"Are Opiates Influencing Property Crimes?"
A Study Using Text Analysis on Vermont State Police Narratives

Submitted to:

Department of Public Safety

Submitted by:

Robin Joy, J.D., Ph.D.
Crime Research Group
P.O. Box 1433
Montpelier, VT 05601
802-230-4768
www.crgvt.org

September 2019

This Project was funded by
BJS 2016 State Statistical Analysis Grant #BJ-CX-K016

Introduction

The public's perception in Vermont has been that there is an unprecedented opiate crisis occurring creating a dramatic spike in drug related property crimes. Anecdotally, law enforcement throughout the state reported increasing suspicion that there is a connection between the number of opiate crimes and drug related property crimes. This research aimed to answer the question of whether the escalation of opiate and other drug use has been influencing property crimes in Vermont. This project examines the role of drugs in property crimes investigated by the Vermont State Police in 2017. New techniques were explored to enhance existing interfaces and data exchange technologies that improve statistical and research access to law enforcement data and data systems. The data exchange capacity established the use of text analysis in R to examine the narratives in administrative and operational law enforcement databases to answer the research question.

Methodology

Several techniques were used to understand the relationship between property crimes and opiates and/or other drugs. First, a key word dictionary was created and matched against the Vermont State Police (VSP) narratives. This helped identify incidents potentially related to drugs. The incidents were matched into NIBRS (National Incident Based Reporting System) for more specific details of the incidents and offenses. If there was an arrest in a property incident, the criminal histories of the defendants were used and categorized their offending. The final step was to conduct a text analysis of the narrative using key words and topic modeling (Appendix A).

The VSP provided the narratives for all property crimes investigated during 2017. The narratives (1,448 were provided, 1,446 were used in the analysis) were read into R, a statistical program, and read against a dictionary of key words that might appear in a drug related case. The dictionary contained words such as "heroin," "opiates," "drug," etc. The dictionary was run against the narratives; the result was a count of how often these words appeared in a narrative. "Heroin" appeared in 27 narratives, and, in one narrative, was mentioned 21 times. Most narratives that contained the word "heroin" used it less than three times. For purposes of analysis, a case was labeled "drug related" if there were at least three occurrences of any key word in the narrative.[1]

Incidents were matched into the NIBRS data, using the FBI's downloadable NIBRS files for 2017. Unfortunately, the FBI incident numbers reported in 2017 are not formulated in the way Vermont incident numbers are formulated. To account for this, the incidents, and date and hour of the offense, were matched by VSP barracks. If the barracks responded to two offenses at the same hour, the property offense was kept as the correct incident. If the barracks responded to two property offenses on the same date and hour, both property offenses were eliminated from the

[1] After reading the incident reports and checking on the matches, three hits on any keyword in a narrative was the minimum number of times to avoid mistaken identification. For example, if the word "drugs" appeared only once, with no other keywords, it was unlikely to be relevant to the narrative.

analysis. The FBI does not include Group B offenses in the data, meaning that offenses for bad checks, a Group B offense, could not be analyzed.

Names were provided to Vermont Crime Information Center (VCIC) to obtain the criminal history of defendants arrested for property crimes. Using the definition of "drug related," criminal history patterns were analyzed for those arrested in drug related incidents compared to those who were arrested for incidents not drug related. Cases were tracked through disposition and sentencing. Topic modeling was applied to all narratives allowing a deeper analysis of the narratives.

NIBRS Analysis

Table 1 depicts the types of offenses associated with the property incidents investigated by the VSP. Incidents may have more than one associated offense. For example, an incident could contain both a burglary and an assault. The *% of Total Within Category* in Table 1 indicates whether incidents labeled as "Drug Related" differ from those labeled "Not Drug Related"[2] For example, of those labeled "drug related" there are proportionally fewer Thefts from Motor Vehicles (2.22%) than those labeled not drug related (11.43%).

Table 1: Drug Related Crimes Compared to Not Drug Related Crimes

Drug	Offense Name	Number of Offenses	% of Total Within Category
Not Drug Related	Aggravated Assault	4.0	0.27%
	All Other Larceny	393.0	26.57%
	Burglary/Breaking & Ente..	576.0	38.95%
	Counterfeiting/Forgery	4.0	0.27%
	Credit Card/Automated T..	2.0	0.14%
	Destruction/Damage/Van..	61.0	4.12%
	Drug/Narcotic Violations	3.0	0.20%
	Embezzlement	1.0	0.07%
	False Pretenses/Swindle/..	5.0	0.34%
	Intimidation	1.0	0.07%
	Motor Vehicle Theft	12.0	0.81%
	Shoplifting	80.0	5.41%
	Simple Assault	11.0	0.74%
	Stolen Property Offenses	18.0	1.22%
	Theft From Building	136.0	9.20%
	Theft From Coin-Operate..	1.0	0.07%
	Theft From Motor Vehicle	169.0	11.43%
	Weapon Law Violations	2.0	0.14%
Drug Related	All Other Larceny	11.0	24.44%
	Burglary/Breaking & Ente..	16.0	35.56%
	Credit Card/Automated T..	1.0	2.22%
	Destruction/Damage/Van..	3.0	6.67%
	Drug/Narcotic Violations	4.0	8.89%
	Shoplifting	1.0	2.22%
	Stolen Property Offenses	3.0	6.67%
	Theft From Building	5.0	11.11%
	Theft From Motor Vehicle	1.0	2.22%

It's important to note that there are very small numbers of drug related offenses in this data set.

[2] We use the term "Not Drug Related" for ease of reading. Properly, it should read: "No Evidence that this case was Drug Related in the Narrative."

Location of Incidents

Table 2: Location of Crimes for Drug Related and Not Drug Related Offenses

Drug	Location Name	Number of Incidents	% of Total in Category
Not Drug Related	Abandoned/Condemned Struct..	1.0	0.08%
	Air/Bus/Train Terminal	1.0	0.08%
	Auto Dealership New/Used	7.0	0.53%
	Bank/Savings and Loan	8.0	0.61%
	Bar/Nightclub	2.0	0.15%
	Camp/Campground	22.0	1.67%
	Church/Synagogue/Temple/Mo..	5.0	0.38%
	Commercial/Office Building	19.0	1.44%
	Construction Site	5.0	0.38%
	Convenience Store	52.0	3.94%
	Department/Discount Store	28.0	2.12%
	Drug Store/Doctor's Office/Hos..	5.0	0.38%
	Farm Facility	15.0	1.14%
	Field/Woods	31.0	2.35%
	Government/Public Building	10.0	0.76%
	Grocery/Supermarket	29.0	2.20%
	Highway/Road/Alley/Street/Si..	49.0	3.71%
	Hotel/Motel/Etc.	10.0	0.76%
	Industrial Site	1.0	0.08%
	Lake/Waterway/Beach	10.0	0.76%
	Liquor Store	1.0	0.08%
	Other/Unknown	24.0	1.82%
	Park/Playground	3.0	0.23%
	Parking/Drop Lot/Garage	62.0	4.69%
	Rental Storage Facility	9.0	0.68%
	Residence/Home	833.0	63.06%
	Restaurant	16.0	1.21%
	School-Elementary/Secondary	1.0	0.08%
	School/College	16.0	1.21%
	Service/Gas Station	35.0	2.65%
	Shopping Mall	1.0	0.08%
	Specialty Store	40.0	3.03%
Drug Related	ATM Separate from Bank	1.0	3.13%
	Commercial/Office Building	1.0	3.13%
	Drug Store/Doctor's Office/Hos..	2.0	6.25%
	Highway/Road/Alley/Street/Si..	4.0	12.50%
	Hotel/Motel/Etc.	1.0	3.13%
	Lake/Waterway/Beach	1.0	3.13%
	Residence/Home	24.0	75.00%
	School/College	1.0	3.13%
	Specialty Store	1.0	3.13%

Table 2 shows the location of the crime. Private residence is the most significant category for location of property crimes regardless of whether these crimes are drug related. Drug related incidents were more likely to take place at a drug store/doctor's office/hospital (6.25% vs. .38%) and on the street/sidewalk (12.5% vs. 3.7%). However, the number of drug incidents identified in the NIBRS data (36) is really too small to draw any conclusions.

Victims

As noted in Table 3, drug related incidents had 55 victims, and not drug related incidents had 1,646 victims. Individual was the most significant category of victim. There can be more than one victim in an incident.

Table 3: Victim Type

	Not Drug Related	Drug Related
Business	245	3
Financial Institution	5	
Government	15	
Individual	1,359	43
Other	6	
Religious Organization	4	
Society/Public	12	9

Victim to Offender Relationship

The victim to offender relationship was missing in all 43 incidents identified as drug related and was missing in 1,321 of the 1,370 incidents identified as not drug related. A victim to offender relationship is only reported when the offender is arrested and the victim is an individual. The offender is not arrested in all cases.

Residency of the Arrestee

NIBRS documents the residency of the arrestee, shown in Table 4. For VSP, a person is a resident if they live in the jurisdiction of the barracks.

Table 4: Residency of the Arrestee

	Not Drug Related	Drug Related
Non-Resident	46	6
Missing	86	18
Resident	194	16

As seen in Table 4, resident status was missing in 104 arrests. For the arrests that identified resident status, the proportion of residents to non-residents were the proportional.

Criminal Histories

VSP provided the names and dates of birth of 216 individuals arrested in connection with the 1,448 incidents in the narratives. VCIC matched 182 of these individuals with criminal histories. Of the 182 individuals, 18 (10%) were identified as being involved in drug related incidents.

Demographics of Arrestees

Age of arrestee: The average age at arrest for both drug related and not drug related incidents was 31.

Table 5: Race and Gender of Arrestee

		Drug Related	Not Drug Related	Grand Total
Asian	Male		1	**1**
Black	Female		1	**1**
	Male		3	**3**
White	Female	5	51	**56**
	Male	12	96	**108**
Unkown	Male		3	**3**
Missing	Female		5	**5**
	Male	1	4	**5**
Grand Total		**18**	**164**	**182**

Only white defendants were arrested for drug related property incidents. Arrest of white females was proportionally the same for both drug related and not drug related incidents. White males were arrested for drug related property offenses at 66.7%, and not drug related at 58.5%.

Prior Incidents

Of the 182 defendants arrested, 53 were first-time offenders in Vermont. Five of the first-time offenders were in identified as being involved in drug related incidents, 48 first-time offenders were not in drug related incidents. The 129 defendants arrested who had a prior criminal history are represented in Table 6.

Table 6 presents the prior criminal histories of the defendants in this study. Thirteen unique defendants were in the drug related incidents and 116 unique defendants were in the not drug related incidents. Keep in mind that one person often has multiple charges.

Table 6: Prior Criminal Histories – Charges and Individuals

	Drug Related		Not Drug Related	
	Number of Charges	Number of People	Number of Charges	Number of People
Missing	54	12	1,072	98
Fish and Game	14	2	17	3
Public Order	196	12	2,380	93
Motor Vehicle	124	6	882	66
Drugs	28	6	174	33
Frauds	76	4	195	22
Thefts	71	8	966	81
GNO	3	1	43	15
DUI	34	7	285	43
Weapons			1	1
Assault	8	4	227	48
VAPO	14	1	87	16
Robbery			15	6
Domestic	11	3	135	34
Sex Offenses			18	10
Grand Total	**633**	**13**	**6,497**	**116**

Table 6 shows the number of charges for each crime category and the number of people responsible for those charges. For example, two individuals in the drug related category were arrested for 14 Fish and Game violations. Almost half of the drug related defendants (6) had a prior arrest for a drug violation, compared to approximately 30% of the non-drug related defendants. Over half of each group had prior arrests for theft offenses.

Dispositions

Of the 182 defendants, 117 had the 2017 base incident disposed of in the criminal court shown in Table 7. For example, three people in a drug related incident had a public order charge dismissed and two people received a misdemeanor conviction.

Table 7: Dispositions for Offenses by Individuals and Charges

		Drug Related			Not Drug Related		
		Dismissed	Felony Conviction	Misdemeanor Conviction	Dismissed	Felony Conviction	Misdemeanor Conviction
Missing	Number of Charges	5.00	4.00		9.00	39.00	11.00
	Number of People	4.00	4.00		8.00	21.00	7.00
Public Order	Number of Charges	3.00		3.00	30.00	5.00	41.00
	Number of People	3.00		2.00	15.00	5.00	27.00
Motor Vehicle	Number of Charges	1.00		1.00	2.00	3.00	4.00
	Number of People	1.00		1.00	2.00	3.00	3.00
Drugs	Number of Charges	1.00	1.00	2.00			1.00
	Number of People	1.00	1.00	2.00			1.00
Frauds	Number of Charges	3.00		1.00	14.00	6.00	4.00
	Number of People	1.00		1.00	4.00	6.00	4.00
Thefts	Number of Charges	7.00		11.00	30.00	15.00	51.00
	Number of People	5.00		11.00	18.00	10.00	42.00
GNO	Number of Charges						1.00
	Number of People						1.00
Assault	Number of Charges				8.00	2.00	3.00
	Number of People				6.00	2.00	3.00
VAPO	Number of Charges						1.00
	Number of People						1.00
Domestic	Number of Charges				3.00	1.00	
	Number of People				3.00	1.00	

Drug related incidents earned five felony convictions and 18 misdemeanor convictions, 12 for theft or fraud. Not drug related incidents earned 71 felony convictions and 71 misdemeanor convictions.

Table 8 shows the sentence, number of charges and number of people by felony or misdemeanor conviction. Deferred and split sentences were used in very few property incident dispositions.

Table 8: Disposition Type by Number of People and Charges

		DrugDefendant / Disposition			
		Drug Related		Not Drug Related	
SENTENCE		Felony Conviction	Misdemeanor Conviction	Felony Conviction	Misdemeanor Conviction
Missing	Number of Charges				1.00
	Number of People				1.00
DEFERRED	Number of Charges	2.00	3.00	6.00	2.00
	Number of People	2.00	2.00	6.00	2.00
FINE ONLY	Number of Charges				10.00
	Number of People				7.00
PROBATION	Number of Charges	1.00	9.00	11.00	42.00
	Number of People	1.00	6.00	8.00	23.00
SPLIT	Number of Charges			3.00	1.00
	Number of People			2.00	1.00
STRAIGHT	Number of Charges	2.00	6.00	51.00	61.00
	Number of People	2.00	6.00	25.00	34.00

Text Analysis of Narratives

VSP provided the full narratives in a .csv text file for 1,448 incidents. Two incidents contained formatting errors which affected the structure of the data, and they were removed from the analysis.

Text analysis aims to understand topics relationships between words in a body of work. For this project, we were interested in the relationship between opiates, drugs, and property crimes. As with any statistical analysis, text analysis requires cleaning and formatting the data to make the analysis meaningful.[3]

First, words are restructured to their roots. The words “garage” and “garages” and “garaged” share the root “garage.” In text analysis, the root of the word provides the meaning and frequencies of the word and is more valuable than the frequencies of all the variants. Second, some words are excluded from analysis. For example, the word “I” is not meaningful in analyzing themes of property crimes. Words that generally hold little meaning in English are called stopwords, and 172 have been identified. Stopwords were removed from the analysis.

[3] The analysis was done in R. The full R script is in Appendix A.

Likewise, there are words that are used so often in formulaic writing that they have no meaning in relationship to the analysis. The most frequent word in the narratives was “advised.” Troopers either “were advised” or “did advise." Similarly, “dob,” (date of birth) was very common. These words add nothing to the intended analysis, so a calculation is used for identification and removal. There is a statistic called the *term frequency inverse document frequency* (tf idf). This statistic calculates a weight of the word. Very frequent words (tf) are weighted less than infrequent words (idf). Words with lower weight were excluded from the analysis.

Key Word Analysis

The hypothesis was that if property crimes were being driven by the opiate crisis, then there would be evidence of that in the narratives. The hypothesis was tested by creating a word dictionary, inclusive of the words “heroin,” “opiates,” and “drug.” The dictionary was processed against the narratives; the result was a count of how often these words appeared in a narrative. “Heroin” appeared in 27 narratives. In one narrative, it occurred 21 times. Most narratives that contained the word “heroin” used the word less than 3 times.

Overall, the dictionary flagged 118 out of 1,446 narratives that mentioned a drug key word. This represents eight percent of the narratives in which officers noted drugs as an issue.

Topic Analysis

Text analysis can also create topic models. A topic model scans the body of work and identifies words that tend to appear together with some probability. We used Latent Dirichlet Allocation (LDA) modeling to create topics and associated words. The model returned some interesting results shown in Chart 1. First, the narratives contained the names of victims, witnesses, and defendants. One will see in the topic chart that names are associated with topics. For example, topic three associates the name, Jason, with terms related to jewelry. Jason is a common name and at least two troopers in the dataset have the name Jason. One would expect that a topic that includes diamonds also includes gold and ring. The presence of this topic in the model suggests that jewelry thefts are common in the data.

Of particular interest are examples contained in topic nine, where the word “juvenile” is strongly correlated with “wallet” and “atm,” and topic ten, associating “school” with “wallet.” This may indicate that youth are more at risk, or exposed to stolen wallets, or that youth steal wallets.

Removing proper names from the dataset proved more challenging than anticipated. The most common names published by the Social Security Administration were added to the list of stopwords. However, when running that function on the data, the system crashed. The error is still being investigated. Until this error is corrected, the model has limited use.

Chart 1 below shows the topic analysis for property crimes:

term
beta
1
shaw
tolp
mcbain
brent
brown
price
chopper
spencer
joann
sport
2
edward
ead
adkinslegrand
bailey
juli
sanborn
hunt
verdon
safe
ball
0.00 0.01 0.02
3
jason
ring
hein
stephani
defend
gold
purs
jewelri
diamond
johnson
4
kateria
labri
bank
niiler
alic
montgomeri
sheltra
jolli
davi
pill
5
thoma
safe
dog
bockus
patrick
low
reed
priscilla
bottl
lawton
6
tool
tiffani
duroch
pillsburi
cooper
abbott
furlow
doug
wheeler
dewalt
7
brown
goodel
farnsworth
bunnel
walsh
trailer
jame
plant
carlen
garag
0.00 0.01 0.02
8
gun
firearm
sargent
gambl
pistol
william
truck
kelli
abbott
guerin
9
juvenil
wallet
atm
credit
baker
bank
max
walmart
cowdrey
ingal
demasi
0.00 0.02 0.04
10
mailbox
king
school
william
credit
wallet
petri
laflamm
cpl
basso
11
smith
churchil
greenwood
legaci
curschmann
savo
truck
turner
brown
macdonald
0.00 0.02 0.04
12
camp
robert
ryan
man
daniel
howard
dog
fish
dna
joseph
13
georg
barn
desjadon
trailer
unit
michael
safe
zachari
brault
storag
14
bank
account
benoit
shawn
jone
duvan
heather
deposit
benjamin
jacob
15
john
erica
megaro
ethan
scott
macnevin
gosselin
greenwood
jacqu
jeep
16
wood
gas
generat
craig
nick
vanc
guertin
gun
larson
pump
17
truck
cook
ashley
greg
rajda
mari
kristina
curti
young
garag
18
alysha
despain
dumont
colbi
justin
lussier
brace
odonnel
blake
larrow
19
storm
chris
christoph
cota
larri
jessica
phillip
horner
hammond
barreda
20
boat
motor
guyett
sullivan
jessica
sweet
martin
trailer
gander
garag

Findings:

1. Using the key word analysis only 8% of the narratives in the VSP data returned an indication that the incidents are drug related crimes. The key words included drug as well as opiates and heroin.

2. Because drugs were so infrequently mentioned in the data, none of the topics yielded the expected result showing a relationship between the key words drugs and heroin with property crimes. Research on this topic has noted that this current opioid crisis has not been associated with a rise in property crime.[4]

3. Reviewing narratives using text analysis may help police include more relevant information in the narratives.

4. There were no discernible differences between incidents labeled drug related and those labeled not drug related.

5. Because the number of drug related incidents were so small, the results are not generalizable.

6. Though the results of this analysis didn't yield the expected result, the intention of this project was also to use new techniques to enhance existing interfaces and data exchange technologies that improve statistical and research access. The data exchange and analysis capacity established in this project was the use of text analysis in R to examine the narratives in administrative and operational law enforcement data.

7. The code written for key word analysis will allow stakeholders in Vermont to obtain this type of analysis for many areas of interest. Long narratives and text documents can now be analyzed quickly. Text and Sentiment Analysis in R allows for large amounts of text-based data to be analyzed for content, strength of relationships between word objects and words conveying emotion or intent.

8. Topic analysis will provide more value once the proper name and a few other technical issues are worked out. It will provide insight into how crime in Vermont may be categorized.

---

[4] Szalavitz, M. Rigg, K.K., Substance Use & Misuse; 2017, Vol. 52, No. 14, 1027-1931; https://doi.org/10.1080/10826084.2017.1376685. Drug epidemics often bring with them an accompanying rise in crime. The heroin wave of the 1970's and crack crisis of the 1980's were each accompanied by major gun violence, including large numbers of murders and violent property crimes. The current United States opioid epidemic, however, has not been associated with either a rise in homicide or in property crime.

Appendix A

```
library(tidytext)
library(tm)
library(ggplot2)
library(dplyr)

#make vector of narrative
narrative_source <- VectorSource(drugincidentswithnarratives$Narrative)
#make corpus
narrative_corpus <- VCorpus(narrative_source)

###dictionary of key words
My_words<-  c(“drug”, “opiate”, “heroin”, “fentanyl”,  “sobriety”, “spoon”, “marijuana”, “cocaine”)
####create document term matrix
narrative_dtm <- DocumentTermMatrix(narrative_corpus, control = list(stemming = TRUE, stopwords =
TRUE,
minWordLength = 2, removeNumbers = TRUE, removePunctuation = TRUE, tolower = TRUE, dictionary =
my_words))
str(narrative_dtm)
# create data.frame from documenttermmatrix
df1 <- data.frame(docs = narrative_dtm$dimnames$Docs, as.matrix(narrative_dtm), row.names = NULL)

######new document term maitrix for LDA analysis
new_dtm <- DocumentTermMatrix(narrative_corpus, control = list(stemming = TRUE, stopwords =
TRUE,
minWordLength = 2, removeNumbers = TRUE, removePunctuation = TRUE))
str(new_dtm)
###  remove some words
term_tfidf <- tapply(new_dtm$v/slam::row_sums(new_dtm)[new_dtm$i], new_dtm$j, mean) *
log2(tm::nDocs(new_dtm)/slam::col_sums(new_dtm > 0))
summary(term_tfidf)
#run of median of.04 (.7 for drug incidents only)
reduced_dtm <- new_dtm[,term_tfidf >= 0.7]
summary(slam::col_sums(reduced_dtm))
#model came back with 0 entry rows on one run.  Find empty rows
rowTotals <- apply(reduced_dtm , 1, sum)
#remove all docs without words
dtm.new <- reduced_dtm[rowTotals> 0, ]
###  LDA model
lda_model2 <- LDA(dtm.new, 20)
p_topics <- tidy(lda_model2, matrix = "beta")
p_topics
####plot them
p_top_terms <- p_topics %>%
group_by(topic) %>%
top_n(10, beta) %>%
```

```
ungroup() %>%
arrange(topic, -beta)
p_top_terms %>%
mutate(term = reorder_within(term, beta, topic)) %>%
ggplot(aes(term, beta, fill = factor(topic))) +
geom_col(show.legend = FALSE) +
facet_wrap(~ topic, scales = "free") +
coord_flip() +
```