

PAWS Analysis with R-Ladies - Question 3 additional analysis

Alice Walsh (updated by KF)

8/20/2018

```
# This code works when you are using RStudio project. Paths are relative and stuff  
library(survival)  
library(survminer)  
library(RColorBrewer)  
suppressPackageStartupMessages(library(tidyverse))
```

Question

What is the typical pattern for volunteer disengagement?

I need to summarise the number of volunteers with various characteristics:

- did orientation, but no shifts
- did orientation, had only 1 shift
- did orientation, had 2+ shifts

What factors are associated with these three groups? location, orientation type, zip code, days to first shift?

For volunteers that do show up, what happens? Lets plot frequencies over the following factors:

- orientation type
- orientation location
- did a shift
- did a second shift
- shift types

How do we define a “disengaged” volunteer?

- My idea: Length of time after last shift is > x months (test different x)

Procedure

1. Load in the data
2. Analyze

Load Data

```
final_data <- readRDS("../Data/merged.Rds")
```

Analysis

Let's try time-to-event analysis. Here we will look at data for volunteers who did orientation between Jan 1, 2018 and May 1, 2018. For this analysis, only volunteers that did at least one shift are considered and a volunteer is defined as "disengaged" if greater than 90 days passes without volunteering.

```
final_data$event <- 0
final_data$event[final_data$days_since_last > 90] <- 1

# create factor for whether they reported for shift within 14 days
final_data$first_shift_group <- case_when(
  as.numeric(final_data$time_to_first) <= 14 ~ "<= 14 days",
  as.numeric(final_data$time_to_first) <= 60 ~ "14 to 60 days",
  as.numeric(final_data$time_to_first) > 60 ~ "> 60 days"
)
final_data$first_shift_group <- factor(final_data$first_shift_group,
  levels = c("<= 14 days", "14 to 60 days", "> 60 days"))

final_data$timebw_shift_group <- case_when(
  as.numeric(final_data$time_bw_median) <= 7 ~ "<= 7 days",
  as.numeric(final_data$time_bw_median) <= 14 ~ "7 to 14 days",
  as.numeric(final_data$time_bw_median) > 14 ~ "> 14 days"
)
final_data$timebw_shift_group <- factor(final_data$timebw_shift_group,
  levels = c("<= 7 days", "7 to 14 days", "> 14 days"))

final_data %>%
  filter(Orientation.Date.Primary <= "2018-05-01", total_shifts > 0) %>%
  group_by(first_shift_group) %>%
  summarise(mean_shifts = mean(total_shifts),
    mean_missed_shifts = mean(missed_shifts),
    freq = n()) %>%
  knitr::kable(caption = "Summary of time orientation to first shift")

## Warning in strptime(xx, f <- "%Y-%m-%d", tz = "GMT"): unknown timezone
## 'zone/tz/2018e.1.0/zoneinfo/America/New_York'
```

Table 1: Summary of time orientation to first shift

first_shift_group	mean_shifts	mean_missed_shifts	freq
<= 14 days	7.767123	0.1643836	73
14 to 60 days	6.089744	0.1410256	78
> 60 days	10.133333	0.1333333	15

```
final_data %>%
  filter(Orientation.Date.Primary <= "2018-05-01", total_shifts > 0) %>%
  group_by(timebw_shift_group) %>%
  summarise(mean_shifts = mean(total_shifts),
    mean_missed_shifts = mean(missed_shifts),
    freq = n()) %>%
  knitr::kable(caption = "Summary of time between shifts")
```

Table 2: Summary of time between shifts

timebw_shift_group	mean_shifts	mean_missed_shifts	freq
<= 7 days	11.740260	0.1948052	77
7 to 14 days	5.038462	0.1153846	26
> 14 days	3.742857	0.1714286	35
NA	1.000000	0.0357143	28

Plot volunteers time between first shift and disengagement (time_last_to_first)

```

model_data <- filter(final_data, total_shifts > 1, Orientation.Date.Primary <= "2018-05-01")
surv.model <- Surv(model_data$time_last_to_first, model_data$event)

surv.fit <- survfit(surv.model ~ first_shift_group, data = model_data)
surv.fit

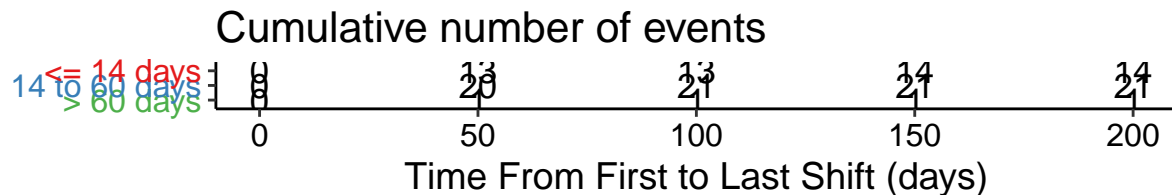
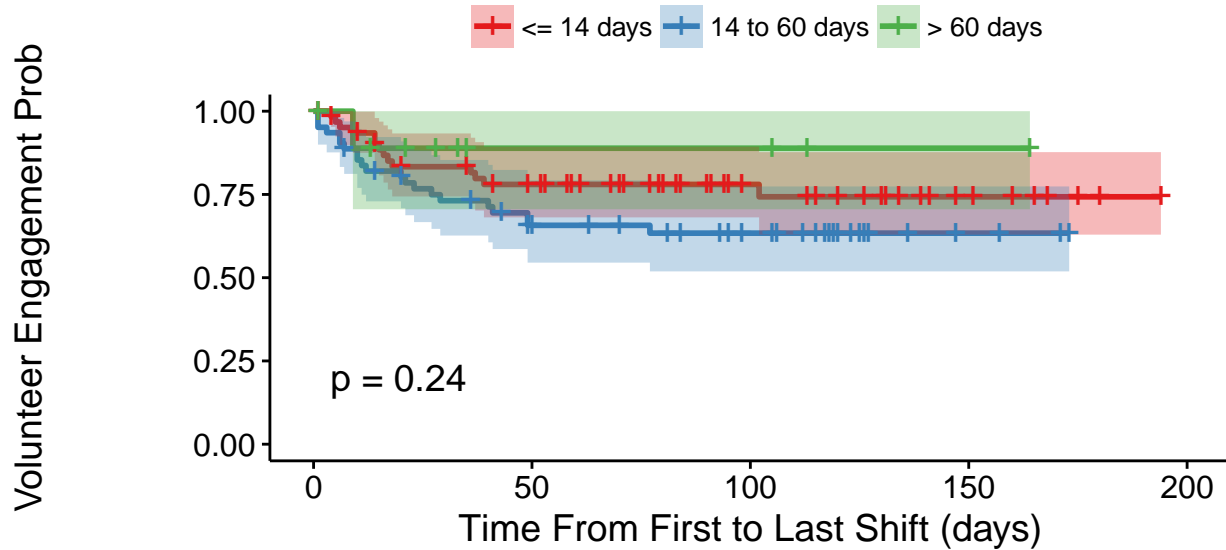
## Call: survfit(formula = surv.model ~ first_shift_group, data = model_data)
##
##              n events median 0.95LCL 0.95UCL
## first_shift_group=<= 14 days   63     14    NA      NA      NA
## first_shift_group=14 to 60 days 62     21    NA      NA      NA
## first_shift_group=> 60 days   10      1    NA      NA      NA

ggsurvplot(surv.fit, data = model_data,
            surv.median.line = "hv",
            pval = TRUE, conf.int = TRUE, cumevents = TRUE,
            ylab = "Volunteer Engagement Prob",
            palette = brewer.pal(n=9, "Set1"),
            xlab = "Time From First to Last Shift (days)",
            title = "Compare how soon after orientation volunteered",
            legend.title = "", legend.labs = levels(factor(model_data$first_shift_group)))

## Warning in .add_surv_median(p, fit, type = surv.median.line, fun = fun, :
## Median survival not reached.

```

Compare how soon after orientation volunteered



```
surv.fit <- survfit(surv.model ~ timebw_shift_group, data = model_data)
surv.fit
```

```
## Call: survfit(formula = surv.model ~ timebw_shift_group, data = model_data)
```

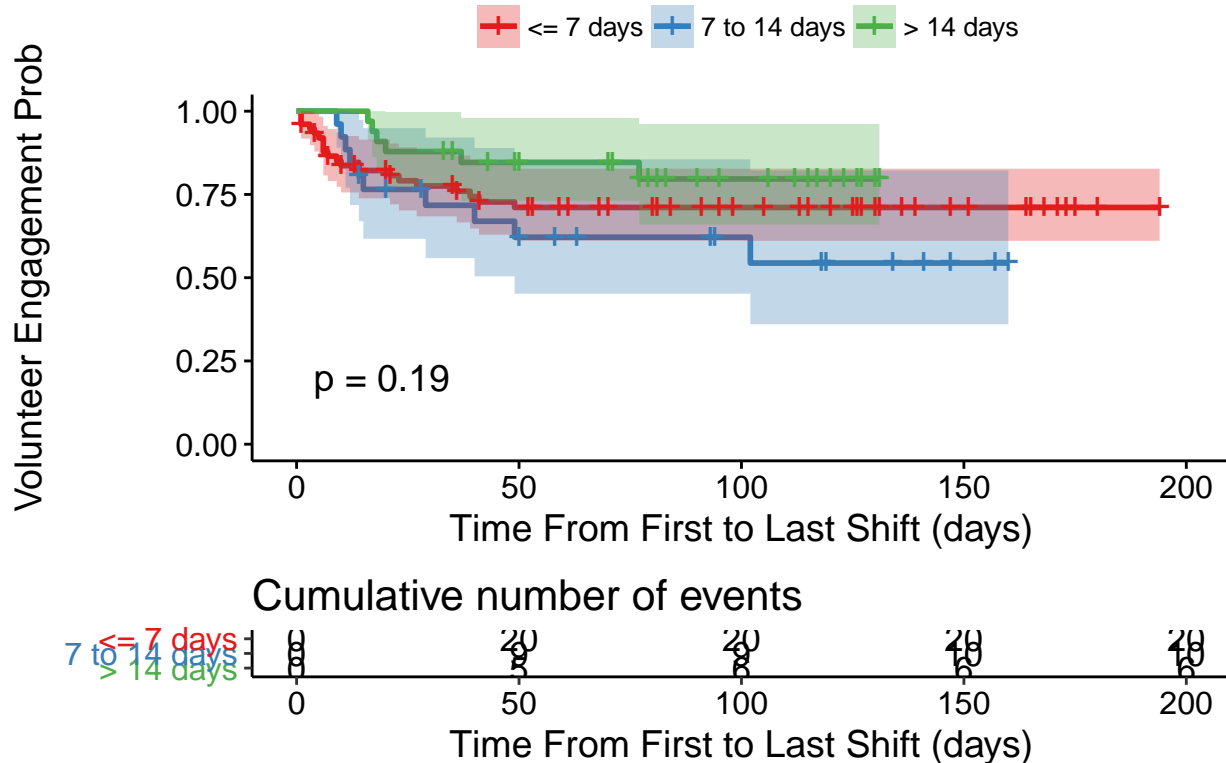
```
##
```

```
##          n events median 0.95LCL 0.95UCL
## timebw_shift_group=<= 7 days    76     20     NA      NA      NA
## timebw_shift_group=7 to 14 days  26     10     NA     49     NA
## timebw_shift_group=> 14 days    33      6     NA     NA     NA
```

```
ggsurvplot(surv.fit, data = model_data,
  surv.median.line = "hv",
  pval = TRUE, conf.int = TRUE, cumevents = TRUE,
  ylab = "Volunteer Engagement Prob",
  palette = brewer.pal(n=9, "Set1"),
  xlab = "Time From First to Last Shift (days)",
  title = "Compare median time between shifts",
  legend.title = "", legend.labs = levels(factor(model_data$timebw_shift_group)))
```

```
## Warning in .add_surv_median(p, fit, type = surv.median.line, fun = fun, :
## Median survival not reached.
```

Compare median time between shifts



We don't see a strong trend that volunteers leave sooner if they tend to volunteer more frequently or if they first volunteer soon or later after they complete orientation.

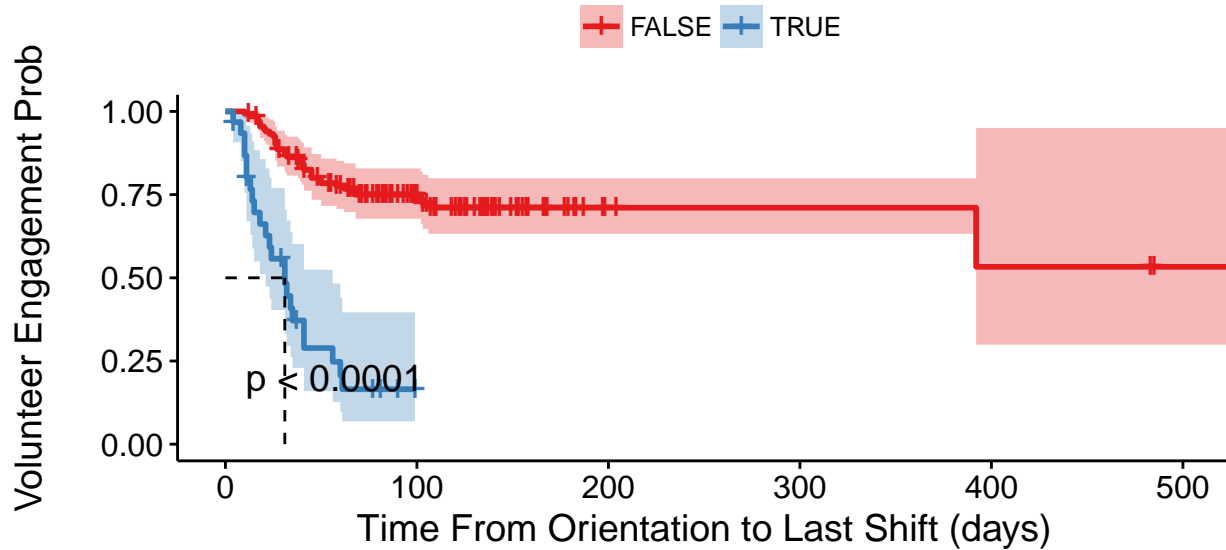
```
# Could plot either total_time - time from orientation to last shift
# or time_last_to_first - time from first shift to last shift

model_data <- filter(final_data, total_shifts > 0, Orientation.Date.Primary <= "2018-05-01")
surv.model <- Surv(model_data$total_time, model_data$event)
surv.fit <- survfit(surv.model ~ only_one_shift, data = model_data)
surv.fit

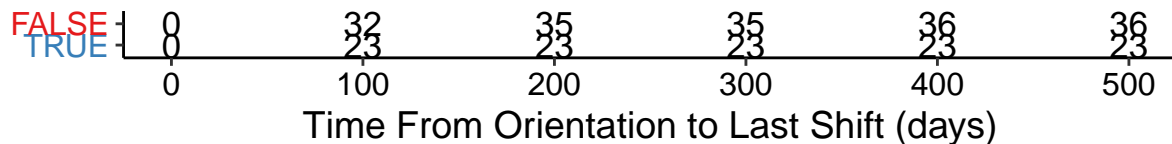
## Call: survfit(formula = surv.model ~ only_one_shift, data = model_data)
##
##              n events median 0.95LCL 0.95UCL
## only_one_shift=FALSE 135     36    NA     392     NA
## only_one_shift=TRUE   31     23    31      21     56

ggsurvplot(surv.fit, data = model_data,
  surv.median.line = "hv",
  pval = TRUE, conf.int = TRUE, cumevents = TRUE,
  ylab = "Volunteer Engagement Prob",
  palette = brewer.pal(n=9, "Set1"),
  xlab = "Time From Orientation to Last Shift (days)",
  title = "Compare whether only did one shift",
  legend.title = "", legend.labs = levels(factor(model_data$only_one_shift)))
```

Compare whether only did one shift



Cumulative number of events



```
# top assignments:
top_assign <- names(sort(table(final_data$primary_type),decreasing = T))[1:3]
model_data <- filter(final_data, total_shifts > 1,
  Orientation.Date.Primary <= "2018-05-01",
  primary_type %in% top_assign)

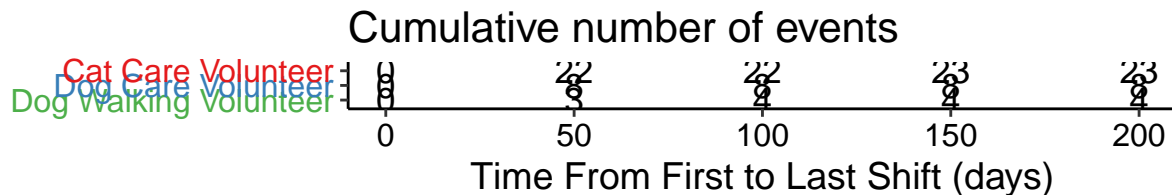
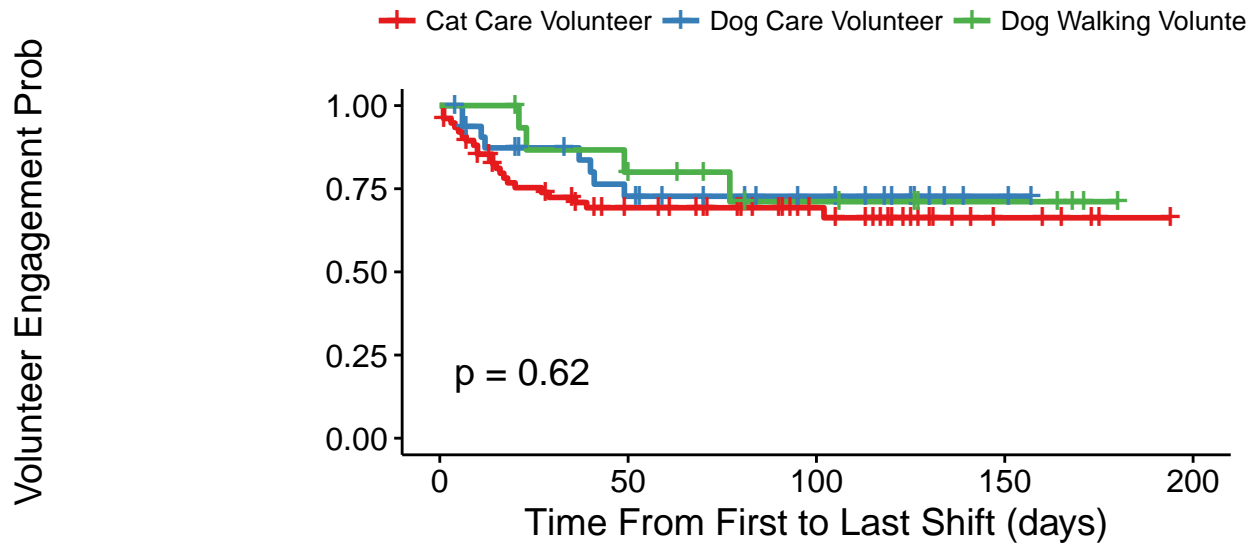
surv.model <- Surv(model_data$time_last_to_first, model_data$event)
surv.fit <- survfit(surv.model ~ primary_type, data = model_data)
surv.fit

## Call: survfit(formula = surv.model ~ primary_type, data = model_data)
##
##
##               n events median 0.95LCL 0.95UCL
## primary_type=Cat Care Volunteer    77     23    NA      NA      NA
## primary_type=Dog Care Volunteer    33      8    NA      NA      NA
## primary_type=Dog Walking Volunteer  16      4    NA      NA      NA

ggsurvplot(surv.fit, data = model_data,
  surv.median.line = "hv",
  pval = TRUE, conf.int = FALSE, cumevents = TRUE,
  palette = brewer.pal(n=9, "Set1"),
  ylab = "Volunteer Engagement Prob", xlab = "Time From First to Last Shift (days)",
  title = "Compare primary assignments",
  legend.title = "", legend.labs = levels(factor(model_data$primary_type)))

## Warning in .add_surv_median(p, fit, type = surv.median.line, fun = fun, :
## Median survival not reached.
```

Compare primary assignments



```
model_data <- filter(final_data, total_shifts > 1, Orientation.Date.Primary <= "2018-05-01")
surv.model <- Surv(model_data$time_last_to_first, model_data$event)
surv.fit <- survfit(surv.model ~ orient_loc.Primary, data = model_data)
surv.fit
```

```
## Call: survfit(formula = surv.model ~ orient_loc.Primary, data = model_data)
```

```
##
```

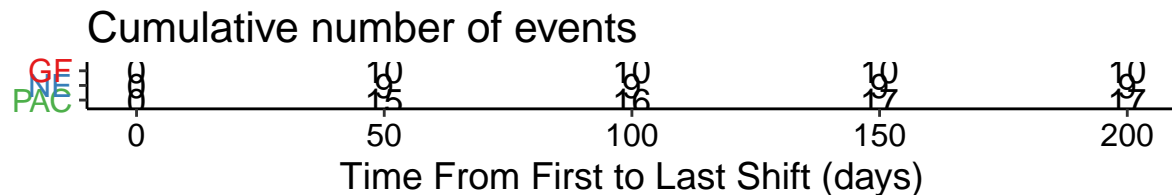
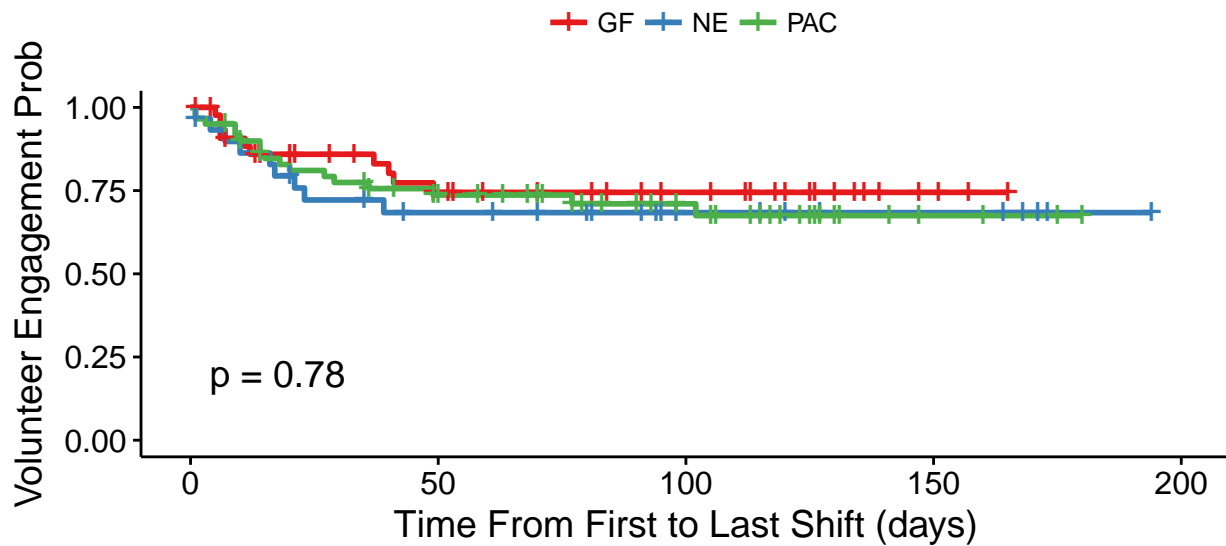
```
##          n events median 0.95LCL 0.95UCL
## orient_loc.Primary=GF  45      10      NA      NA      NA
## orient_loc.Primary=NE  30       9      NA      NA      NA
## orient_loc.Primary=PAC 60      17      NA      NA      NA
```

```
ggsurvplot(surv.fit, data = model_data,
  surv.median.line = "hv",
  pval = TRUE, conf.int = FALSE, cumevents = TRUE,
  palette = brewer.pal(n=9, "Set1"),
  ylab = "Volunteer Engagement Prob", xlab = "Time From First to Last Shift (days)",
  title = "Compare primary orientation location",
  legend.title = "", legend.labs = levels(factor(model_data$orient_loc.Primary)))
```

```
## Warning in .add_surv_median(p, fit, type = surv.median.line, fun = fun, :
```

```
## Median survival not reached.
```

Compare primary orientation location



```
model_data <- filter(final_data, total_shifts > 1, Orientation.Date.Primary <= "2018-05-01")
surv.model <- Surv(model_data$time_last_to_first, model_data$event)
surv.fit <- survfit(surv.model ~ city_clean, data = model_data)
surv.fit
```

```
## Call: survfit(formula = surv.model ~ city_clean, data = model_data)
```

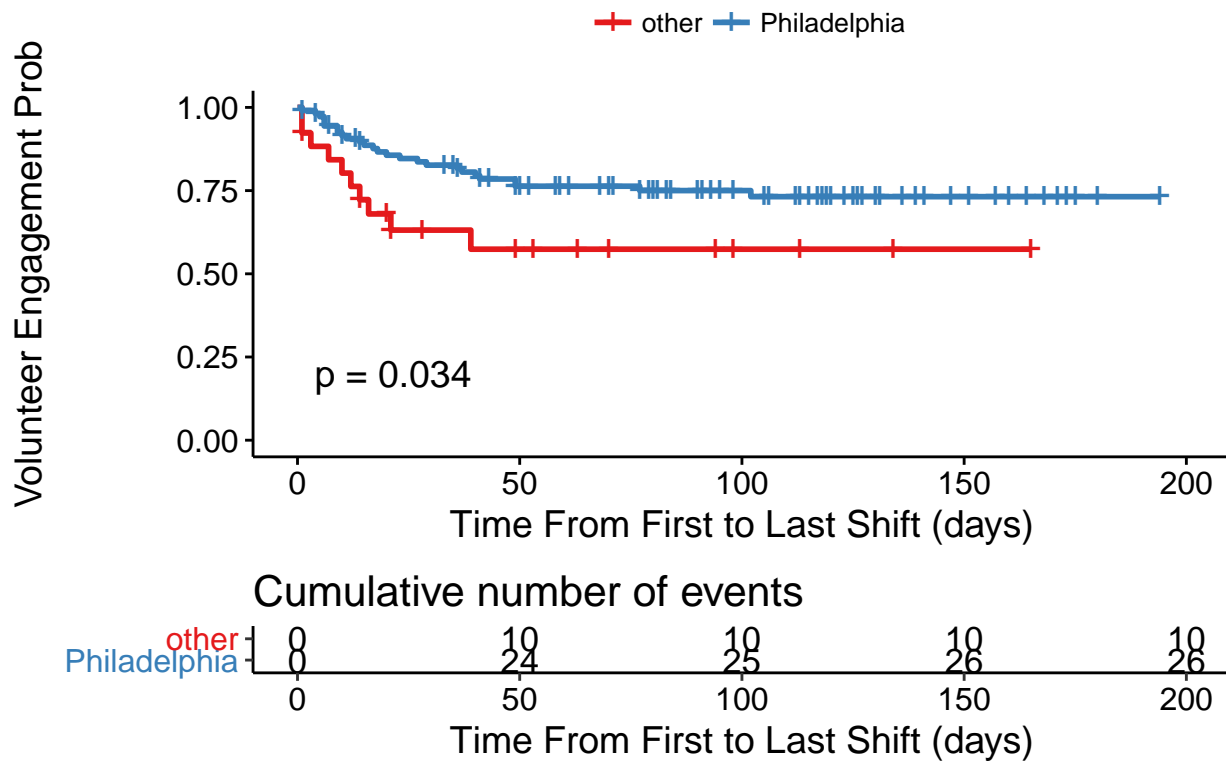
```
##
```

```
##              n events median 0.95LCL 0.95UCL
## city_clean=other      26      10      NA      21      NA
## city_clean=Philadelphia 109      26      NA      NA      NA
```

```
ggsurvplot(surv.fit, data = model_data,
  surv.median.line = "hv",
  pval = TRUE, conf.int = FALSE, cumevents = TRUE,
  palette = brewer.pal(n=9, "Set1"),
  ylab = "Volunteer Engagement Prob", xlab = "Time From First to Last Shift (days)",
  title = "Compare primary orientation location",
  legend.title = "", legend.labs = levels(factor(model_data$city_clean)))
```

```
## Warning in .add_surv_median(p, fit, type = surv.median.line, fun = fun, :
## Median survival not reached.
```


Compare primary orientation location



There is not a clear trend for orientation location or whether someone lives in philadelphia or outside philadelphia.

There is a trend that volunteers that typically volunteer as dog walking volunteers stay engaged longer than cat care of dog care volunteers. Cat care is at all three sites, dog care only at Grays Ferry and dog walking is either PAC or NE.