

# Safer at home? Domestic Violence During the COVID-19 Pandemic

Samuel, Natalie, & Stacy

## INTRODUCTION AND DATA

### Safer at home? Domestic Violence During the COVID-19 Pandemic

In what has been dubbed “a pandemic within a pandemic,” a new public health crisis has emerged in the wake of the coronavirus - increasing rates of domestic violence, affecting those in abusive relationships around the globe. This trend is not necessarily surprising, since rates of domestic violence tend to increase whenever families spend more time together - such as during Christmas and summer vacations - according to Bristol University sociologist Marianne Hester, but still troubling nonetheless, especially in countries (like the United States) where the virus doesn’t seem to be disappearing anytime soon. Without a strong support network due to social distancing and lockdown restrictions, it is much more difficult for victims of domestic violence to get help or escape (quarantined in their home), while the situation also enables the abuser by giving them greater power and control over what their partner can and can not do - suggesting that perhaps not everyone is truly “safer at home.”

In order to further investigate the rate of domestic violence during COVID-19, we decided to pulled existing data from the City of Chicago Data Portal, which contains data related to reported incidents of crime (with the exception of murders where data exists for each victim) that occurred in the City of Chicago from 2001 to present. The data contained in this dataset were extracted from the Chicago Police Department’s CLEAR (Citizen Law Enforcement Analysis and Reporting) system.

-introduce your general research question and your data (where it came from, how it was collected, what are the cases, what are the variables, etc.). -Define what we mean by domestic violence

## METHODOLOGY

-variables used to address your research question -useful visualizations or summary statistics -introduce and justify the statistical method(s) that you believe will be useful in answering your research question.

```
## Rows: 231,002
## Columns: 23
## $ CASE. <chr> "JD163753", "JD212847", "JC497784", "JC459410...
## $ DATE..OF.OCCURRENCE <chr> "02/24/2020 08:15:00 PM", "04/10/2020 10:56:0...
## $ BLOCK <chr> "031XX W LEXINGTON ST", "005XX W 103RD ST", "...
## $ IUCR <chr> "1153", "0560", "0860", "0560", "0810", "0820...
## $ PRIMARY.DESCRPTION <chr> "DECEPTIVE PRACTICE", "ASSAULT", "THEFT", "AS...
## $ SECONDARY.DESCRPTION <chr> "FINANCIAL IDENTITY THEFT OVER $ 300", "SIMPL...
## $ LOCATION.DESCRPTION <chr> "", "RESIDENCE", "DEPARTMENT STORE", "SIDEWAL...
## $ ARREST <chr> "N", "N", "N", "N", "N", "N", "N", "N", "N", "N", ...
## $ DOMESTIC <chr> "N", "N", "N", "N", "N", "N", "N", "N", "N", "N", ...
## $ BEAT <int> 1134, 2232, 1924, 122, 123, 2433, 312, 914, 3...
## $ WARD <int> 24, 9, 44, 4, 25, 48, 20, 11, 5, 26, 27, 37, ...
## $ FBI.CD <chr> "11", "08A", "06", "08A", "06", "06", "08A", ...
## $ X.COORDINATE <int> NA, 1174583, NA, NA, NA, NA, 1180030, 1171590...
## $ Y.COORDINATE <int> NA, 1836593, NA, NA, NA, NA, 1862317, 1887793...
```

```

## $ LATITUDE      <dbl> NA, 41.70700, NA, NA, NA, NA, 41.77747, 41.84...
## $ LONGITUDE     <dbl> NA, -87.63629, NA, NA, NA, NA, -87.61556, -87...
## $ LOCATION      <chr> "", "(41.707000821, -87.636288063)", "", "", ...
## $ MONTH         <int> 2, 4, 11, 10, 5, 12, 5, 5, 4, 5, 4, 5, 5, ...
## $ DAY           <int> 24, 10, 3, 4, 24, 5, 7, 3, 28, 7, 25, 7, 7, 5...
## $ YEAR          <int> 2020, 2020, 2019, 2019, 2020, 2019, 2020, 202...
## $ DATEINT       <int> 202002, 202004, 201911, 201910, 202005, 20191...
## $ isPM          <dbl> 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 1, 0, 0, 0, 1, ...
## $ HOUR          <int> 8, 10, 11, 6, 9, 2, 12, 2, 6, 5, 3, 8, 6, 12,...

## # A tibble: 430 x 2
##   SECONDARY.DESCRPTION      n
##   <chr>                    <int>
## 1 $500 AND UNDER          20888
## 2 ABUSE / NEGLECT - CARE FACILITY      5
## 3 ABUSE/NEGLECT: CARE FACILITY      10
## 4 AGG CRIM SEX ABUSE FAM MEMBER      78
## 5 AGG CRIMINAL SEXUAL ABUSE        63
## 6 AGG PO HANDS ETC SERIOUS INJ       8
## 7 AGG PO HANDS NO/MIN INJURY      593
## 8 AGG PRO EMP HANDS SERIOUS INJ      16
## 9 AGG PRO.EMP: HANDGUN              26
## 10 AGG PRO.EMP: OTHER DANG WEAPON     85
## # ... with 420 more rows

## Rows: 231,002
## Columns: 31
## $ CASE.          <chr> "JD163753", "JD212847", "JC497784", "JC459410...
## $ DATE..OF.OCCURRENCE <chr> "02/24/2020 08:15:00 PM", "04/10/2020 10:56:0...
## $ BLOCK          <chr> "031XX W LEXINGTON ST", "005XX W 103RD ST", "...
## $ IUCR           <chr> "1153", "0560", "0860", "0560", "0810", "0820...
## $ PRIMARY.DESCRPTION <chr> "DECEPTIVE PRACTICE", "ASSAULT", "THEFT", "AS...
## $ SECONDARY.DESCRPTION <chr> "FINANCIAL IDENTITY THEFT OVER $ 300", "SIMPL...
## $ LOCATION.DESCRPTION <chr> "", "RESIDENCE", "DEPARTMENT STORE", "SIDEWAL...
## $ ARREST         <chr> "N", "N", "N", "N", "N", "N", "N", "N", "N", ...
## $ DOMESTIC       <chr> "N", "N", "N", "N", "N", "N", "N", "N", "N", ...
## $ BEAT           <int> 1134, 2232, 1924, 122, 123, 2433, 312, 914, 3...
## $ WARD           <int> 24, 9, 44, 4, 25, 48, 20, 11, 5, 26, 27, 37, ...
## $ FBI.CD         <chr> "11", "08A", "06", "08A", "06", "06", "08A", ...
## $ X.COORDINATE    <int> NA, 1174583, NA, NA, NA, NA, 1180030, 1171590...
## $ Y.COORDINATE    <int> NA, 1836593, NA, NA, NA, NA, 1862317, 1887793...
## $ LATITUDE       <dbl> NA, 41.70700, NA, NA, NA, NA, 41.77747, 41.84...
## $ LONGITUDE      <dbl> NA, -87.63629, NA, NA, NA, NA, -87.61556, -87...
## $ LOCATION       <chr> "", "(41.707000821, -87.636288063)", "", "", ...
## $ MONTH          <int> 2, 4, 11, 10, 5, 12, 5, 5, 4, 5, 4, 5, 5, ...
## $ DAY            <int> 24, 10, 3, 4, 24, 5, 7, 3, 28, 7, 25, 7, 7, 5...
## $ YEAR           <int> 2020, 2020, 2019, 2019, 2020, 2019, 2020, 202...
## $ DATEINT        <int> 202002, 202004, 201911, 201910, 202005, 20191...
## $ isPM           <dbl> 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 1, 0, 0, 0, 1, ...
## $ HOUR           <int> 8, 10, 11, 6, 9, 2, 12, 2, 6, 5, 3, 8, 6, 12,...
## $ isbeforecovid   <dbl> 0, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
## $ isprelockdown   <dbl> 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
## $ islockdown      <dbl> 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1, ...
## $ isphase2        <dbl> 0, 0, 0, 0, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1, 0, ...
## $ isphase3        <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...

```

```

## $ isphase4          <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
## $ isdomviolence     <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, ...
## $ is9_5             <dbl> 0, 0, 1, 0, 0, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, ...

##   isbeforecovid      n
## 1              0 90305
## 2              1 140697

##   isprelockdown      n
## 1              0 196383
## 2              1  34619

##   islockdown         n
## 1              0 214160
## 2              1  16842

##   isphase2           n
## 1              0 212424
## 2              1  18578

##   isphase3           n
## 1              0 218280
## 2              1  12722

##   isphase4           n
## 1              0 223458
## 2              1   7544

##   isdomviolence      n
## 1              0 206105
## 2              1  24897

## [1] isdomviolence n
## <0 rows> (or 0-length row.names)

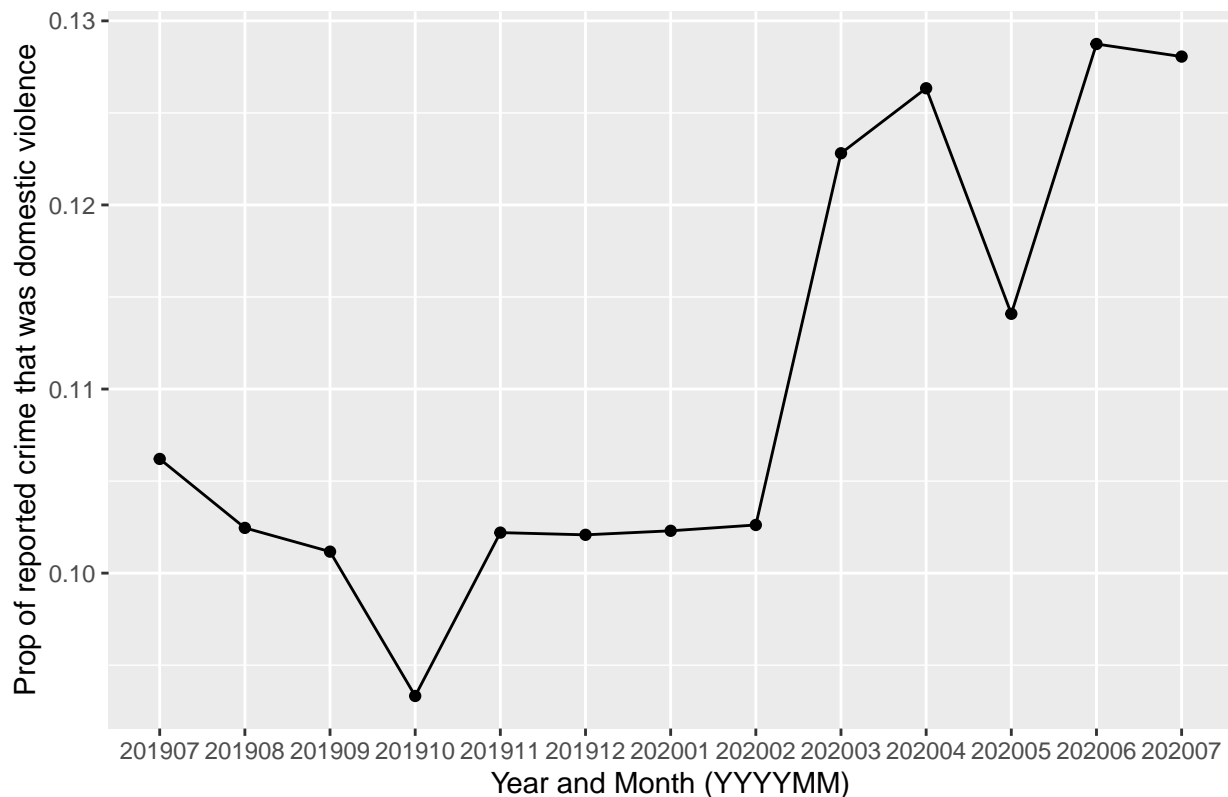
##   isdomviolence      n
## 1              0 15502
## 2              1  1842

```

## RESULTS

-Showcase how -Provide the main results from your analysis

Monthly domestic violence proportion seem to have shot up in March, whe



### Chi-Square Test

```
##
##      lockdown    post    pre
## 0      31109  17599 157397
## 1       4311   2667  17919
##
## Pearson's Chi-squared test
##
## data:  table(domvio_mut$isdomviolence, domvio_mut$chisquare_indicators)
## X-squared = 247.63, df = 2, p-value < 2.2e-16
```

Since the data (Table 1) satisfies the independent sampling assumption and is large enough (i.e. each cell > 10), we will be performing a chi-square test at the  $\alpha = 0.05$  significance level. We test the two hypotheses below:

$H_0$  : The frequency of domestic violence cases in Chicago is unrelated to the phases of the pandemic.  $H_1$  : The frequency of domestic violence cases in Chicago is related to the phases of the pandemic.

Under the null hypothesis, our test statistic has a chi-square distribution with 2 degrees of freedom. We performed the test and obtained a chi-square value of 247.63, which corresponds to a p-value of  $< 0.001$ . Thus, at an  $\alpha = 0.05$  significance level, we reject the null hypothesis; there is sufficient evidence to suggest that the frequency of domestic violence cases in Chicago is related to the phases of the pandemic.

**Step Down 2 Proportion Z-Tests** Since the overall Chi-square test was significant, we stepped down to identify where the differences are. We conducted three 2 proportion z-tests. To account for multiple

comparisons, we will perform the Bonferroni correction and thus assess our results relative to the adjusted  $\alpha = 0.05/3$  level.

```
##
## 2-sample test for equality of proportions with continuity correction
##
## data:  c(4311, 2667) out of c(35420, 20266)
## X-squared = 11.411, df = 1, p-value = 0.0007303
## alternative hypothesis: two.sided
## 95 percent confidence interval:
## -0.01569439 -0.00408326
## sample estimates:
##      prop 1      prop 2
## 0.1217109 0.1315997

##
## 2-sample test for equality of proportions with continuity correction
##
## data:  c(2667, 17919) out of c(2667 + 17599, 17919 + 157397)
## X-squared = 166.3, df = 1, p-value < 2.2e-16
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.024497 0.034283
## sample estimates:
##      prop 1      prop 2
## 0.1315997 0.1022097

##
## 2-sample test for equality of proportions with continuity correction
##
## data:  c(4311, 17919) out of c(35420, 17919 + 157397)
## X-squared = 118.55, df = 1, p-value < 2.2e-16
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.01579582 0.02320653
## sample estimates:
##      prop 1      prop 2
## 0.1217109 0.1022097

## [1] 0.01666667
```

We found that all three pairwise difference in proportions are significant at the adjusted significance level.

**Regression Analysis** In order to model the relationship between the probability that a given case reported to the police was an incidence of domestic violence (considered to be a “success” in the context of our analysis), and the phase of the pandemic in which it was reported, we created a logistic regression model. We chose to use a logistic regression model because we had a binary “success” outcome, where the variable “isdomvio” was either a 0 or 1, with 1 being classified as a case of domestic violence, and 0 otherwise - a scenario in which it is inappropriate to use multiple linear regression, due to the challenges of modeling probabilities with this method. In our model, we included a series of dummy variables under the category “pandemic\_time” in reference to the specific phase of lockdown that the crime was reported in - either before the COVID-19 pandemic (before\_covid), prior to when the first official stay at home order was issued (pre-lockdown), during the strict stay at home order itself (lockdown), or during either phase 2, phase 3, or phase 4 of reopening (phase 2, phase 3, and phase 4, respectively). The exact dates we used to define each of these time periods are provided in the introduction, for reference. The variable, “before\_covid” was used as the baseline category for the series of pandemic\_time dummy variable, to which the difference of the effect of a given estimated

parameter on the predicted outcome of success was given in reference to. In our model, we also adjusted for the time of day in which the crime was reported to the police, using a binary indicator “is9\_5” where a 1 indicates that the crime was reported during the 9-5 workday, 0 otherwise. We chose to include this predictor in our model and adjust for its effect with the effect of the COVID-19 pandemic on the workforce and normal working hours in mind. We hypothesized that before the pandemic, one would be more likely to see crimes of domestic violence reported to the police outside of the 9 to 5 working day, given that individuals would most likely be at work, less at risk of being subjected to violent encounters with an abusive partner, rather than at home (outside of the 9 to 5 time-frame). However, during the pandemic, as more individuals were forced to work from home and the concept of a traditional “9 to 5” work day disappeared for most, we hypothesized that we would see a difference in the number of cases of domestic violence reported during these two time periods, and more specifically, that we would see more cases reported during lockdown during the 9 to 5 workday than before.

Adjusting for time, we hypothesized that the probability of success that a case reported to the police would be related to domestic violence would increase during the initial lockdown phase, and begin to recede slightly as the state of Chicago and the city of Illinois began to gradually reopen. We based this hypothesis on our understanding of the nature of domestic violence (i.e. that it often occurs more often in the home, as the time individuals spend with abusive partners increases), and using evidence from the spike in the proportion of cases of domestic violence as the stay at home order went into effect that we previously confirmed graphically.

After running our code, we also determined the estimated odds ratio for each beta coefficient(given by  $\exp(\beta)$ ) in our model and calculated the associated confidence interval for each of these odds ratios (with the appropriate confidence multiplier using a standard normal distribution and the standard error from our model output) for use in further analysis. In this context,  $\exp(\beta)$  corresponds to the multiplicative effect of a 1 unit change (i.e. whether a dummy condition is satisfied or not) in a given predictor on the odds of a crime reported to the police being related to domestic violence, while holding all other confounders constant. Additionally, when we report confidence intervals and say we are 95% confident, what we mean is that if we were to take new samples independently from the original population and repeatedly construct logistic regression models and confidence intervals for  $\exp(\beta)$  in the same way, we would expect 95% of such intervals to cover the true parameter.

```
## # A tibble: 7 x 5
##   term                estimate std.error statistic  p.value
##   <chr>                <dbl>    <dbl>    <dbl>    <dbl>
## 1 (Intercept)         -2.07      0.0101   -205.    0.
## 2 pandemic_timelockdown  0.303     0.0244    12.4  1.87e- 35
## 3 pandemic_timephase 2  0.105     0.0249     4.23  2.37e- 5
## 4 pandemic_timephase 3  0.301     0.0276    10.9  8.43e- 28
## 5 pandemic_timephase 4  0.262     0.0355     7.37  1.77e- 13
## 6 pandemic_timepre-lockdown 0.0528    0.0196     2.69  7.11e- 3
## 7 as.factor(is9_5)1    -0.314     0.0142   -22.2  1.00e-108
```

Our logistic regression model corresponding to the probability of success that a reported crime for a given time period will be one of domestic violence is as follows:

$$\hat{\beta}_0 + \hat{\beta}_1 * (\text{pandemic\_time} == \text{lockdown}) + \hat{\beta}_2 * (\text{pandemic\_time} == \text{phase 2}) + \hat{\beta}_3 * (\text{pandemic\_time} == \text{phase 3}) + \hat{\beta}_4 * (\text{pandemic\_time} == \text{phase 4}) + \hat{\beta}_5 * (\text{pandemic\_time} == \text{pre-lockdown}) + \hat{\beta}_6 * (\text{is9\_5})$$

At the  $\alpha = 0.05$  significance level, the true  $\beta$  coefficients corresponding to the logit of the probability that a case reported during that time period was related to domestic violence for each of our dummy variables (all relative to the baseline/reference category of before the COVID-19 pandemic) were statistically significant, as was the beta coefficient corresponding to the logit of the probability of success (see above) during the 9-5 work day, as compared to outside this timeframe. We were specifically interested in  $\beta_1$ , which corresponds to the logit of the probability of success (that a reported case was related to domestic violence) during lockdown, as compared to before the pandemic began, and whether there was a relationship between the probability of a reported crime being related to domestic violence during lockdown, when compared to before the pandemic.

$H_0: \beta_1 = 0$  (There is no relationship between our predictor and the probability of success that a reported case is related to domestic violence, while holding time of day (either during the 9-5 workday or outside this timeframe) constant. The true population parameter  $\beta_1$  is equal to 0).

$H_1$ : (There is a relationship between our predictor and the probability of success that a reported case is related to domestic violence, while holding time of day (either during the 9-5 workday or outside this timeframe) constant. The true population parameter  $\beta_1$  is not equal to 0).

Under the null hypothesis, our test statistic follows a standard normal distribution. The value of our test statistic is equal to approximately 12.426, which corresponds to a p-value of less than 0.001. Thus, at the  $\alpha = 0.05$ , we reject our null hypothesis; we have sufficient evidence to suggest that the true value of  $\beta_1$  is not equal to 0, such that this predictor tells us something about the probability of success of our outcome, while holding time of day (either during the 9-5 workday or outside this timeframe) constant.

We see from the model output above that the estimated  $\hat{\beta}_1$  coefficient, corresponding to the difference of being in the “lockdown” phase of time in quarantine - relative to the time period before COVID-19 - on the logit of the probability of success that a reported case will be related to domestic violence, is approximately 0.303. This corresponds to an odds ratio of  $\exp(0.303)$ , which is approximately 1.354. Therefore, we would expect crimes reported during the lockdown phase of the stay-at-home order in the city of Chicago to have 1.354 times the odds of being related to domestic violence, relative to cases that were reported before the COVID-19 pandemic began, while adjusting for time of day, relative to the 9-5 work period. Moreover, we are 95% confident that the true population parameter,  $\exp(\beta_1)$  lies within the interval (3.694, 4.059).

```
## [1] 1.05443
## [1] 1.353914
## [1] 1.110711
## [1] 1.351209
## [1] 1.298228
## [1] 1.013896
## [1] 1.096583
## [1] 1.291702
## [1] 1.419123
## [1] 1.057599
## [1] 1.16649
## [1] 1.279054
## [1] 1.427435
## [1] 1.209783
## [1] 1.393138
```

## DISCUSSION

-summary of what you have learned about your research question along -statistical arguments supporting your conclusions -critique your own methods and provide suggestions for improving your analysis (Issues pertaining to the reliability and validity of your data and appropriateness of the statistical analysis) Make sure to mention possible confounders we did not include in our model, perhaps also the fact that we were just looking at Chicago, vs. the US as a whole or even multiple countries. -what you would do differently -what you would do next if you were going to continue work on the project Are rates of domestic violence during lockdown different in different parts of the world and/or US?