### eAppendix 1

Place level covariates were derived from the sources detailed below.

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Source</th>
<th>Temporal level</th>
<th>Geographic level</th>
<th>Analytic steps</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weather</td>
<td>Center WRC. WestMap Climate Analysis, PRISM Climate Mapping Program. <a href="http://www.cefa.dri.edu/Westmap/">http://www.cefa.dri.edu/Westmap/</a>, 2016.</td>
<td>Monthly</td>
<td>County</td>
<td>Census places cross county boundaries. Using a county-to-place geographic crosswalk based on 2010 census data (<a href="http://mcdc.missouri.edu/websas/geocorr14.html">http://mcdc.missouri.edu/w ebsas/geocorr14.html</a>), we assigned covariate values for each census place corresponding to the county making up the largest proportion of the population in the census place. Precipitation, temperature and unemployment were entered in the models as continuous variables.</td>
</tr>
</tbody>
</table>
| Civil unrest events       | [https://en.wikipedia.org/wiki/List_of_incidents_of_civil_unrest_in_the_United_States](https://en.wikipedia.org/wiki/List_of_incidents_of_civil_unrest_in_the_United_States) All California events listed were verified based on news reports:  
  - Anaheim, CA, July 2012  
- Los Angeles, CA, May 2007
- Los Angeles, CA, September 2010
- Oakland, CA, January 2009
- Oakland, CA, July 2010
- Oakland, CA, November 2010
- Oakland, CA, October 2011
- San Bernardino, CA, March, 2006
  - [http://articles.latimes.com/2006/mar/06/local/me-riot6](http://articles.latimes.com/2006/mar/06/local/me-riot6)
- Santa Cruz, CA, May 2010
  - [https://www.mercurynews.com/2010/05/02/at-least-18-santa-cruz-businesses-suffered-damage-during-may-day-riot/](https://www.mercurynews.com/2010/05/02/at-least-18-santa-cruz-businesses-suffered-damage-during-may-day-riot/)
<table>
<thead>
<tr>
<th>Socioeconomic and demographic characteristics</th>
<th>American Community Survey 5-year and 1-year estimates</th>
<th>Yearly</th>
<th>Census place</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Percent of households receiving food stamps</td>
<td></td>
<td></td>
<td>We used the 1-year ACS estimates for places where they were available—this includes places with population size over approximately 100,000.</td>
</tr>
<tr>
<td>• Percent of families with no workers in the past 12 months</td>
<td></td>
<td></td>
<td>We used the 2005-2009 year file for years 2005-2009 when there was no 1-year file for that place. Going forward, we used the 2006-2010 data for 2010, 2007-2011 data for 2011, and 2008-2012 data for 2012.</td>
</tr>
<tr>
<td>• Percent of population age 25+ with a bachelor's degree as highest level of educational attainment</td>
<td></td>
<td></td>
<td>Due to changes in the definitions of census places before and after 2010, for 4.2% of places we had to carry covariate data forward or backwards to include all of the years.</td>
</tr>
<tr>
<td>• Percent of population age 15+ never married</td>
<td></td>
<td></td>
<td>All socioeconomic and demographic characteristics were entered in the models as continuous variables.</td>
</tr>
<tr>
<td>• Percent of population white alone, non-Hispanic</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
To identify the American Community Survey covariates most important to control when examining community violence, we used a combination of LASSO and random forest algorithms to identify the subset of variables that were most predictive of community violence. We used cross-validation to avoid over-fitting. The top five categories represented in the most predictive set of variables were 1) poverty, 2) unemployment, 3) marital status, 4) racial/ethnic composition, 5) educational attainment. Consequently, we controlled these strongest predictors of community violence by adjusting for the ACS variables above.
### eTable 1. Autocorrelation Functions (ACF) for Kalman Smoother Residual Outcomes and the Community Violence Exposure at Different Lags, and the Extent of Persistent Autocorrelation \(^{a}\), across places, California 2005-2013

<table>
<thead>
<tr>
<th>Condition</th>
<th>Outcome</th>
<th>ACF at 1-month lag</th>
<th>ACF at 2-month lag</th>
<th>ACF at 3-month lag</th>
<th>ACF at 12-month lag</th>
<th>% with Persistent Autocorrelation (^{a})</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median % &gt; 0.20 or &lt; -0.20</td>
<td>Median % &gt; 0.20 or &lt; -0.20</td>
<td>Median % &gt; 0.20 or &lt; -0.20</td>
<td>Median % &gt; 0.20 or &lt; -0.20</td>
<td>Median % &gt; 0.20 or &lt; -0.20</td>
<td></td>
</tr>
<tr>
<td>Anxiety disorders</td>
<td>Combined fatal and non-fatal</td>
<td>0.040</td>
<td>1.0%</td>
<td>0.048</td>
<td>1.1%</td>
<td>0.054</td>
</tr>
<tr>
<td>Episodic mood disorders</td>
<td>Combined fatal and non-fatal</td>
<td>0.031</td>
<td>0.5%</td>
<td>0.053</td>
<td>1.4%</td>
<td>0.057</td>
</tr>
<tr>
<td>Substance use</td>
<td>Combined fatal and non-fatal</td>
<td>0.040</td>
<td>0.6%</td>
<td>0.051</td>
<td>1.1%</td>
<td>0.056</td>
</tr>
<tr>
<td>Substance use</td>
<td>Fatal only</td>
<td>0.030</td>
<td>9.2%</td>
<td>0.038</td>
<td>10.1%</td>
<td>0.040</td>
</tr>
<tr>
<td>Asthma</td>
<td>Combined fatal and non-fatal</td>
<td>0.077</td>
<td>7.9%</td>
<td>0.053</td>
<td>2.2%</td>
<td>0.058</td>
</tr>
<tr>
<td>Asthma</td>
<td>Fatal only</td>
<td>0.030</td>
<td>21.2%</td>
<td>0.035</td>
<td>21.9%</td>
<td>0.030</td>
</tr>
<tr>
<td>COPD</td>
<td>Combined fatal and non-fatal</td>
<td>0.117</td>
<td>18.4%</td>
<td>0.071</td>
<td>6.7%</td>
<td>0.054</td>
</tr>
<tr>
<td>COPD</td>
<td>Fatal only</td>
<td>0.043</td>
<td>1.0%</td>
<td>0.049</td>
<td>2.2%</td>
<td>0.057</td>
</tr>
<tr>
<td>Acute MI</td>
<td>Combined fatal and non-fatal</td>
<td>0.036</td>
<td>0.3%</td>
<td>0.050</td>
<td>1.1%</td>
<td>0.058</td>
</tr>
<tr>
<td>Acute MI</td>
<td>Fatal only</td>
<td>0.035</td>
<td>0.6%</td>
<td>0.055</td>
<td>2.1%</td>
<td>0.053</td>
</tr>
<tr>
<td>Heart failure</td>
<td>Combined fatal and non-fatal</td>
<td>0.051</td>
<td>2.9%</td>
<td>0.053</td>
<td>1.4%</td>
<td>0.055</td>
</tr>
<tr>
<td>Heart failure</td>
<td>Fatal only</td>
<td>0.036</td>
<td>1.4%</td>
<td>0.051</td>
<td>2.9%</td>
<td>0.057</td>
</tr>
<tr>
<td>Community Violence</td>
<td>Combined fatal and non-fatal</td>
<td>0.033</td>
<td>0.2%</td>
<td>0.046</td>
<td>0.6%</td>
<td>0.057</td>
</tr>
</tbody>
</table>

\(^{a}\) persistent autocorrelation defined as having 6 or more of the first 12 month lags of the ACF > 0.20 or < -0.20
### Table 2. Sensitivity Analysis Omitting Place-Outcomes that Exhibit Persistent Autocorrelation: Analysis of the Relationships of Acute Violence Changes with Health Outcomes, Estimated with Linear Regression Analyses of Monthly Residual Outcome Rates with Fixed Effects on Place, California 2005-2013

<table>
<thead>
<tr>
<th>Health Outcome</th>
<th>Residual rate differences associated with violence spikes $^b$</th>
<th>Residual rate differences associated with acute violence increases $^{b,c}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RD $^d$</td>
<td>95% CI</td>
</tr>
<tr>
<td>Anxiety disorders $^a$</td>
<td>Combined fatal and non-fatal</td>
<td>0.30</td>
</tr>
<tr>
<td>Episodic mood disorders $^a$</td>
<td>Combined fatal and non-fatal</td>
<td>0.08</td>
</tr>
<tr>
<td>Substance use</td>
<td>Combined fatal and non-fatal</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td>Fatal only</td>
<td>0.00</td>
</tr>
<tr>
<td>Asthma</td>
<td>Combined fatal and non-fatal</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>Fatal only</td>
<td>0.00</td>
</tr>
<tr>
<td>Chronic obstructive pulmonary disease</td>
<td>Combined fatal and non-fatal</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>Fatal only</td>
<td>0.00</td>
</tr>
<tr>
<td>Acute myocardial infarction</td>
<td>Combined fatal and non-fatal</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>Fatal only</td>
<td>0.09</td>
</tr>
<tr>
<td>Heart failure</td>
<td>Combined fatal and non-fatal</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>Fatal only</td>
<td>-0.02</td>
</tr>
</tbody>
</table>

RD: Rate Difference per 100,000; 95%CI: 95% Confidence Interval

$^a$ There were too few cases of these fatal outcomes to examine them separately (fatal anxiety disorder cases: 46, fatal episodic mood disorder cases: 322)

$^b$ Residuals based on fitting a Kalman smoother to the outcome rates in each place to remove predictable temporal patterning

$^c$ RD corresponding to 10 per 100,000 increase in acute violence residual

$^d$ Controlled for local monthly average precipitation, average temperature, unemployment, and civil unrest events; controlled for local annual measures of poverty, unemployment, marital status, racial/ethnic composition, and educational attainment
Analyses with place fixed effects assume that past exposures do not affect future outcomes within a place, and that past outcomes do not affect future exposures within a place.\textsuperscript{75} In considering the first part of this assumption, while the underlying health conditions likely developed over long period and may have been affected by prior acute violence, we are examining only acute changes in manifestations of these health conditions which are far less likely to be affected by prior acute violence. We evaluated the second part of this assumption by examining the relations of the health outcomes with subsequent acute violence changes (at one month lag) within a place, treating the health outcomes as the “exposures” or independent variables and community violence rate as the outcome or dependent variable. Analyses were conducted with linear regression models with fixed effects on place. We adjusted the same covariates as the main analyses. As shown in Table 1, the associations of each health measure with the rate of violence in the subsequent month were largely null. These findings suggest that reliance on this assumption is reasonable in our study.

Table 1. Analysis of the Relationships of Health Outcomes with Subsequent Acute Violence Changes, Estimated with Linear Regression Analyses of Monthly Residual Outcome Rates with Fixed Effects on Place, California 2005-2013

<table>
<thead>
<tr>
<th>Health Outcome</th>
<th>Residual rate differences in community violence associated with acute health outcome increases (b, c)</th>
<th>RD(^b)</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anxiety disorders</td>
<td>Combined fatal and non-fatal(^a)</td>
<td>0.01</td>
<td>0.00, 0.02</td>
</tr>
<tr>
<td>Episodic mood disorders</td>
<td>Combined fatal and non-fatal(^a)</td>
<td>0.00</td>
<td>-0.01, 0.01</td>
</tr>
<tr>
<td>Substance use</td>
<td>Combined fatal and non-fatal</td>
<td>0.00</td>
<td>-0.01, 0.01</td>
</tr>
<tr>
<td></td>
<td>Fatal only</td>
<td>0.08</td>
<td>-0.04, 0.20</td>
</tr>
<tr>
<td>Asthma</td>
<td>Combined fatal and non-fatal</td>
<td>0.00</td>
<td>-0.01, 0.00</td>
</tr>
<tr>
<td></td>
<td>Fatal only</td>
<td>-0.08</td>
<td>-0.27, 0.12</td>
</tr>
<tr>
<td>Chronic obstructive pulmonary disease</td>
<td>Combined fatal and non-fatal</td>
<td>0.00</td>
<td>-0.01, 0.01</td>
</tr>
<tr>
<td></td>
<td>Fatal only</td>
<td>0.01</td>
<td>-0.02, 0.04</td>
</tr>
<tr>
<td>Acute myocardial infarction</td>
<td>Combined fatal and non-fatal</td>
<td>-0.01</td>
<td>-0.02, 0.00</td>
</tr>
<tr>
<td></td>
<td>Fatal only</td>
<td>-0.04</td>
<td>-0.07, -0.01</td>
</tr>
<tr>
<td>Heart failure</td>
<td>Combined fatal and non-fatal</td>
<td>0.00</td>
<td>-0.01, 0.01</td>
</tr>
<tr>
<td></td>
<td>Fatal only</td>
<td>0.04</td>
<td>-0.01, 0.09</td>
</tr>
</tbody>
</table>

RD: Rate Difference per 100,000
\(a\) There were too few cases of these fatal events to examine them separately (fatal anxiety disorder cases: 46, fatal episodic mood disorder cases: 322)
b Residuals based on fitting a Kalman smoother to the rates in each place to remove predictable temporal patterning

c RD corresponding to 1 per 100,000 increase in each health measure “exposure”

d Controlled for local monthly average precipitation, average temperature, unemployment, and civil unrest events; controlled for local annual measures of poverty, unemployment, marital status, racial/ethnic composition, and educational attainment
Analysis code in R

```r
# Load packages and define functions used throughout
library(forecast)
library(KFAS)
library(reshape2)
library(foreign)
library(DataCombine)
library(MASS)
library(doBy)
library(zoo)
library(sandwich)

### Define Kalman Filter Smoothing function

kfs.out <- function(ts) {
  fit <- auto.arima(ts,allowdrift=T,allowmean=T)
  coef <- fit$coef
  # get coefficients
  try(rm(ar))
  if (is.na(coef["ar1"])) ar <- NULL
  if (!is.na(coef["ar1"])) ar <- coef["ar1"]
  if (!is.na(coef["ar2"])) ar <- c(ar,coef["ar2"])
  if (!is.na(coef["ar3"])) ar <- c(ar,coef["ar3"])
  if (!is.na(coef["ar4"])) ar <- c(ar,coef["ar4"])
  if (!is.na(coef["ar5"])) ar <- c(ar,coef["ar5"])
  try(rm(ma))
  if (is.na(coef["ma1"])) ma <- NULL
  if (!is.na(coef["ma1"])) ma <- coef["ma1"]
  if (!is.na(coef["ma2"])) ma <- c(ma,coef["ma2"])
  if (!is.na(coef["ma3"])) ma <- c(ma,coef["ma3"])
  if (!is.na(coef["ma4"])) ma <- c(ma,coef["ma4"])
  if (!is.na(coef["ma5"])) ma <- c(ma,coef["ma5"])
  diff <- fit$arma[length(fit$arma)-1]
  # fit Kalman smoother
  model <- SSModel(ts ~ SSMarima(ar=ar,ma=ma,d=diff))
  fit <- fitSSM(model,init=c(0,0))
  out.kfs <- KFS(fit$model)
  # get residuals
  resid <- ts - out.kfs$muhat
  # locate shocks identified by procedure
  idsh <- as.numeric(abs(resid)>2*sd(resid))
}
```
# save and return
results <- list(resid, idsh)
return(results)
}

# Created Kalman Filtered Violence Residuals
rm(list=setdiff(ls(), "kfs.out")) # Clear the workspace except for the Kalman Filter function

### Read in data

data <- read.dta("E:/Data/analysis files/1 - Violence/1 - Overall Violence/2 - Violence Rates/Combined OSHPD VS/Place/monthly_ipv_sdv_total_allyrs.dta")

# Check unique identifiers
nrow(data)
nrow(unique(data[,c('year','place')]))

# Keep only vars of interest
data <- data[,c('year','place',paste0('r_ipv_',1:12),paste0('ppop_',1:12))]

# Make sure variables are numeric
for (i in names(data)[names(data)!='place']) {
data[,i] <- as.numeric(data[,i])
}

# Restrict to Places with at least 5000 residents for all years
length(unique(data$place)) # 1516 places
for (i in sort(unique(data$place))) {
  residents <- data[data$place==i,c(paste0("ppop_",1:12))]
  above5000 <- (residents >= 5000)
  if (sum(as.numeric(above5000))<9*12) data <- data[data$place!=i,]
}
length(unique(data$place)) # 631 places remain
data <- data[,!names(data) %in% paste0('ppop_',1:12)]

# Format data long
data <- melt(data, id.vars = c('year','place'))
data$month <- as.numeric(substr(data$variable,7,8))
data$<t> <- (as.numeric(data$year)-2005)*12 + data$month
data <- data[order(data$place,data$year,data$month),]

#### Fit Kalman and identify violence spikes ('shocks') and residuals

data$residual <- data$shock <- NA
for (i in sort(unique(data$place))) {
  print(paste('Place: ',i))
  v <- as.ts(data$value[data$place==i])
  res <- kfs.out(v)
data$residual[data$place==i] <- res[[1]]
data$shock [data$place==i] <- res[[2]]
}

#### Save

final <- data
names(final) <- c('year','place','month','violence_rate','t','shock','residual')
write.csv(final, "Kalman_residuals.csv", row.names = F)

########################################################################
# Created Combined Exposure and Outcome Dataset
########################################################################
rm(list=setdiff(ls(), "kfs.out")) # Clear the workspace except for the Kalman Filter function

### Pull and clean outcome counts- nonfatal

setwd("E:/Data/analysis files/7 - Violence Shocks Outcomes/1 - Counts/OSHPD/Place/")
data <- NULL
for (y in 2005:2013) {
    print(y)
    temp <- read.csv(paste0('oshpd_place_violshock_cnts_',y,'.csv'))
    temp <- temp[!is.na(temp$place),]
data <- rbind(data, temp)
}
data <- data[,!names(data) %in% c('n_yr_nf_asthma', 'n_yr_nf_acute_MI',
    'n_yr_nf_heart_fail', 'n_yr_nf_subst_use',
    'n_yr_nf_episod_md_dis','n_yr_nf_anxi_strss_dis',
    'n_yr_nf_copd', 'all',
    'county')]
for (i in 1:12) {
    data[[paste0('n_nf_all_',i)]] <- rowSums(data[,paste0(c('n_nf_asthma_','n_nf_acute_MI_','n_nf_heart_fail_','n_nf_subst_use_','n_nf_episod_md_dis_','n_nf_anxi_strss_dis_','n_yr_nf_copd_'),rep(i,7))])
}
data <- melt(data, id.vars=c('year','place'))
data$month <- gsub('[^0-9]', '', data$variable)
data$variable <- gsub('\d', '', data$variable)
data$month <- as.numeric(data$month)
data$variable <- gsub('_nf_','_', data$variable)
names(data) <- c('year','place','condition','cases_nf','month')

### Pull and clean outcome counts- fatal

setwd("E:/Data/analysis files/7 - Violence Shocks Outcomes/1 - Counts/VS/Place/")
dat2 <- NULL
for (y in 2005:2013) {
    print(y)
    temp <- read.csv(paste0("dcf_place_violshock_cnts_","y",".csv"))
    temp <- temp[!is.na(temp$place),]
dat2 <- rbind(dat2, temp)
}
dat2 <- dat2[,!names(dat2) %in% c('n_yr_f_asthma', 'n_yr_f_acute_MI',
    'n_yr_f_heart_fail', 'n_yr_f_subst_use',
    'n_yr_f_episod_md_dis','n_yr_f_anxi_strss_dis','n_yr_f_copd',
    'all','county')]
for (i in 1:12) {
    dat2[[paste0('n_f_all_',i)]] <- rowSums(dat2[,paste0(c('n_f_asthma_','n_f_acute_MI_','n_f_heart_fail_','n_f_subst_use_','n_f_episod_md_dis_','n_f_anxi_strss_dis_','n_yr_f_copd_'),rep(i,7))])
}
dat2 <- melt(dat2, id.vars=c('year','place'))
dat2$month <- gsub('[^0-9]', '', dat2$variable)
dat2$variable <- gsub('\d', '', dat2$variable)
dat2$month <- as.numeric(dat2$month)
dat2$variable <- gsub("_f","_",dat2$variable, fixed=T)
names(dat2) <- c("year","place","condition","cases_f","month")

### Pull and clean denominators

setwd("E:/Data/geocoding and denominators/denominators/5 - Denominators/2 - Data files/Place/")
denom <- read.dta("denom_place_all.dta")
denom <- denom[denom$agegrp=="Total" & denom$race=="Total" & denom$sex=="Total",]
denom <- denom[,names(denom) %in% c('agegrp','race','sex',paste0('ppop0',5:9),paste0('ppop',10:14))]
denom$place <- as.numeric(denom$place)
denom$month <- as.numeric(substr(gsub("[^_0-9]","", denom$variable),4,5))
denom$year <- as.numeric(substr(gsub("[^_0-9]","", denom$variable),1,2))+2000
denom$variable <- gsub("\"d","", denom$variable)
denom <- denom[,names(denom)!="variable"]
names(denom) <- c("place","placename","pop","month","year")

# Make a data frame that is square on place, month, year AND condition, so we don't end up with spotty
# missingness later
square <- expand.grid(place=unique(denom$place), year=2005:2013, month=1:12,
condition=unique(data$condition))
square$condition <- as.character(square$condition)
denom <- merge(square, denom, by=c('place','year','month'), all=T)
head(denom)
summary(denom)

### Merge outcomes and denominators

# Merge non-fatal to denominators
dim(data)
dim(denom)
dat3 <- merge(data, denom, by=c('place','year','month','condition'), all=T)
dim(dat3)

summary(dat3)

# Places not in the count data are missing because there weren't any cases.
# Fill these in with zeros:
dat3$cases_nf[is.na(dat3$cases_nf)] <- 0
summary(dat3)

# Merge fatal to denominators
dim(dat2)
dim(denom)
dat4 <- merge(dat2, denom, by=c('place','year','month','condition'), all=T)
dim(dat4)

summary(dat4)
dat4$cases_f[is.na(dat4$cases_f)] <- 0
summary(dat4)

# Merge fatal and non-fatal
dim(dat3); dim(dat4)
dat5 <- merge(dat3, dat4,
  by=c('place','year','month','condition','placename','pop'), all=T)
dim(dat5)
summary(dat5)
table(is.na(dat5$condition))
table(is.na(dat5$placename))

# Make case and rate variables for combined fatal and nonfatal
dat5$cases <- dat5$cases_f + dat5$cases_nf
dat5$rate <- dat5$cases/dat5$pop *100000
dat5$rate_f <- dat5$cases_f/dat5$pop *100000
dat5$rate_nf <- dat5$cases_nf/dat5$pop *100000
summary(dat5)

rm(data, dat2, dat3, dat4, denom, square, temp, y)

# Check that data are square
dim(dat5)
dim(unique(dat5[,c('year','place','month','condition')]))

#### Pull in and merge on violence spikes ('shocks') and residuals
shocks <- read.csv("E:/Data/analysis files/1 - Violence/1 - Overall Violence/3 - Violence Residuals/Place/Kalman_residuals.csv")

# Reassign the negative shocks to non-shocks— we only want to evaluate positive shocks
table(shocks$shock)
shocks$shock[shocks$shock==1 & shocks$residual<0] <- 0
table(shocks$shock)
names(shocks)[names(shocks)="residual"] <- "viol_residual"

# Keep vars of interest
shocks <- shocks[,c("year","month","place","shock","viol_residual")]
dim(dat5)
dim(shocks) # Note that shocks were only computed for the 631 places with pop of at least 5,000
dat6 <- merge(dat5, shocks, by=c("year","month","place"), all.y=T)
dim(dat6)
dim(unique(dat5[,c('year','place','month','condition')]))

summary(dat6)
table(is.na(dat6$condition))
table(is.na(dat6$placename))

rm(dat5, shocks)

length(unique(dat6$place))

# Data now restricted to the 631 places with pop of at least 5,000

#### Apply Kalman Filter to the rate for each outcome and month
dat6$rate_residual <- dat6$rate_kalman.fit <-
dat6$rate_f_residual <- dat6$rate_f_kalman.fit <-
dat6$rate_nf_residual <- dat6$rate_nf_kalman.fit <-NA
for (c in unique(dat6$condition)) {
  for (i in sort(unique(dat6$place))) {
    for (f in c("","_f","_nf")) {
      temp <- dat6[dat6$condition==c & dat6$place==i,]
      v <- as.ts(temp[[paste0('rate',f)])
      kalmanfit <- kfs.out(v)
      dat6[dat6$condition==c & dat6$place==i, paste0('rate',f,'_residual')] <-
      as.numeric(kalmanfit$resid)
dat6[dat6$condition==c & dat6$place==i, paste0('rate',f,'_kalman.fit')] <-
as.numeric(kalmanfit$fitted)

# Checks
head(dat6)
summary(dat6)
dim(dat6)
dim(unique(dat6[,c('year','place','month','condition')]))

# Save data
setwd("E:/Data/analysis files/100 - Papers/5 - Violence shocks")
save(dat6, file="combined_working_data_noNegShocks_detrended_Kalman.Rdata")

########################################################################
# Add Covariates
########################################################################
rm(list=ls()) # Clear workspace

### Load exposure and outcome data
setwd("E:/Data/analysis files/100 - Papers/5 - Violence shocks")
load(file="combined_working_data_noNegShocks_detrended_Kalman.Rdata")

### Prep time-varying covariates
setwd("E:/Data/analysis files/100 - Papers/5 - Violence shocks")

# Crosswalk file
cross <- read.csv("county_to_place_crosswalk.csv")
cross <- cross[nrow(cross),c('county','placefp','afact','pop10')]
for (j in 1:ncol(cross)) cross[,j] <- as.numeric(as.character(cross[,j]))

# Remove counties without corresponding place
dim(cross)
cross <- cross[!is.na(cross$placefp),]
dim(cross)

# Select for each place, the county with the largest allocation factor (approximation)
for (p in unique(cross$placefp)) {
  max <- max(cross$pop0[!is.na(cross$placefp)==p] & cross$placefp==p)
cross$county2[!is.na(cross$placefp)==p] <- cross$county[!is.na(cross$pop0)==max &
cross$placefp==p]
rm(max)
}
cross2 <- unique(cross[,c('placefp','county2')])
dim(cross2)
length(unique(cross$placefp))
names(cross2) <- c('place','county')
rm(cross)

# Covariates - temperature, precipitation, unemployment
avgt <- read.csv("CA-PRISM-county-avgt.txt")
avgt <- melt(avgt, id.vars="YEAR.MO")
avgt$year <- as.numeric(substr(avgt$YEAR.MO,1,4))
avgt$month <- as.numeric(substr(avgt$YEAR.MO,6,7))
avgt$county <- as.numeric(substr(avgt$variable,2,6))
avgt$avgt <- avgt$value
avgt <- avgt[avgt$year>=2004,c('county','year','month','avgt')]

precip <- read.csv("CA-PRISM-county-pcp.txt")
precip <- melt(precip, id.vars="YEAR.MO")
precip$year <- as.numeric(substr(precip$YEAR.MO,1,4))
precip$month <- as.numeric(substr(precip$YEAR.MO,6,7))
precip$county <- as.numeric(substr(precip$variable,2,6))
precip$precip <- precip$value
precip <- precip[precip$year>=2004,c('county','year','month','precip')]

unemp <- read.csv("SeriesReport-20151110124206_b868ae.csv", stringsAsFactors=F)
unemp$county <- as.numeric(substr(unemp$Series.ID,6,10))
unemp <- unemp[unemp$View.Description=="Original Data Value",]
unemp <- melt(unemp, id-vars=c('Series.ID','View.Description'))
unemp$month <- substr(unemp$variable,1,3)
unemp$year <- as.numeric(substr(unemp$variable,5,8))
i <- 0
for (m in unique(unemp$month)) {
  i <- i+1
  unemp$month.num[unemp$month==m] <- i
}
table(unemp$month, unemp$month.num)
unemp$unemp <- unemp$value
unemp <- unemp[unemp$year>=2004 & unemp$year<=2013,c('county','year','month.num','unemp')]
names(unemp)[names(unemp)="/"month.num"] <- "month"
covs <- merge(avgt, precip, by=c('county','year','month'))
dim(avgt); dim(precip); dim(covs)
covs <- merge(covs, unemp, by=c('county','year','month'))
dim(unemp); dim(covs)
covs <- merge(covs, cross2, by=c('county'))
dim(covs)
10*12*1516 # 10 years, 12 months, 1516 places. Data is square.

### Add on yearly covariates and prep data for regression analysis

# Reshape data long
drop <- c('cases','cases_f','cases_nf','rate','rate_f','rate_nf','rateKalman.fit','rateKalman.fit','rate_nfKalman.fit','pop','placename')
long <- melt(dat6[,!names(dat6) %in% drop], id-vars=c('year','month','place','t','condition','shock','viol_residual'))
long$ftype <- 'all'
long$ftype[grep("_f",long$variable)] <- 'fatal'
long$ftype[grep("_nf",long$variable)] <- 'nonfatal'
# Checks
table(long$condition)
table(long$condition, long$variable)
table(table(long$place))
table(long$ftype)
table(long$year, long$month)
summary(long)
dim(long)
dim(unique(long[,c('place','t','condition','ftype')]))

# Merge on monthly covariates
dim(long)
long <- merge(long, covs, by=c('place','year','month'), all.x=T)
dim(long)

# Merge on riots
riots <- read.csv("E:/Data/analysis files/100 - Papers/5 - Violence shocks/shocks.data.frame.riots.csv")
riots <- riots[,names(riots)!="placename"]
dim(long)
long <- merge(long, riots, by=c("place","year","month"), all.x=T)
dim(long)
summary(long)

# ACS covariates
acs <- read.csv("E:/Data/analysis files/100 - Papers/5 - Violence shocks/place_covariates05_13_v2.csv")
acs$geoid <- acs$geoid - 600000
acs <- acs[acs$geoid %in% unique(long$place),] # restrict to relevant places
length(unique(acs$geoid))*length(unique(acs$year))
dim(acs) # data is not square

# Make ACS data square
square <- expand.grid(geoid = unique(long$place), year=2005:2013)
dim(square)
acs <- merge(acs, square, by=c("geoid","year"), all=T)
dim(acs)

select <-
c("hhfoodstamps","fam_noworkers","pct_colgrad25pl","pctnevermarried","r_whitenh","r_blacknh","r_hispanic")
acs <- acs[,c("geoid","year",select)]
dim(acs)
dim(unique(acs[,c('geoid','year')]))
names(acs)[names(acs)=="geoid"] <- "place"

summary(acs) # 4.2% missing due to changes in Place definitions / boundaries in 2010

# Carry missing data forward and backward
acs <- acs[order(acs$place, acs$year),]
for (p in unique(acs$place)) {
  print(p)
  for (cov in select) {
    acs[acs$place==p, cov] <- na.locf(acs[acs$place==p, cov], na.rm=F)
    acs[acs$place==p, cov] <- na.locf(acs[acs$place==p, cov], na.rm=F, fromLast=T)
  }
}

summary(acs)
dim(acs)
dim(unique(acs[,c('place','year')]))

dim(long)
long <- merge(long, acs, by=c("place","year"), all.x=T)
dim(long)

# Checks
head(long)
table(long$condition)
table(long$condition, long$variable)
table(long$place)
table(long$ftype)
table(long$year, long$month)
table(long$condition, long$shock)
summary(long)
table(covs$place %in% long$place)
table(long$place %in% covs$place)

rm(dat6, covs, avgt, cross2, i, j, m, p, precip, unemp, drop, acs, cov, select, square, riots)
setwd("E:/Data/analysis files/100 - Papers/5 - Violence shocks")
write.csv(long, 'panel_analysis_detrended_analysis_data.csv',row.names=F)

############################
# Regression Analysis
############################

rm(list=ls())
contExp <- T # Toggle T/F for continuous exposure (vs binary)?
setwd("E:/Data/analysis files/100 - Papers/5 - Violence shocks")
long <- read.csv('panel_analysis_detrended_analysis_data.csv', stringsAsFactors=F)
results <- NULL
for (condit in c('n_acute_MI_','n_asthma_','n_copd_','n_anxi_strss_dis_','n_episod_md_dis_','n_hear_t_fail_','n_subst_use_')) {
  for (ftype in c('all','fatal')) {
    print(paste0("condition: ",condition," , ftype: ",ftype))
    if ((condition=='n_episod_md_dis_' | condition== "n_anxi_strss_dis_") & ftype=='fatal') next
    temp <- long[long$ftype==ftype & long$condition==condition,]
    # Rescale continuous covariates to avoid convergence issues
    for (n in c('avgt','precip','unemp','hhfoodstamps','fam_noworkers','pct_colgrad25pl','pctnever_married','r_whitenh','r_blacknh','r_hispanic')) {
      temp[[n]] <- (temp[[n]] - mean(temp[[n]])) / sd(temp[[n]])
    }
    # Model: Normally distributed rate residual outcome, identity link, place as fixed effect, with covariates
    if (!contExp) formula <- as.formula('value ~ factor(shock) +avgt +precip +unemp +riot +hhfoodstamps +fam_noworkers +pct_colgrad25pl +pctnevermarried +r_whitenh +r_blacknh +r_hispanic +factor(place)')
    if ( contExp) formula <- as.formula('value ~ viol_residual +avgt +precip +unemp +riot +hhfoodstamps +fam_noworkers +pct_colgrad25pl +pctnevermarried +r_whitenh +r_blacknh +r_hispanic +factor(place)')
    try(reg <- glm(formula, data=temp))
    if (exists("reg")) {
      se <- sqrt(diag(vcovHC(reg, type = "HC"))) # sandwich estimator for SE's
      tab <- (cbind(Est = coef(reg), LL = coef(reg)-1.96*se, UL = coef(reg)+1.96*se))
      results_temp <- data.frame(condition=condition, ftype=ftype)
      results_temp[,c('beta','beta_lower','beta_upper')] <- c(tab[2,])
      results <- rbind(results, results_temp)
      rm(reg, se, tab, results_temp)
    }
  }
}
print(results)
write.csv(results, paste0('regression_results',ifelse(contExp,"_contExp",""),'.csv'), row.names=F)

############################
# Sensitivity analysis

```
rm(list=ls())

setwd("E:/Data/analysis files/100 - Papers/5 - Violence shocks")
long <- read.csv('panel_analysis_detrended_analysis_data.csv', stringsAsFactors=F)

# Restrict to outcomes of interest
long <- long[long$ftype!='nonfatal' & long$condition!='n_all_' &
!(long$condition=='n_anxi_strss_dis_' & long$ftype=='fatal') &
!(long$condition=='n_episod_md_dis_' & long$ftype=='fatal'),]

### Restrict to places that do not have persistent autocorrelation
# Define persistent autocorrelation as those with 6 or more of 12 lags > 0.2 or <- -0.2
# Bring in and merge on ACF stats
# These are the autocorrelation at 1-21 month lags for adjusted models WITHOUT the exposure in the model.
# (want to determine how much autocorrelation the outcomes have before we consider its relation to exposure.)
acf <- read.csv('acf_pacf_outcomes_modeled.csv', stringsAsFactors=F)

# Restrict to models and outcomes of interest and ACF stats for lags 2-13 (i.e. lags of 1-12 months)
acf <- acf[acf$timeFE==0 & acf$lag>=2 & acf$lag<=13 &
!(acf$outcome=='n_anxi_stress_dis_' | acf$outcome=='n_episod_md_dis_') &
acf$fatality=='fatal',
c('outcome','fatality','place','lag','acf')]

# confirm unique identifiers
nrow(acf)
acf <- unique(acf[,c('outcome','fatality','place','lag')])

# identify high autocorrelation
acf$acf.high <- as.numeric(abs(acf$acf) > 0.2)

# identify persistently high autocorrelation
acf <- summaryBy(acf.high ~ outcome + fatality + place, data=acf, FUN=sum)
acf$acf.high.persistent <- as.numeric(acf$acf.high.sum >=6)

summary(acf) # 2.9% of outcome-places with persistently high autocorrelation

# get names of census places
name <- read.dta("E:/Data/geocoding and denominators/denominators/5 - Denominators/2 - Data files/Place/denom_place_all.dta")
name <- unique(name[,c('place','placename')])
nname$place <- as.numeric(name$place)
nrow(acf)
acf <- merge(acf, name, by="place", all.x=T)
nrow(acf)

# Which outcomes and places have persistently high autocorr
table(acf$outcome[acf$acf.high.persistent==1],
acf$fatality[acf$acf.high.persistent==1])
table(acf$placename[acf$acf.high.persistent==1])
table(acf$placename[acf$acf.high.persistent==1][table(acf$placename[acf$acf.high.persistent==1])>1])

# merge onto analysis data to drop
names(acf)[names(acf)=='outcome'] <- "condition"
names(acf)[names(acf)=='fatality'] <- "ftype"

nrow(acf)
acf <- unique(acf[,c('condition','ftype','place')])
```

nrow(long)
long <- merge(long, acf, by=c('condition','ftype','place'), all.x=T)
nrow(long)

summary(long)

nrow(long)
long <- long[long$acf.high.persistent!=1,]
nrow(long)

### Run models
results <- NULL

for (condition in 
c('n_acute_MI_','n_asthma_','n_copd_','n_anxi_strss_dis_','n_episod_md_dis_','n_hear t_fail_',
   'n_subst_use_') ) {
  for (ftype in c('all','fatal')) {
    print(paste0("condition: ",condition,"", ftype: ",ftype))

    if ((condition=='n_episod_md_dis_' | condition== "n_anxi_strss_dis_") &
        ftype=='fatal') next

    temp <- long[long$ftype==ftype & long$condition==condition,]

    # Rescale continuous covariates to avoid convergence issues
    for (n in 
c('avgt','precip','unemp','hhfoodstamps','fam_noworkers','pct_colgrad25pl','pctnever married',
        'r_whitenh', 'r_blacknh', 'r_hispanic')) {
      temp[[n]] <- (temp[[n]] - mean(temp[[n]])) / sd(temp[[n]])
    }

    # Model: Normally distributed rate residual outcome, identity link, place as
    # fixed effect, with covariates
    if (!contExp) formula <- as.formula('value ~ factor(shock) +avgt +precip +unemp +riot +hhfoodstamps +fam_noworkers +pct_colgrad25pl +pctnevermarried +r_whitenh +r_blacknh +r_hispanic +factor(place)')
    if ( contExp) formula <- as.formula('value ~ viol_residual +avgt +precip +unemp +riot +hhfoodstamps +fam_noworkers +pct_colgrad25pl +pctnevermarried +r_whitenh +r_blacknh +r_hispanic +factor(place)')

    try(reg <- glm(formula, data=temp))
    if (exists("reg")) {
      se <- sqrt(diag(vcovHC(reg, type = "HC"))) # sandwich estimator for SE's
      tab <- (cbind(Est = coef(reg), LL = coef(reg)-1.96*se, UL =
                   coef(reg)+1.96*se))
      results_temp <- data.frame(condition=condition, ftype=ftype)
      results_temp[,c('beta','beta_lower','beta_upper')] <- c(tab[2,])
      results <- rbind(results, results_temp)
      rm(reg, se, tab, results_temp)
    }

    print(results)
    write.csv(results, paste0('regression_results',ifelse(contExp, "_contExp",""), '_sens.csv'), row.names=F)
  }

# END