The Rise of Traffic Fatalities During the Pandemic: Seatbelts, Speeding, DUIs

According to press releases by the U.S. Department of Transportation's National Highway Traffic Safety Administration (NHTSA), fatal traffic accidents increased during the COVID-19 Pandemic. According to these prereleases, the NHTSA attributes this rise as positively correlated to the rise of DUI/DWI's, lack of seatbelt usage, and speeding.

This project examines 2020 and 2021 traffic accident data from the NHTSA to determine whether the variables Restraint Violation, DUI Violation, and Speed Violation were significant in explaining the claimed increase in accident mortality during the COVID-19 Pandemic.

The variables examined in this project are as follows:

- Fatality
- Pandemic
- Restraint Violation
- DUI Violation
- Speed Violation
- Pandemic
- Urban
- Rollover
- Surface Conditions

Every variable used in our models is a dummy variable (0 or 1), representing if there was a presence of that variable (1) or not (0).

Fatality is our model's dependent variable. Fatality represents if at least one person died directly from the auto accident.

The Pandemic variable represents the time period of the data set where the Pandemic occurred (March 2020 through December 2020).

The Restraint Violation can be operationalized by the "Failure to require restraint use (by self or passengers)" violation cited in the data.

The DUI Violation was determined as inclusive of the following marked violations:

- Alcohol, drug, or impairment violations generally
- Drinking while operating
- Driving under the influence of a substance not intended to intoxicate
- Driving while intoxicated (alcohol or drugs) or BAC above the limit (any detectable BAC for CDLs)
- Driving with detectable alcohol

The Speed Violation was also determined as inclusive of the following marked violations:

- Exceeding the special speed limit (e.g., for trucks, buses, cycles, or on a bridge, in a school zone, etc.)
- Speed greater than reasonable & prudent (not necessarily over the limit)
- Speed related violations, generally
- Speeding (above the speed limit)

The Urban Variable represents an accident that occurred in an area considered to be urban by the NHTSA, non-urban otherwise.

The Rollover Variable represents an accident that included at least one or more vehicles that rolled over during an auto accident.

The Surface Conditions variable represents the presence of surface conditions that are considered hazardous. The hazardous conditions

marked included under the Surface Conditions variable was determined as inclusive of the following marked conditions:

- Ice/Frost
- Mud, Dirt, or Gravel
- Oil
- Sand
- Slush, Snow
- Water (Standing or Moving), Wet

Tests of Proportions

The following graphic is a set of confusion matrices to show the count of a specific violation by column and whether or not they were pre or post-pandemic by row.

Fatality	0	1	Grand Total
Pre-Pandemic	64592	1237	65829
Pandemic	42196	1129	43325
Grand Total	106788	2366	109154

DUI Violation	0	1	Grand Total
Pre-Pandemic	64154	1675	65829
Pandemic	42117	1208	43325
Grand Total	106271	2883	109154

Speed Violation	0	1	Grand Total
Pre-Pandemic	63668	2161	65829
Pandemic	42062	1263	43325
Grand Total	105730	3424	109154

Restraint Violation	0	1	Grand Total
Pre-Pandemic	65580	249	65829
Pandemic	43119	206	43325
Grand Total	108699	455	109154

Urban	0	1	Grand Total
Pre-Pandemic	14741	51088	65829
Pandemic	11235	32090	43325
Grand Total	25976	83178	109154

The above tables demonstrate that all population sizes are much larger than n=30.

To confirm the base claims made in the NHTSA's report, we will conduct a series of tests of proportions.

The claims to be tested are as follows:

1. There was a significant increase in the proportion of fatal auto accidents after the start of the pandemic.
2. There was a significant increase in the proportion of DUI Violations issued during an auto accident after the start of the pandemic.
3. There was a significant increase in the proportion of Speeding Violations issued during an auto accident after the start of the pandemic.
4. There was a significant increase in the proportion of Restraint Violations issued during an auto accident after the start of the pandemic.

In addition to those claims, we will also test a fifth claim.

5. There was a significant decrease in the proportion of auto accidents in urban areas after the start of the pandemic.

1. *Test of Proportion for Fatalities, Pre vs. Post-Pandemic*

To test whether there was a significant difference in fatalities after the pandemic's start, the author did a proportion test. The following is an R-generated print-out of that proportion test:

```
> prop.test(x = c(1237,1129), n = c(65829, 43325), alternative = "less",
+           correct = TRUE)

        2-sample test for equality of proportions with continuity correction

data:  c(1237, 1129) out of c(65829, 43325)
X-squared = 64.739, df = 1, p-value = 4.275e-16
alternative hypothesis: less
95 percent confidence interval:
 -1.000000000 -0.005718022
sample estimates:
    prop 1     prop 2
0.01879111 0.02605886
```

As we can see, there was indeed a significant increase in traffic mortality after the start of the pandemic to any level of significance (p-val = 4.275e-16).

2. *Test of Proportion for DUI Violations, Pre vs. Post-Pandemic*

To test whether there was a significant difference between DUI violations after the pandemic's start, the author utilized a proportion test. The following is an R-generated print-out of that proportion test:

```
> prop.test(x = c(1675,1208), n = c(65829,43325), alternative = "less",
+           correct = TRUE)

        2-sample test for equality of proportions with continuity correction

data:  c(1675, 1208) out of c(65829, 43325)
X-squared = 5.943, df = 1, p-value = 0.007388
alternative hypothesis: less
95 percent confidence interval:
 -1.0000000000 -0.0007716842
sample estimates:
    prop 1     prop 2
0.02544471 0.02788229
```

There was a moderately significant increase in DUI citations issued after the pandemic's start to at least a significance level of α=0.01 (p-val = 0.007388).

3. *Test of Proportion for Speed Violations, Pre vs. Post-Pandemic*

To test whether there was a significant difference between speeding violations after the pandemic's start, the author utilized a proportion test. The following is an R-generated print-out of that proportion test:

```
> prop.test(x = c(2161,1263), n = c(65829,43325), alternative = "less",
+           correct = TRUE)

        2-sample test for equality of proportions with continuity correction

data:  c(2161, 1263) out of c(65829, 43325)
X-squared = 11.498, df = 1, p-value = 0.9997
alternative hypothesis: less
95 percent confidence interval:
 -1.000000000  0.005447649
sample estimates:
    prop 1     prop 2
0.03282748 0.02915176

> prop.test(x = c(2161,1263), n = c(65829,43325), alternative = "greater",
+           correct = TRUE)

        2-sample test for equality of proportions with continuity correction

data:  c(2161, 1263) out of c(65829, 43325)
X-squared = 11.498, df = 1, p-value = 0.0003484
alternative hypothesis: greater
95 percent confidence interval:
 0.001903785 1.000000000
sample estimates:
    prop 1     prop 2
0.03282748 0.02915176
```

Instead of seeing an increase for citations given post-pandemic, we saw a significant decrease in speeding citations given post-pandemic at an α=0.01 (p-val =0.0003484).

4. *Test of Proportion for Restraint Violations, Pre vs. Post-Pandemic*

To test whether there was a significant difference between speeding violations after the pandemic's start, the author utilized a proportion test. The following is an R-generated print-out of that proportion test:

```
> prop.test(x = c(249,206), n = c(65829, 43325 ), alternative = "less",
+           correct = TRUE)

        2-sample test for equality of proportions with continuity correction

data:  c(249, 206) out of c(65829, 43325)
X-squared = 5.7178, df = 1, p-value = 0.008397
alternative hypothesis: less
95 percent confidence interval:
 -1.0000000000 -0.0002819905
sample estimates:
     prop 1      prop 2
0.003782527 0.004754761
```

A moderately significant increase in traffic restraint violations was issued after the pandemic's start to a significance level of at least α=0.01 (p-val = 0.008397).

5. *Test of Proportion for Urban Auto Accidents, Pre vs. Post-Pandemic*

To test whether there was a significant difference between the number of auto accidents in urban areas pre and post-pandemic, the author implemented a proportion test. The following is an R-generated print-out of that proportion test:

```
> prop.test(x = c(51088,32090), n = c(65829, 43325 ), alternative = "greater",
+           correct = TRUE)

        2-sample test for equality of proportions with continuity correction

data:  c(51088, 32090) out of c(65829, 43325)
X-squared = 180.27, df = 1, p-value < 2.2e-16
alternative hypothesis: greater
95 percent confidence interval:
 0.03099672 1.00000000
sample estimates:
   prop 1    prop 2
0.7760713 0.7406809
```

There was an extremely significant decrease in the proportion of auto accidents in urban areas after the pandemic's start to any level of significance (p-val < 2.2e-16). This means that the probability of an auto accident to occur in an urban area (as opposed to a rural area) decreased after the pandemic began, supporting the evidence that the migration of people leaving urban areas for socially-distanced rural areas may be a possible explanation for a change in fatality rate.

Comparing Regressions Pre and Post Pandemic

```
> glmmodpre = glm(Fatality ~  DUI_Vio + Restraint_Vio + + Speed_Vio  +
+                   Urban + Rollover +SurfaceCond , family = binomial)
> summary(glmmodpre)

Call:
glm(formula = Fatality ~ DUI_Vio + Restraint_Vio + +Speed_Vio +
    Urban + Rollover + SurfaceCond, family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.5220  -0.1739  -0.1739  -0.1739   3.3900

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   -3.84399    0.06030 -63.744  < 2e-16 ***
DUI_Vio       -0.51886    0.20786  -2.496   0.0126 *
Restraint_Vio -0.23000    0.42463  -0.542   0.5881
Speed_Vio     -1.31231    0.28162  -4.660 3.16e-06 ***
Urban         -0.34068    0.06578  -5.179 2.23e-07 ***
Rollover       1.91954    0.07685  24.978  < 2e-16 ***
SurfaceCond   -0.24570    0.08251  -2.978   0.0029 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 11502  on 62565  degrees of freedom
Residual deviance: 10969  on 62559  degrees of freedom
AIC: 10983

Number of Fisher Scoring iterations: 7
```

```
> glmmodpan = glm(Fatality ~  DUI_Vio + Restraint_Vio + + Speed_Vio  +
+                   Urban + Rollover +SurfaceCond , family = binomial)
> summary(glmmodpan)

Call:
glm(formula = Fatality ~ DUI_Vio + Restraint_Vio + +Speed_Vio +
    Urban + Rollover + SurfaceCond, family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.5191  -0.2097  -0.2097  -0.2097   3.4367

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   -3.64563    0.05809 -62.761  < 2e-16 ***
DUI_Vio       -0.34628    0.18232  -1.899   0.0575 .
Restraint_Vio -0.61777    0.45997  -1.343   0.1792
Speed_Vio     -1.26038    0.29315  -4.299 1.71e-05 ***
Urban         -0.16059    0.06410  -2.505   0.0122 *
Rollover       1.70915    0.07406  23.079  < 2e-16 ***
SurfaceCond   -0.21852    0.09142  -2.390   0.0168 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 11238  on 46587  degrees of freedom
Residual deviance: 10790  on 46581  degrees of freedom
AIC: 10804

Number of Fisher Scoring iterations: 7
```

	Pre-Pandemic		Pandemic	
	Estimated Coefficient	Pr(>|z|)	Estimated Coefficient	Pr(>|z|)
(Intercept)	- 3.84399	< 2e-16 ***	-3.64563	< 2e-16 ***
DUI_Vio	-0.51886	0.0126 *	-0.34628	0.0575 .
Restraint_Vio	-0.23000	0.5881	-0.61777	0.1792
Speed_Vio	-1.31231	3.16e-06 ***	-1.26038	1.71e-05 ***
Urban	-0.34068	2.23e-07 ***	-0.16059	0.0122 *
Rollover	1.91954	< 2e-16 ***	1.70915	< 2e-16 ***
SurfaceCond	-0.24570	0.0029 **	-0.21852	0.0168 *

To test whether the variables were significant for both periods, we first compare two regressions: one with the data from January 2019 through February 2020 (Pre-Pandemic) and another from March 2020 through December 2020 (Pandemic).

The table above directly compares the estimated coefficients and significance levels from both regressions. As we can see, most variables tested are very similar in coefficient estimate and significance, with a few exceptions.

The first such exception would be Restraint Violations. Pre-Pandemic restraint violations were not a significant explanatory variable under any significance level (p=0.5881), but did gain significance during the Pandemic. However, the significance gained is not enough to consider the variable statistically relevant (p=0.1792). Another point of interest

is to compare the estimated coefficient pre-pandemic (-0.23000) and during the pandemic (-0.61777). The estimated coefficient became more negative during the pandemic, meaning that given the presence of a citation issued for a failure to properly use a restraint (such as a seatbelt), the likelihood of an auto accident resulting in a fatality became even less likely during the pandemic. This evidence is contrary to the NHTSA's conclusions.

Also contrary to the NHTSA's conclusions are the negative values of the estimated coefficients for DUI Violations and Speeding Violations, for both the pre-pandemic and pandemic sets. This means that the presence of a citation issued for either of these violations during an auto accident indicates the accident is less likely rather than more likely to result in a fatality, contrary to what the NHTSA would lead us to expect. Both of these variables have been found to be extremely significant variables for both the pre-pandemic and pandemic models.

The Urban variable became a less significant variable to predict a fatal auto outcome after the start of the pandemic. We see that urban areas had a negative correlation (-0.34068) with auto fatalities pre-pandemic and still was negative after the pandemic but less so (-0.16059). This could mean that while urban areas became less deadly for auto accidents, they also became a weaker predictor of whether an accident would become deadly.

We see that Rollovers were an extremely significant predictor of fatal auto accident outcomes both pre-pandemic and during (both at p-val<2e-16). Both regressions showed positive correlations between the rollovers and fatal accidents, meaning that a rollover during an accident is more likely to predict a fatal outcome. We saw that the Rollover coefficient decreased from a pre-pandemic estimate of 1.91954 to a pandemic 1.70915. This could mean that rollovers became less deadly during the pandemic.

Dangerous surface conditions seemed to remain just about the same pre and post pandemic, both in estimated coefficient and statistical significance.

	Mod 0		Mod 1		Mod 2		Mod 3	
	Estimated Coefficient	Pr(>|z|)	Estimated Coefficient	Pr(>|z|)	Estimated Coefficient	Pr(>|z|)	Estimated Coefficient	Pr(>|z|)
(Intercept)	- 3.82894	< 2e-16 ***	-3.87579	< 2e-16 ***	- 3.79875	< 2e-16 ***	- 3.85860	< 2e-16 ***
DUI_Vio	- 0.42094	0.00212 **	-0.42496	0.001928 **	-1.38415	7.92e-05 ***	- 1.33192	9.03e-05 ***
DUI_Vio:Pandemic					0.07674	0.778658		
DUI_Vio:Urban					1.23121	0.000925 ***	1.21182	0.001079 **
Restraint_Vio	- 0.42141	0.17643	-0.43421	0.163770	0.50145	0.242891	- 0.41008	0.188532
Restraint_Vio:Pandemic					- 1.24308	0.075556 .		
Restraint_Vio:Urban					- 1.01126	0.147035		
Speed_Vio	- 1.29282	1.93e-10 ***	-1.28464	2.52e-10 ***	- 1.40768	0.000279 ***	-1.28128	2.81e-10***
Speed_Vio:Pandemic					- 0.34607	0.412372		
Speed_Vio:Urban					0.39926	0.359445		
Pandemic			0.27209	1.09e-10 ***	0.15135	0.047063 *	0.27269	9.94e-11 ***
Urban	- 0.26056	1.38e-08 ***	- 0.24898	5.96e-08 ***	- 0.36393	5.90e-08 ***	- 0.27442	3.16e-09 ***
Urban:Pandemic					0.19255	0.034320 *		
Rollover	1.82553	< 2e-16 ***	1.81081	< 2e-16 ***	1.81053	< 2e-16 ***	1.81123	< 2e-16 ***
SurfaceCond	- 0.25479	3.03e-05 ***	- 0.22668	0.000214 ***	- 0.22950	0.000183 ***	- 0.22764	0.000202 ***

Creating The Final Regression

A series of regressions were created to create the best explanatory model. A table of the 4 regressions (0 through 3) is shown above, with their respective estimated coefficients and levels of significance.

The Mod 0 is a printout of the summary results for a regression of all the data, but without the Pandemic as its own variable. After adding Pandemic as its own variable in Mod 1, we can see that there is something about the presence of the Pandemic that is significant in explaining the rise in fatal car crashes.

In Mod 2, we look to see what level of collinearity is present among the variables. A moderate or lower level of collinearity is acceptable in a multivariable regression and can be ignored for our final regression.

For DUI Violations, we see an insignificant amount of collinearity with the Pandemic Variable but an extremely significant amount of collinearity with the Urban Variable. DUI_Vio:Urban has a positive estimated coefficient (1.23121), which means that the interaction between DUI Violations and in an Urban area is more likely to result in a fatal auto accident, all other things being equal.

For Restraint Violations, we see that both the Pandemic and Urban variables have a low and an insignificant amount of collinearity, respectively.

For Speed Violations, we see that both the Pandemic and Urban variables share an insignificant amount of collinearity.

Lastly, we also find only a moderate level of collinearity between the Urban and Pandemic variables.

From this version of the model, we conclude that we need to include the collinear DUI_Vio:Urban in our final model.

Mod 3 is our final regression model to describe how our chosen explanatory variables correlated to fatal auto accidents.

1. DUI Violations

   This variable was highly significant, but negatively correlated with Fatal accidents. For every auto-accident where a DUI citation is issued, the odds that the accident also includes a fatality decreases.

2. Restraint Violations

   Unlike what was stated by the NHTSA, we did not find restraint violations to have even a moderate level of significance to explain auto fatalities as a result of an accident.

3. Speeding Violations

   This variable was highly significant, but negatively correlated with Fatal accidents. For every auto-accident where a SpeedingI citation is issued, the odds that the accident also includes a fatality decreases.

4. Urban

   This variable was also highly significant, but negatively correlated with Fatal accidents. For every auto-accident that occurs in an urban area, the odds that the accident also includes a fatality decreases. Inversely,for every auto-accident that occurs in an non-urban area, the odds of the accident also being fatal increases.

5. Surface Conditions

This variable was highly significant, but also negatively correlated with Fatal accidents. For every auto-accident where a dangerous surface condition was present, the odds that the accident also includes a fatality decreases.

6. DUI:Urban

   The collinearity between DUI and Urban was statistically significant in this model and positively correlated to Fatality. Oddly enough, if a DUI was issued and the accident occurred in an urban area, the odds of a fatality also occurring increases.

7. Rollover

   This variable was highly significant, and our sole theoretical value that was positively correlated with Fatal accidents. For every auto-accident where a rollover occurs, the odds that the accident also includes a fatality increases.

8. Pandemic

   Whatever aspect of the Pandemic changed driving/road conditions, we could not eliminate the presence of the pandemic through exploration of multicollinearity of our theoretical variables alone. As a highly significant variable, the era of the pandemic was positively correlated with auto fatalities.

Conclusion

While we were able to find support that two of the variables put forth by the NHTSA (DUI and Speeding violations) were strong variables in predicting fatal auto accidents, we found that it had the opposite expected outcome than was offered by the NHTSA.

Also contrary to what was offered by the NHTSA, we did not find enough evidence to include Restraint violations as a significant variable in explaining the rise in Auto-accidents.

Links:

1. https://www.nhtsa.gov/press-releases/early-estimate-2021-traffic-fatalities
2. https://www.nhtsa.gov/press-releases/2020-fatality-data-show-increased-traffic-fatalities-during-pandemic
3. https://www.nhtsa.gov/file-downloads?p=nhtsa/downloads/CRSS/2020/
4. https://www.nhtsa.gov/file-downloads?p=nhtsa/downloads/CRSS/2021/