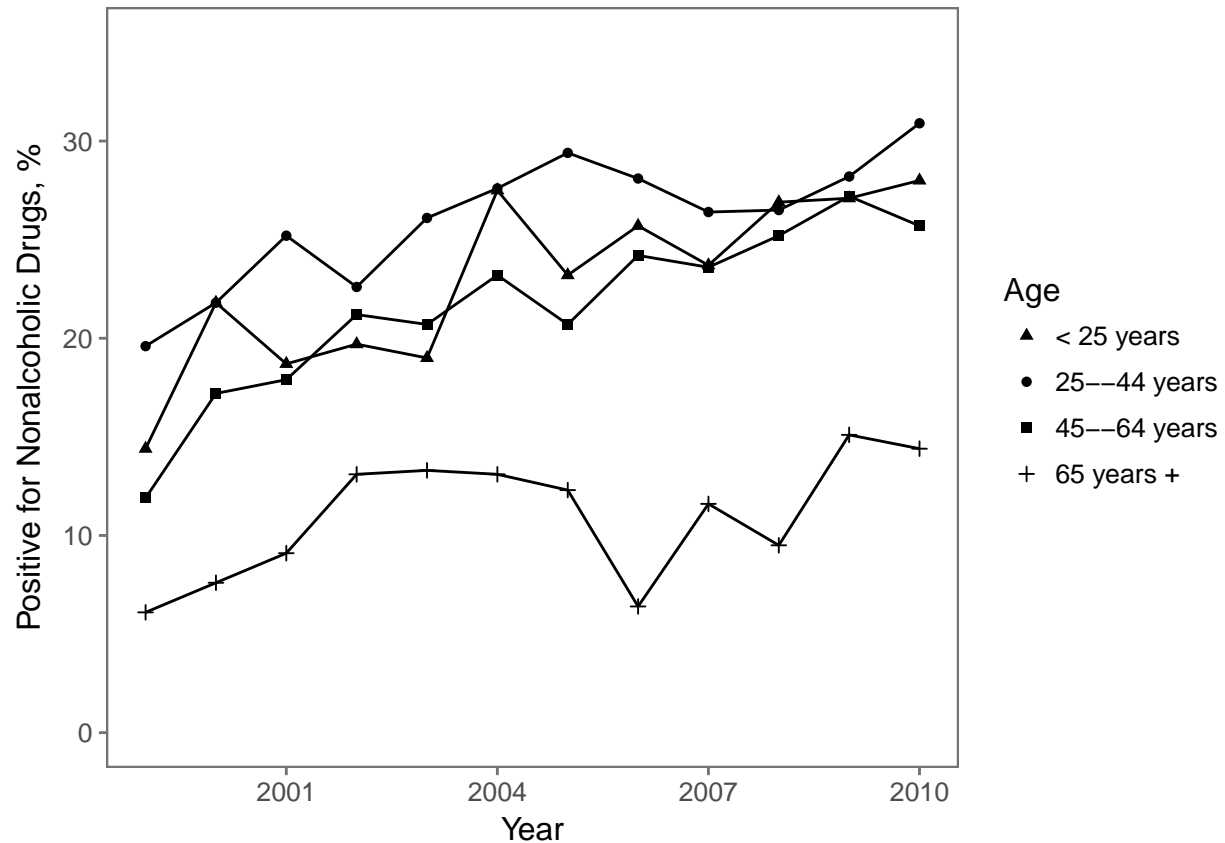
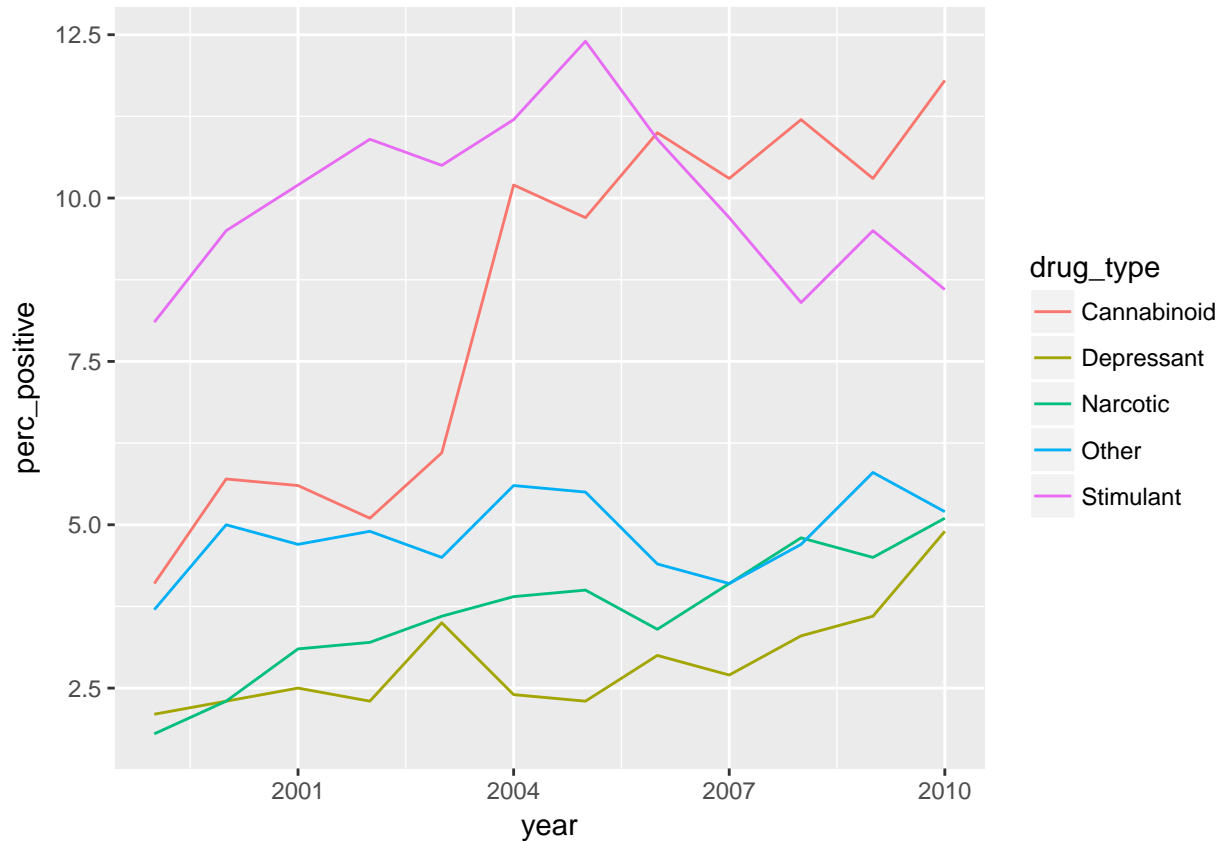


Figure 2: Prevalence of nonalcohol drugs in fatally injured drivers by year and drug type

```
fig_two <- clean_fars %>%
  filter(!is.na(positive_for_drug)) %>%
  filter(drug_type != "Alcohol") %>%
  group_by(year, drug_type) %>%
  summarize(n_non_missing = sum(!is.na(positive_for_drug)),
            positive_test = sum(positive_for_drug, na.rm = TRUE),
            perc_positive = round(100 * positive_test / n_non_missing, 1))

fig_two_plot <- fig_two %>%
  ggplot(aes(x = year, y = perc_positive, colour = drug_type)) +
  geom_line()
fig_two_plot
```



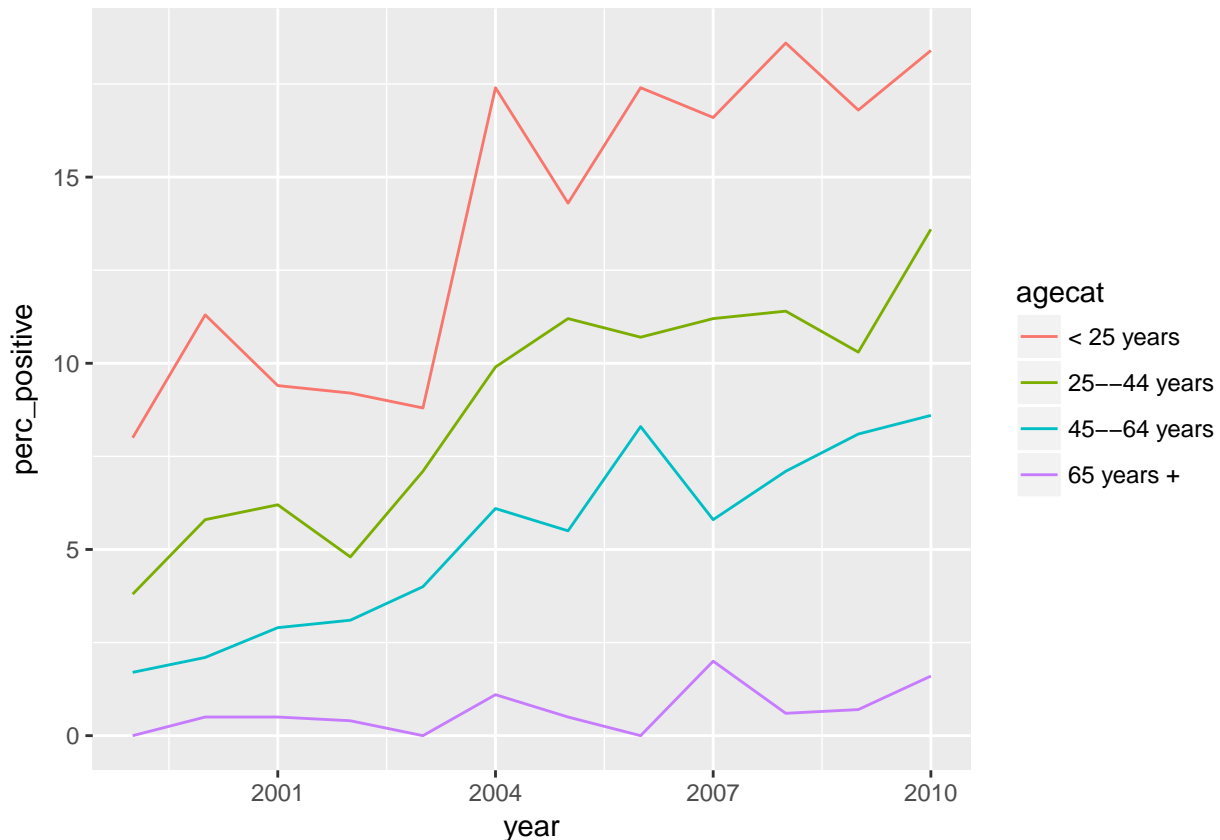
**Figure 3: Prevalence of cannabinoid drugs in fatally injured drivers by year and age group**

```
fig_three <- clean_fars %>%
  filter(!is.na(agecat)) %>%
  filter(drug_type == "Cannabinoid") %>%
  group_by(year, agecat) %>%
  summarize(n_non_missing = sum(!is.na(positive_for_drug)),
    positive_test = sum(positive_for_drug, na.rm = TRUE),
    perc_positive = round(100 * positive_test / n_non_missing, 1))
fig_three
```

```
## # A tibble: 48 x 5
## # Groups:   year [?]
##   year agecat n_non_missing positive_test perc_positive
##   <int> <fctr>         <int>         <int>         <dbl>
## 1 1999 < 25 years      473             38            8.0
## 2 1999 25--44 years  760             29            3.8
## 3 1999 45--64 years  362              6            1.7
## 4 1999 65 years +    202              0            0.0
## 5 2000 < 25 years      453             51           11.3
## 6 2000 25--44 years  757             44            5.8
## 7 2000 45--64 years  438              9            2.1
## 8 2000 65 years +    197              1            0.5
```

```
## 9 2001 < 25 years 511 48 9.4
## 10 2001 25--44 years 800 50 6.2
## # ... with 38 more rows
```

```
fig_three_plot <- fig_three %>%
  ggplot(aes(x = year, y = perc_positive, colour = agecat)) +
  geom_line()
fig_three_plot
```



## Prevalence of drugs in fatally injured drivers for 1999 and 2010 by drug type

```
prev_one <- clean_fars %>%
  filter(year %in% c("1999", "2010")) %>%
  group_by(drug_type, year) %>%
  summarize(positive = sum(positive_for_drug, na.rm = TRUE),
            trials = sum(!is.na(positive_for_drug)))

prev_two <- data.frame(prevalence = perc_cis(x = prev_one$positive,
                                             n = prev_one$trials))

prev_one <- tibble::rowid_to_column(as.data.frame(prev_one), var = "rowid")
prev_two <- tibble::rowid_to_column(as.data.frame(prev_two), var = "rowid")

prev_final <- prev_one %>%
  full_join(prev_two, by = "rowid") %>%
```

```

select(drug_type, year, prevalence) %>%
spread(key = year, value = prevalence) %>%
rename(`Drug type` = drug_type)
knitr::kable(prev_final)

```

Drug type	1999	2010
Alcohol	38.7% (36.5%, 40.9%)	39.1% (36.7%, 41.5%)
Cannabinoid	4.1% (3.1%, 5%)	11.8% (10.2%, 13.4%)
Depressant	2.1% (1.5%, 2.8%)	4.9% (3.8%, 5.9%)
Narcotic	1.8% (1.2%, 2.5%)	5.1% (4%, 6.2%)
Other	3.7% (2.9%, 4.6%)	5.2% (4.1%, 6.3%)
Stimulant	8.1% (6.8%, 9.4%)	8.6% (7.2%, 9.9%)

Statistics for testing for trend in prevalence of drugs over study years by drug type using Cochran-Armitage trend test

```

drug_list <- c("Alcohol", "Nonalcohol", "Narcotic", "Depressant",
              "Stimulant", "Cannabinoid", "Other")
drug_trend_tests_ca <- lapply(drug_list, test_trend_ca)
drug_trend_tests_ca <- dplyr::bind_rows(drug_trend_tests_ca) %>%
  dplyr::mutate(drug = drug_list) %>%
  dplyr::select(drug, Z, p.value)
drug_trend_tests_ca %>% knitr::kable()

```

drug	Z	p.value
Alcohol	1.2	0.228
Nonalcohol	9.9	0.000
Narcotic	6.7	0.000
Depressant	4.7	0.000
Stimulant	0.5	0.604
Cannabinoid	13.6	0.000
Other	1.4	0.157

Statistics for testing for trend in prevalence of drugs over study years by drug type using Wald test of logistic regression coefficient for “year”

```

drug_list <- c("Alcohol", "Nonalcohol", "Narcotic", "Depressant",
              "Stimulant", "Cannabinoid", "Other")
drug_trend_tests_log_reg <- lapply(drug_list, test_trend_log_reg)
drug_trend_tests_log_reg <- dplyr::bind_rows(drug_trend_tests_log_reg) %>%
  dplyr::mutate(drug = drug_list) %>%
  dplyr::select(drug, Z, p.value)
drug_trend_tests_log_reg %>% knitr::kable()

```

drug	Z	p.value
Alcohol	1.2	0.228
Nonalcohol	9.9	0.000
Narcotic	6.6	0.000

drug	Z	p.value
Depressant	4.7	0.000
Stimulant	-0.5	0.604
Cannabinoid	13.5	0.000
Other	1.4	0.158