Simulation experiment for "*An evaluation of the Suspect Target Management Plan, October 2020*" Study

Dr. Gordana Popovic, UNSW, 5 November 2020

Aim

The purpose of this simulation is to see what data and analysis might look like if there was no effect of STMP on offending rates of individuals, both in terms of the plots (Figure 1a and b) and any analysis.

Executive summary

- Synthetic data was simulated in which offending did not change after entry to STMP
- Offenders in the synthetic data were recruited to STMP (a random number of days) following an offence
- Analysis from the Study was recreated with this synthetic data
- Though there was explicitly **no change** in offending following recruitment in the synthetic data, the analysis nevertheless showed strong evidence of a significant decrease in offending after recruitment, and figures consistent with the Study.
- The result in the synthetic data is an artifact of sampling, as offenders were recruited after an offence, the rate of offending just prior to recruitment is artificially increased.
- A further simulation demonstrated that even when offending **increases** post recruitment, the analysis will show strong evidence of a decrease, and figures consistent with the Study.

Method

I simulate a constant rate of offending for 10000 individuals, using an exponential distribution. The time point where individuals start STMP will be allocated a random waiting time after a randomly chosen contact with the justice system (offence), however I will not alter the offending rate of individuals in any way after they start STMP.

Some preliminaries:

- we model 10000 individuals.
- we allow each individual to have their own rate of offending, with an average over individuals of `lambda = 200` days between offences, and a standard deviation of `sig=10` between individual offending rates. Rates are kept constant over time for each individual.
- we model `n_off=20` offences to have enough to cover the two year time span, we then remove any data outside the one year before and after STMP window.
- each individual is put onto STMP an average of `wait=100` days after a randomly selected offence.

```
library(dplyr)
library(tidyr)
library(ggplot2)
N=10000 # number of idividuals
lambda=200 #average number of days beween offences
sig=10 #individual variation in ofending rate,
# can be set to 0 for constant offending rate across individuals
```

```
n_off=20 #number of offenses modelled
wait=100 #average number of days from random offence to being put on STMP
```

Simulation

Simulating days of offences relative to entry into STMP, with no effect of STMP on offending.

```
d=matrix(NA,N,(n_off))
for(i in 1:N){
  lambda_ind=abs(rnorm(1,lambda,sig)) #individual offending rate centered around lambda
  abs=round(cumsum(rexp(n_off, rate=1/lambda_ind)),0) #day of each offense
  d_stmp<-round(sample(abs,1)+rexp(1,1/wait),0) #date of going on STMP
  rel=abs-d_stmp #day relative to STMP starting
  d[i,]=rel
}
dat=data.frame(id=1:N,d)
```

Data organisation

We then reorganize the data to long format, where there is now one column for id, one for the relative time and one for offending occasion.

```
dat_long <- dat %>%
  pivot_longer(-c(1:2), values_to= "days_before_after") %>%
  mutate(occasion=as.numeric(substr(name,2,3))) %>%
  mutate(time=ifelse(days_before_after>0,1,0)) %>%
  select(id,occasion,days_before_after,time)   %>%
  filter(days_before_after>-365) %>% #subset to 1 year before and after STMP
  filter(days_before_after<365)
dat_long
```

```
## # A tibble: 41,392 x 4
##       id occasion days_before_after  time
##    <int>    <dbl>             <dbl> <dbl>
##  1     1       14              -109     0
##  2     1       15               -27     0
##  3     1       16               300     1
##  4     2       17              -304     0
##  5     2       18              -186     0
##  6     2       19              -184     0
##  7     2       20               -11     0
##  8     3       18              -209     0
##  9     3       19                -7     0
## 10     3       20               178     1
## # ... with 41,382 more rows
```

Results

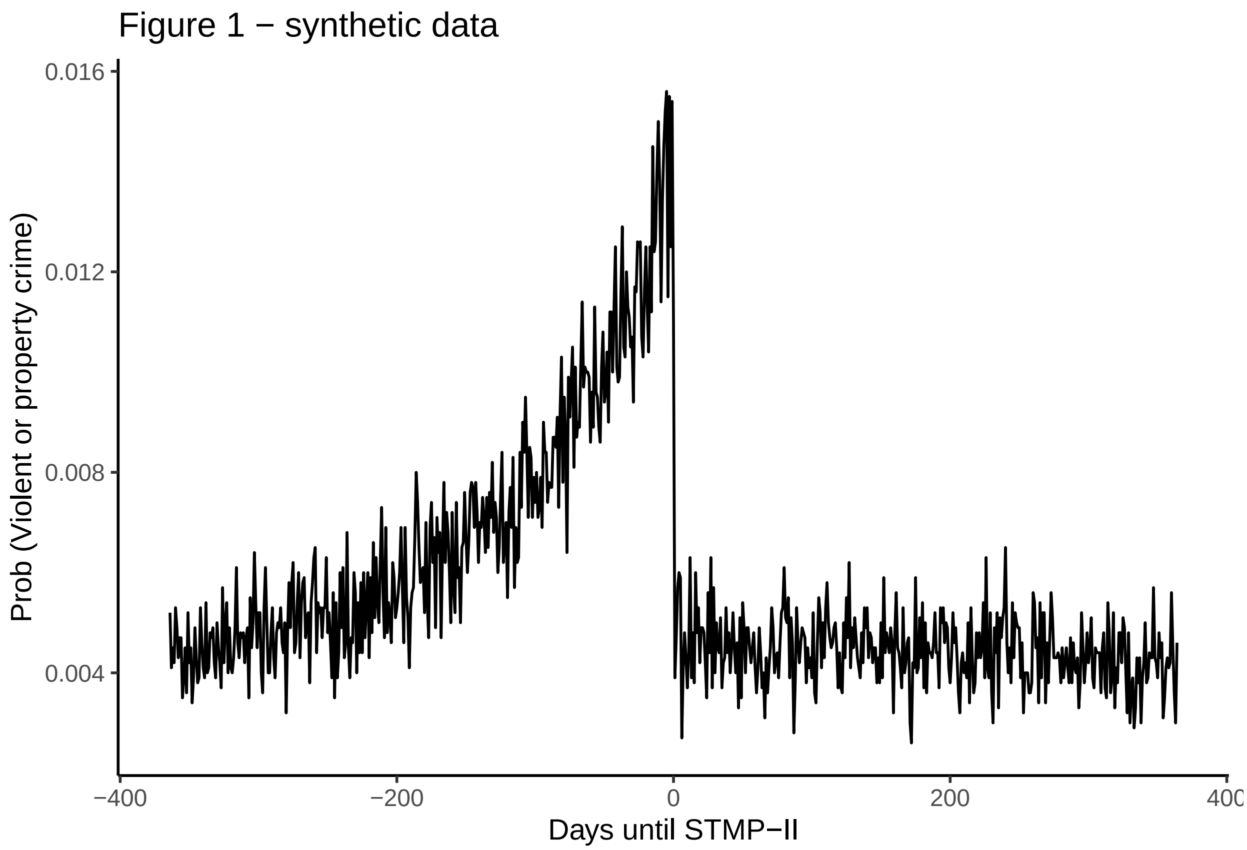
Figure 1

It is now straight forward to calculate the proportion of individuals who offended on each day.

```
crime_rates<-dat_long %>%
  group_by(days_before_after) %>%
  count()
```

And plot this.

```r
crime_rates %>%
  ggplot(aes(days_before_after,n/N))+
  geom_line()+
  xlab("Days until STMP-II")+
  ylab("Prob (Violent or property crime)")+
  ggtitle("Figure 1 - synthetic data")+
  xlim(-365,365)+
  theme_classic()
```



Regression

We first calculate for each individual whether they offended in the year leading up to, and the year following them starting STMP.

```r
mod_dat=dat_long %>%
  group_by(id) %>%
  summarise(before=(sum(time==0)>0)*1, #any offence in year before STMP
            after=(sum(time==1)>0)*1) %>% #any offence in year after STMP
  pivot_longer(-1, names_to = "time",values_to="offended") %>%
  mutate(time=relevel(factor(time), ref="before"))

head(mod_dat)
```

```
## # A tibble: 6 x 3
##      id time   offended
##   <int> <fct>     <dbl>
```

```
## 1      1 before        1
## 2      1 after         1
## 3      2 before        1
## 4      2 after         0
## 5      3 before        1
## 6      3 after         1
```

We then reproduce (a simplified version) of the analysis in the Study.

```
summary(glm(offended~ id+time, data=mod_dat, family=binomial) )
```

```
##
## Call:
## glm(formula = offended ~ id + time, family = binomial, data = mod_dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.6317   0.2537   0.2564   0.6994   0.7078
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.389e+00  6.884e-02  49.233   <2e-16 ***
## id           4.193e-06  7.744e-06   0.541    0.588
## timeafter   -2.133e+00  6.205e-02 -34.371   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 14939  on 19827  degrees of freedom
## Residual deviance: 13203  on 19825  degrees of freedom
## AIC: 13209
##
## Number of Fisher Scoring iterations: 6
```

We find strong evidence ($p<0.001$) of significant decrease in offending after STMP.

Conclusion

Despite simulating data with **no** effect of STMP, we see very similar results from this simulation and the Study.

- The shape of *Figure 1 - synthetic data* is very similar to Figure 1a, and particularly Figure 1b from the Study.
- There is strong evidence ($p<0.001$) of significant decrease in offending after STMP for the synthetic data, though we know that offending does not change.

We conclude that the results in the Study are consistent with a population without any effect of STMP. It seems likely the effect is an artifact of the timing of entry to STMP being after contact with the justice system.

Appendix

Simulation 2 - increased offending after STMP

This simulation aims to find a reason why in Figure 1 the rate of offending after STMP reduces over time. It seems plausible that this is because recruiting offenders to STMP temporarily increases offending, or makes them more likely to be caught, for a period following their recruitment, which thereafter goes back to their usual offending rate. The simulation below explores this scenario.

```
d=matrix(NA,N,(4*n_off))
for(i in 1:N){
  lambda_ind=abs(rnorm(1,lambda,sig)) #individual offending rate normal centered around lambda
  between_pre=rexp(n_off/2, rate=1/lambda_ind) #usual offending rate
  between_post1<-rexp(n_off/4, rate=2/(lambda_ind)) # double offending rate
  between_post2<-rexp(n_off/4, rate=1/(lambda_ind)) #usual offending rate
  abs=round(cumsum(c(between_pre,between_post1,between_post2)),0) #day of each offense
  d_stmp<-round(sum(between_pre)+rexp(1,1/wait),0) #date of going on STMP
  rel=abs-d_stmp #day relative to STMP starting
  d[i,]=rel
}
dat=data.frame(id=1:N,d)
```

And plot this.

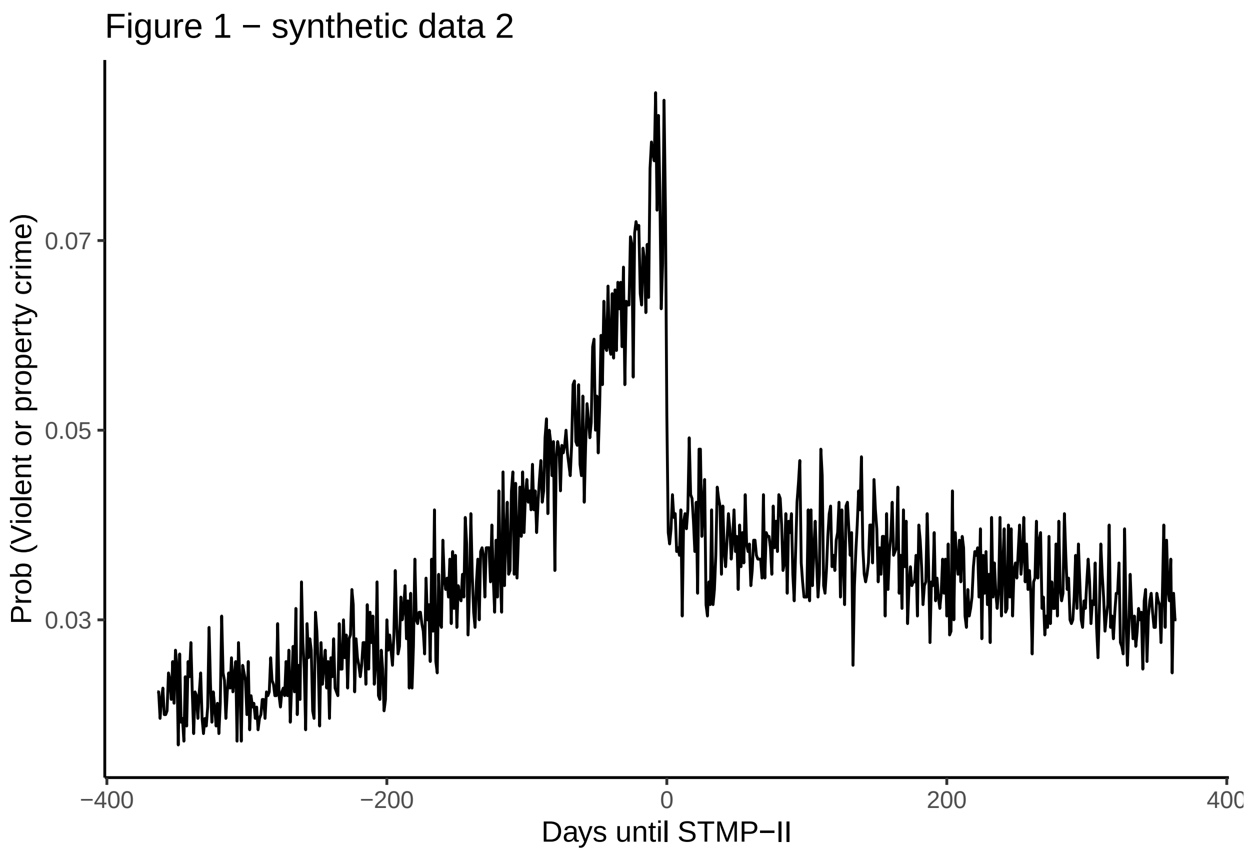


Figure 1 - synthetic data 2 has many of the properties of Figure 1a and 1b in the Study, including the apparent increase in offending prior to STMP, an apparent sharp decrease after STMP starts, and a gradual decrease in the following year. This figure was produced with synthetic data where there is no change in offending prior to STMP, then a temporary **increase** in offending after recruitment to STMP.

```
summary(glm(offended~ id+time, data=mod_dat, family=binomial) )
```

```
##
## Call:
## glm(formula = offended ~ id + time, family = binomial, data = mod_dat)
```

```
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.7318   0.0467   0.0505   0.2487   0.2694
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  6.969e+00  3.191e-01  21.835   <2e-16 ***
## id          -3.142e-05  1.980e-05  -1.586    0.113
## timeafter   -3.356e+00  3.072e-01 -10.924   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3254.1  on 19919  degrees of freedom
## Residual deviance: 2903.2  on 19917  degrees of freedom
## AIC: 2909.2
##
## Number of Fisher Scoring iterations: 9
```

We still find strong evidence ($p<0.001$) of significant **decrease** in offending after STMP, even though the rate of offending has actually **increased**.

This result is consistent with the findings of the matched group analysis (page 24 of the Study) which found that *individuals subject to STMP offend at much higher rates than their matched counterparts.*

Last compiled on 05 November, 2020