Does Coverage of Sexual Assault Cases Ease the Reporting Decision? Evidence from FBI Data and Google Trends

Harry Elworthy

Professor Patrick Bayer, Faculty Advisor

Honors Thesis submitted in partial fulfillment of the requirements for Graduation with Distinction in Economics in Trinity College of Duke University.

Duke University

Durham, North Carolina

2019

Acknowledgements

I am grateful for the advice of Professor Patrick Bayer, along with Professors Kent Kimbrough and Grace Kim.

Replication files are available at `www.github.com/harryelworthy/Thesis`

Abstract

This paper investigates the effect that national news coverage related to sexual assault has on the reporting decision of assault victims. Estimates are based on time series data of reports made to police stations in the US from 2008 to 2016 and Google Trends data of search volume, along with an identification strategy that uses a number of high profile sexual assault allegations and events, also collected using Google Trends. By removing events that occurred on the day that they were reported, I plausibly estimate only the effect of coverage on the reporting of assaults, and not on assaults themselves. A significant positive effect is found using several specifications. Back-of-the-envelope calculations suggest that there were between 31 and 121 additional reports of sexual assault for each of the 38 high profile events captured. No evidence is found to suggest that these additional reports have different arrest rates to other reports, indicating that there are not a significant number of false reports. This paper adds to current literature on the sexual assault reporting decision considering the effect of news coverage and by using new methods of inference.

JEL Codes: D91;J16;K42;L86;Z13

Keywords: Crime; Gender; Sexual Harassment

Introduction

An estimated 18.3% of women and 1.4% of men in the United States are sexually assaulted at some point in their lives, with more than a third of these assaults occurring before the victim turns 18 (Black et al., 2011). About 20% to 25% of women nationally are sexually assaulted at some point during their college careers (Fisher, Cullen, & Turner, 2000). Despite this, very few assaults are reported to police. According to Fisher et al. (2000), "fewer than 5% of attempted or completed sexual assault against college age women are reported to law enforcement. 66% of victims tell friends but not family or school officials." Resnick et al. (2000) finds that 20% of women that experience sexual assault report it to police, while Greenfield (1997) estimates this number at 30% of women. Because these studies use self-reported survey data, there are still almost certainly still non-reporting women and thus these estimates are likely upper bounds on the proportion of sexual assault victims that report their crimes. Victims do not report to police for a number of reasons, discussed more in-depth below, including self-blame, guilt, fear of the perpetrator or fear of not being believed (Du Mont, Miller, & Myhr, 2003).

There are many reasons to seek to increase the proportion of sexual assaults that are reported. Some victims of sexual assault may not want to report their crime to police at all because they do not wish their perpetrator to face justice, often because they are friends or are otherwise close. However, many victims would prefer to report their crime to the police, and balance this desire against the costs they see in reporting (Du Mont et al., 2003). There cannot be a criminal investigation if the crime is not reported to the police, and so women may wish to report in order to see justice for their assaulter. Victims of rape that report the crime to the police are also 9 times more likely to receive medical care than those that do not (Resnick et al., 2000) as well as more likely to receive psychological care (Sable, Danis, Mauzy, & Gallagher, 2006) meaning that it is often in the interest of the victim's health for them to report the crime to police. There is also a benefit to society of reporting. Bachman (1998) notes that "an unreported incident of rape eliminates the possibility that an offender will be arrested or convicted. This may, in turn, reduce the perceived likelihood that rape and sexual assault, in general, will be punished." This reduction in perceived likelihood of punishment has significant consequences. A number of studies have found that the propensity of men to commit sexual assault is significantly decreased by the threat of formal sanctions such as arrest (Bachman, Paternoster, & Ward, 1992; Antunes & Hunt, 1974). Abel et al. (1987) finds that non-incarcerated rapists have high recidivism rates, suggesting that unreported assaults pose a serious threat to others. Finally, increased reporting can be a sign that the barriers to reporting that victims feel, barriers often steeped in misogyny, are declining in severity. This is surely a thing to strive for.

Previous research has focused on many factors affecting the reporting decision, including relations of the victim to the perpetrator, whether the act was violent and whether alcohol was involved. This research is discussed more thoroughly in the next section. One factor not previously investigated, however, is the response of reporting behaviour to news coverage of sexual assault allegations and cases. The #metoo movement that followed the Harvey Weinstein allegations focused on women coming forward with their sexual assault stories because they saw others come forward with theirs - thus the 'me too.' This idea highlights an important question: are victims of sexual assault encouraged to report to police or other authorities by coverage of other victims reporting?

In this paper, I explore these questions using incident-level FBI data of crime reports from 2008 to 2016, along with search-volume data from Google Trends. I supplement this with instrumental-variable analysis using a novel dataset of high-profile sexual assault allegations.

Literature Review

Since Becker outlined his economic model of crime, illicit activities have maintained a place in the economic literature. Sexual assault has received a share of this attention, although perhaps less than other crimes. One reason for this deficit is the difficulty present in gathering accurate data on sexual assault. Crime is under-reported in general, sexual assault especially (Fisher et al., 2000). This under-reporting makes the study of sexual assault inherently difficult, because no study will ever be able to accurately gauge how many assaults occurred over a given time period, only how many victims report. Studies that use surveys to estimate this under-reporting are always subject to under-reporting of their own, as even simply answering a survey question about sexual assault is difficult for many victims (Du Mont et al., 2003). When observing an increase in reports, the inability to observe assault numbers also creates issues of inference. Are reports increasing because people feel more safe reporting their assaults, or because assaults themselves are increasing? Different papers approach these issues in different ways when dealing with sexual assault.

Recently, several papers in economics have focused on sexual assault and harassment (Lindo, Siminski, & Swensen, 2018; Bisschop, Kastoryano, & van der Klaauw, 2017; Borker, 2018). Only one recent economics paper, however, focuses on the decision for a victim of sexual assault to report: Allen (2007) investigates the factors that influence an individual's decision to report a rape to law enforcement using survey data from the National Crime Victimization Survey and finds that 'social support' and ancillary evidence will increase the probability that the victim reports their crime. The victim being married, their offender being a stranger, their neighborhood being predominantly divorced households or households that have lived in the area for more than five years, and the victim's family income are all considered as factors that contribute to 'social support' for the victim, and each increase the propensity for a victim to report.

More attention is paid to the reporting decision by other fields, especially law and psychology. A number of studies find that a victim's propensity to report is increased when the offender uses physical force (Bachman, 1993; Du Mont et al., 2003). They believe this is likely because of both a desire to receive medical care - as seen above, those that report their assault to the police are much more likely to receive medical care than those that are not - and because physical violence often provides the victim evidence of the crime. Bachman (1993) also investigates whether the reporting decision is affected by the relation of the victim to the offender or by whether the event occurred at the victim's house and find no significant effects. All of these studies use the same National Crime Victimization Survey data for their studies, over different periods of time. Other survey-style studies exist too. Greenberg and Ruback (1992) use survey data from rape crisis centers and find that victims are more likely to report their sexual assault if the attack occurred outdoors, if the assailant was African American, or if "degrading acts" were part of the assault.

These studies have their limitations. Because they all use self-reported surveys, they suffer from selection bias, in that there will always be people that are asked but do not participate. There will also always be people that report things incorrectly. In the case of sexual assault surveys, there are very likely victims that do not report their assaults even to surveyors because doing so may be traumatic. This selection bias and misreporting is a problem that is hard to manage in a survey-style study, because there is no way to know if the characteristics of these people are the same as those that fill out the survey accurately. Consider an example. Suppose people who are violently assaulted but do not report their assault to police are ashamed of their non-reporting, and so when asked by surveyors they say they have never been assaulted. This behavior would undermine the results of Bachman (1993), which finds that victims report to police more often when their assault involves violence, because we cannot see the non-reporters of violent attacks. In addition to this selection bias, these studies tend to struggle to convincingly infer causality. As an example, Du Mont et al. (2003) concludes that women that visit the hospital after an assault are more likely to report their assault to the police. Resnick et al. (2000), using the same dataset, concludes that women that report their assault to the police are more likely to receive medical care. Because these studies use single time period survey data, there is no simple way to solve this issue of causality.

My paper aims to contribute to the literature by considering the effect that national coverage of sexual assault cases and allegations have on sexual assault reporting. This research question is interesting because it has not been approached before, but also because it lends itself to a different type of analysis: time-series and instrumental variable analysis of actual reports to police. This data would not be able to answer many of the questions above because it does not have the detailed information that the NCVS has on victims, nor does it have any indication of victims that do not report. However, for something like news coverage, which I measure at a national level and for which a number of exogenous shocks can be identified, the police data can work well. By using actual reports to police instead of reports and non-reports to police from survey data, selection bias is not an issue in the way that it is with survey data, as I have data on every report made to police in the given time period and state. I am also able to address causality more clearly, through instruments using exogenous shocks. There are clear limitations to my approach as well, which I discuss in my conclusion section, but it does avoid many of the problems discussed above.

This question of the effect that news coverage has is important. Most important is its sign, because a negative effect of news coverage could have serious policey implications. If research found that coverage of sexual assault cases had a significant depressing effect on the reporting of new sexual assaults, this might be cause for policy either within the media industry or from government to reduce such coverage. An allegory that comes to mind is the coverage of suicide cases. It is well established that media coverage of suicide tends to have a copycat effect (Stack, 2003). This is undesirable for a society that wishes to minimize loss of life, and so many countries such as New Zealand, Australia and Canada have strict guidelines on how the media can report on suicide. In the US, there are a number of organizations that advocate for reduced and more careful reporting of suicide[1]. If reporting on sexual assault produced the opposite of a copycat effect, stopping people from reporting their assaults, one might be inclined to recommend similar policy for the coverage of sexual assault cases. Moreover, these results should be interesting to providers of media coverage, as on the margin they should influence whether stories relating to sexual assault are published or not.

Previous economic research has used search volume from Google Trends in similar ways. There are many papers that use weighted indices of a number of words related to a subject, monetary policy for example, to forecast various statistics (Wohlfarth, 2018; Yu, Zhao, Tang, & Yang, 2019). More closely related to my analysis, however, is Berger, Chen, and Frey (2018), who investigates the effects that the introduction of Uber had on taxi drivers in different cities using Uber's staggered rollout as an instrument. They use search volume for "Uber" in different cities as a proxy variable for the intensity of Uber's rollout in that city to allow variable treatment effects of their instrument. This is very similar in methodology to my second instrumental variable specification discussed below.

Theoretical Framework

The decision for a victim to report can be considered as a probabilistic decision that factors in both costs and benefits of reporting (Allen, 2007). Our analysis then essentially is asking: does news coverage related to sexual assault increase

[1] `reportingonsuicide.org` is the most prominent of these organizations.

or decrease the perceived costs of reporting for a victim of sexual assault? If the effect we find is positive, it indicates that coverage related to sexual assaults decreases perceived costs of reporting and vice versa.

It is not clear what the sign of this effect should be expected to be. Coverage of sexual assault stories might decrease the social stigma that victims see associated with reporting a sexual assault, and thus ease the reporting decision for them. It might inspire victims to see other victims step forward. It might simply remind victims that they can report their assualt as a crime. These ideas are what lie behind the #metoo movement, where a high-profile set of allegations against Harvey Weinstein encouraged a number of other victims to come forward and discuss their assaults. However, we might expect negative effects too. Victims might see the way that prominent reporters of sexual assault are treated - for example, the flurry of derision directed at Christine Blasey Ford by those in even the highest seats of power after she came forward about her assault by Justice Kavanaugh - and conclude that the costs of reporting are higher than they initially thought. They might feel that reporting their assault in the wake of a high profile case could reduce the chance of people believing them. Thus there is no clear prediction for the sign of the measured effect.

The data that I use allows the examination of subgroups by age, race and by whether the incident involved alcohol, so it is worth considering how we might expect effects to vary across these categories. Regardless of the sign of effect, we might expect the magnitude to be greater for younger than for older victims for a number of reasons. It seems reasonable to assume that for any of the possible effects discussed above, magnitudes would diminish after the first time a victim saw coverage related to sexual assault - that is, one might be inspired more by the first news story they read about a sexual assault allegation than by the tenth. There is no clear prediction to be made about how the effect might vary across races. One might expect the effect to be more positive among incidents that don't involve alcohol than those that do, as in those events the victims generally have greater ancillary evidence (Allen, 2007), although this prediction is not very strong.

Data Summary

I have two main data sources for this project: The National Incident-Based Reporting System (NIBRS) and relative search volumes from Google Trends. The NIBRS has data on individual reports of crime to police stations, from 1991 to 2016. About 25% of the population is covered, as some states only report summary statistics while others do not give data at all. This percentage increased since 1991 as more stations have begun reporting. There are timestamps for both report and incident date and time, and the data includes some auxiliary information about the victim in question, including their age, race, and whether the incident involved alcohol. Because the data is by incident, it can be collapsed to any specification: nationally daily, by state, etc. The categories of crime that fall under the umbrella of sexual assault are "Forcible Rape," "Forcible Sodomy," "Sexual Assault With An Object," and "Forcible Fondling."[2] When collapsing, I remove from each day incidents that were both reported and occurred that day - i.e. I only keep reports of events from before the day in question. In this way, I hope to avoid findings that include shifts in assaulting behavior from coverage of sexual assault cases, and only include shifts in reporting behavior. A summary of this data by subgroup is shown in Table 1.

The data from Google Trends is relative search volume of a given term, scaled so that 100 is the highest volume over that time period. I collect data in six month periods, as this is required to get daily values, and then scale the data to the values in the first half of 2008 using values from overlapping dates. I use the

[2] These are offence codes 11A, 11B, 11C and 11D.

search term “sexual assault”[3]. Google Trends can provide data nationally or by state, and I use both in my analysis. I merge the two datasets by day for national data and by day and state for state data.

I consider the search volume data from Google Trends as a way to indirectly approximate the intensity of news coverage about sexual assault. High search volumes indicate a high desire for media related to sexual assault, which almost certainly means that there has been some national coverage of the issue already to spark the searches, and that there is or will be coverage supplied to match the high demand. Not only this, but high search volume is a sign of a high awareness of sexual assault coverage, which is arguably a more interesting variable than intensity of the coverage, as it shows that people are receiving the coverage, and thus if they are victims that it can influence their reporting decisions.

A significant issue with this type of study of sexual assault, moreso than survey-based studies, is that I have no sense of how numbers of actual assaults are changing over time. This could mean that an increase in reports of sexual assaults is being driven by an increase in the reporting of assaults or in the number of actual assaults. These two outcomes would have very different meanings and would lead to very different policy responses. It is thus very important to attempt to correct for this. I attempt to do this in my paper by dropping all events whose incident and report-dates are the same - that is, events that are reported on the day that they are committed. Using this measure of reports, if we see high search volume days having high numbers of reports, it is reasonable to assume that the increase is not being driven by increases in actual assaults (as these assaults would have had to happen during the high trend, i.e. on that day, and thus they would not have been counted). This is not a perfect correction. Most incidents

[3] I decided on “sexual assault” as “rape” tended to have a lot of unsavoury related searches, mostly pornography related, whereas searches for sexual assault tended to be related to cases of sexual assault. The two were highly correllated, but “rape”1 tended to be a lot noisier than “sexual assault” in general and especially around times of a high profile sexual assault case.

(74%) only have report dates, presumably because the victim wasn't sure or didn't want to report the date that the incident occurred on. Some of these may have occurred on the day they were reported, and so my strategy of deletion is likely not complete. Thus, an assumption of this paper is that the non-dated incidents are not systematically different from those that are dated, as the majority of dated incidents are reported after the day that they occur on, and thus my estimates should not include sizable changes in behavior of offenders and potential offenders. Another assumption is that if incident dates are misreported, they are not biased away from the report date - that is, victims do not systematically shift the incident dates they report earlier than the true incident dates. If this were not true, then there could be a number of reports included in my analysis of assaults that happened on the day they were reported, and this would hurt my ability to talk about reporting behaviors. These assumptions are not easily verifiable as true, but there is no good reason to suggest that they are not at least approximately true: there is no reason to think that reports with missing incident dates are inherently different than those with included incident dates, nor any reason to think that victims would systematically shift incident dates earlier than the true dates.

Empirical Methodology

A Time Series

In order to estimate the sign and magnitude of the population equivalent of the effect discussed above, I run a series of time series regressions and supplement these results with instrumental variable analysis. My initial regressions are time series regressions at the daily level, and are of the form:

$$reports_t = \beta_0 + \sum_{b=-7}^{7} \delta_b volume_{t+b} + \vec{\gamma_t} + \varepsilon_t$$

Where $reports_t$ is the natural log of the daily number of reports for the subgroup in question; $volume_{t+b}$ is the natural log of daily search volume for "sexual assault" along with a set of leads and lags, and γt is a vector including day-of-week, week-of-year and year fixed-effects[4]. These fixed effects should take care of most seasonality in the data. I include 7 leads and lags as I found this to be a good number to both account for events happening in close proximity to each other and to not reduce degrees of freedom unnecessarily. Changing the number of leads and lags did not affect my results significantly.

B Instrumental Variable Analysis

To accompany these time series results, I find events that are plausibly exogenous shocks to the volume of coverage of sexual assault-related topics, and use these shocks to instrument for the effect of an increase in such coverage on numbers of reports of sexual assault. These events are collected using Google Trend's "Related Queries" function, that collects searches that are made in conjunction with the term in question over a specified time period. I look at times at which the Google Trend for 'sexual assault' is above 60% of its 6-month maximum for the 9 years in question, which gives 563 days. The intuition here is that I want to find significant events that are nationally salient, and thus search volume must be relatively high. For each day, I look for distinct related queries. For example, on November 19, 2014, 'Bill Cosby' is the top related query, as he is for the next several days. I count only the first occurrence of these terms as a distinct 'high profile event.' Using this method, I find 37 events in the 9 year period in question.

[4]I use log form throughout my explanation as its results are more easily interpreted and extended than those of other functional forms. Running regressions without logs did not significantly change any results.

To use these events as instruments for the Google Trend, they must be correlated with the Google Trend and uncorrelated with the error term. I analyse the first requirement in the results section. The second requirement - that these events are uncorrelated with the error term - requires some qualitative justification, as components of the error term are inherently unobservable. This requirement can be reworded as: high profile events cannot affect report rates through any mechanism except through intensity of news coverage, here measured by search volume. It is possible that these events affect reporting in other ways, but I argue that it is not likely in any significant way. It is possible that these events themselves involve reports to police on the days of news coverage - i.e. that an accuser of a prominent case would report to the police and this would be counted in that day's reports - but this should not contribute in any significant way to total daily reports, given that more than 100 reports of rape are filed every day, and none of the events that I include involved many different reports at the same time. Any other channel would likely be caught in the proxy of search volume. Thus this assumption about the instrument seems reasonable.

I use these events as instruments in a number of two stage least squares regressions. To take advantage of the 3 day window of significant effects on search volume, shown below, and to reduce the standard errors of results, I bin data into the three day period including the event and two lag variables. I use two different instrumental variable specifications to obtain results. The first uses a binary variable *event_bin* as the instrument, which is equal to 1 if a high-profile event occurred in the past 3 days and 0 otherwise. The second uses a vector of dummy variables for each individual event's bin as the instrument. This means that while the first method obtains an estimate using the variation in search volume and reports between event- and non-event-dates, the second uses the variation between individual events by allowing for differing treatment affects. These are denoted IV1 and IV2 in my tables.

The first stage for the first IV specification is thus modeled as:

$$volume_t = \delta_0 + \delta_1 event\_bin_t + \epsilon_t$$

And for the second IV specification is modeled as:

$$volume_t = \delta_0 + \delta_1 events\_\vec{f}actor_t + \epsilon_t$$

Where $events\_factor$ is a vector of dummy variables for each individual event. The second stage for both regressions is then:

$$reports_t = \beta_0 + \beta_1 vol\hat{u}me_t + \vec{\gamma_t} + \varepsilon_t$$

Where $vol\hat{u}me_t$ is the predicted values of the given first stage and γ_t is a vector including week and day-of-week fixed effects[5].

C Panel Data

To consider variation by state, I run a panel regression of the form:

$$reports_{i,t} = \beta_0 + \sum_{b=-7}^{7} \delta_b volume_{i,t+b} + \vec{\alpha_i} + \vec{\gamma_t} + \varepsilon_{i,t}$$

Where $reports_{i,t}$ are reports in state i on date t; $volume_{i,t+b}$ is the relative search volume in state i on date t along with a set of leads and lags, α_i is a fixed effect at the level of the state, and γt is a vector including year fixed-effects,

[5]I use week fixed effects here instead of the year and week-of-year fixed effects used in the OLS specification as using these fixed effects introduced collinearity issues to both models. Week fixed effects should still control for seasonality.

day-of-week and week-of-year fixed effects.

D Arrest Data

Following Lindo et al. (2018), I examine whether probability of arrest changes for reports made in high-coverage periods, running the same time series and instrumental variable specifications as above but replacing the dependent variable with the percentage of reports from each day that resulted in arrest. This is an attempt to indirectly discern whether events that occur after during high-coverage periods are systematically different from those that occur during low-coverage period. It is also an attempt to test for a rise in false reporting - the idea that prominent sexual assault allegations might encourage people to report assaults that did not actually happen. It should be noted that a negative coefficient here does not necessarily indicate false reporting. It is possible that events for which the victim has a lot of evidence are reported more often than those that don't, as the findings of Allen (2007) indicate. In this case, it would be expected that reports that are inspired by coverage related to sexual assault would be less "low hanging," in terms of evidence, than those that are reported at other times, and so these reports might have lower arrest rates than during low coverage times, even without false reports.

Results and Discussion

I begin by running a time-series regression of reports of sexual assault to the FBI by report-date on national Google Trends for 'sexual assault.' The results of this are shown in Figure 1. There is a clear, statistically significant effect on the first lag variable. This indicates that news coverage likely has a positive effect on reports. It also helps to establish causality as these results show trend rising and then reports increasing: an effect on a lag variable would likely not be present if

Emphasize more uniqueness of paper and talk how compares to other results causality ran in the other direction. I show this effect, as well as the same effect broken down by subgroups of age and race of the victim and whether the event was reported as involving alcohol, in the first column of Table 2.

To complement this time series analysis, I perform the same subgroup analysis with two different instrumental variable specifications as discussed in the methodology section. To validate the instrument, I test the effect of these 'high profile events' on search volume by graphing the response of search volume before and after an event takes place, including fixed effects for year, week of year and day of week. This is shown in Figure 2. As can be seen, these events have sizable effects on search volume for 'sexual assault' that stays significant for the event date and the 2 days following, and thus that the instrument seems valid. To check whether reports seem to be impacted by these events occurring, I look at reports to police before and after these events in three day bins, including the same fixed effects as above. This is shown in Figure 3. We see that reports do indeed increase around the date of an event, albeit with large standard errors. The results of the instrumental variable regressions are shown in columns 2 and 3 of Table 2.

Does this match my predictions?

All three specifications give positive coefficients for the overall effect, ranging in magnitude from 0.0568 to 0.219. To contextualize this, the mean daily number of reports per day from 2008 to 2016 was 110. According to the low estimate, a 1% increase in the search volume for 'sexual assault' for one day leads to an increase of 0.062 reports, while the high estimate estimates this increase at 0.24 reports. Given that the NIBRS data only covers about 25% of the population of the US, if we assume a homogeneous effect across the country, these estimates become 0.25 and 0.96 reports respectively. The magnitude of these results is further discussed in the conclusion.

This increase appears to be driven by young reporters, as predicted above. Depending on the specification, there is also a significant effect (higher than the effect for young people) among reporters over 50, something non hypothesized. All specifications agree on only small positive or even some negative effects for victims between 20 and 50, although none of the effects are statistically significant. White reporters increase more than average under all specifications, and for the naive time series so do Black reporters, although the IV models fail to find significance here. Events with no alcohol involved increase at greater rates than average for all models, with events involving alcohol having null effects. There are, however, significant limitations with this sub-group analysis. As in the summary statistics table, most victims are young, white and reporting events that did not involve alcohol. It is perhaps not surprising, then, that these are the categories where we find statistical significance. Without more observations, it is difficult to say whether we are seeing differences in groups or simply large standard errors for subgroups with lower numbers of observations.

I next look at variation in effect by state, to try to determine whether local coverage or differences in local salience of events might affect reporting states with relatively higher trends on a given day have higher numbers of reports. I consider this using the panel regression outlined above. Results are shown in Table 3. I fail to find significant results, indicating that coverage at a more local level may not have the effects that national coverage does. Again, however, this may be a case of individual states not having enough observations to give significance.

To consider whether assaults reported during high-coverage times are different in nature to those reported during low-coverage times, I run the same three specifications as above, but use $percent\_arrest_t$ as the dependent variable, where $percent\_arrest_t$ is the percentage of incidents reported on date t that resulted in arrest. The results are shown in Table 4. All three specifications give estimates close to zero and fail to reject the null. This indicates that reports made during high-coverage times are not significantly more or less likely to result in arrest than those in low-coverage times. However, standard errors are large enough that we cannot rule out fairly large effects.

To give further context to these results, the increases in search volume depicted in Figure 2 give an average increase in search volume of 42% over the day of the event and the two days following. Given these increases, one of these events occurring results in a spike of between 31 and 121 reports of sexual assault nationally, according to my lowest and highest estimates respectively. For events with the largest effect, those where the google trend reached its 6-month maximum - the Bill Cosby case, for example - these estimates are 46 and 178 reports nationally.

It is worth noting that the spike in reports during high-coverage times are likely not all "new" reports, in the sense that news coverage may have caused people to report assaults earlier than they otherwise would have, but that still would still have reported in the counterfactual world without news coverage. These expedited reports are not as interesting as reports that would not have happened without coverage of sexual assault related news, as they do not represent an increase in total reports, but rather a change in distribution of when reports are made. Without data on non-reporters this distinction is hard to make.

Conclusion

My results indicate that reports of sexual assault increase substantially when there is significant national coverage of sexual assault-related news. These increases seem to come primarily from reporters younger than 20 reporting events that do not involve alcohol, with some evidence that reporters over 50 are also reporting more. I find no evidence that state-level coverage differences impact reporting. I find no evidence that reports made during high-coverage periods are more or less likely to end in arrest than those made during low-coverage periods, indicating that there are likely not a significant number of false reports.

These results indicate that policy should not seek to reduce the coverage of allegations or cases involving sexual assault. Moreover, they indicate that media providers should, at the margin, consider increasing the coverage they give to stories involving sexual assault. The results also add to the body of literature considering the choice of a victim of sexual assault to report their crime, given that the effect of media coverage on the reporting decision has not been investigated before. The methods that I use here also provide an example of a way to investigate the reporting decision without relying on self-reported survey data and the inherent flaws that that entails.

The data does not allow me to investigate the impact that this reporting has on actual assaults. This is an area that would be interesting to investigate in the future, as much of the impetus for increasing reports comes from the pretext that doing so should hopefully decrease assaults by increasing the likelihood of punishment for a sexual assaulter. The conditions that affect potential assaulters will always be a tougher area of research than the conditions that affect potential reporters, because most data on actual assaults must come one way or another from reporting of those assaults. Nonetheless, this is an area with important policy implications.

Figures and Tables

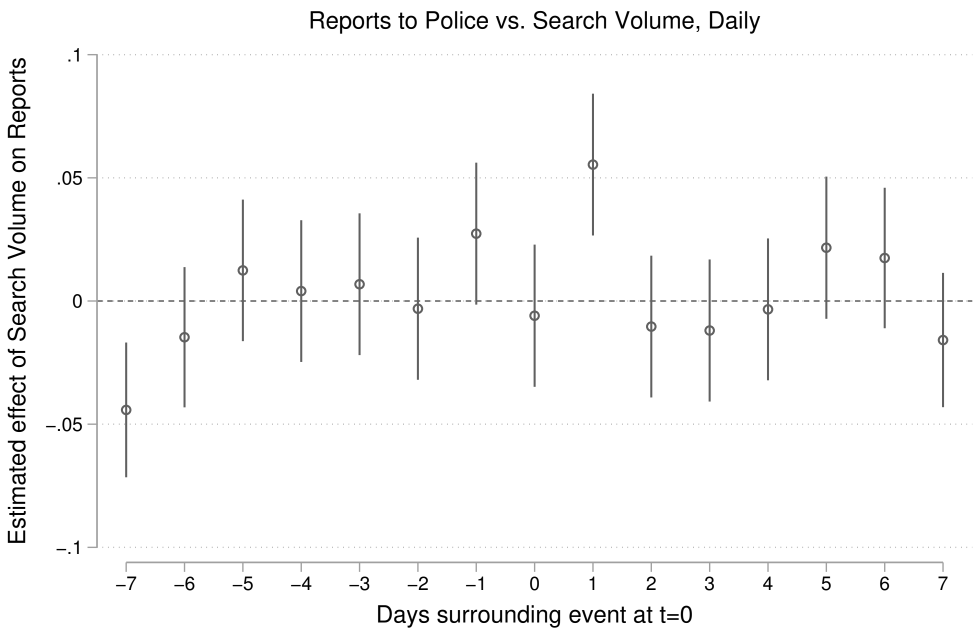

FIGURE 1. TIME-SERIES REGRESSION OF FBI REPORTS OF SEXUAL ASSAULT ON SEARCH VOLUME FOR 'SEXUAL ASSAULT', BOTH VARIABLES IN LOG FORM

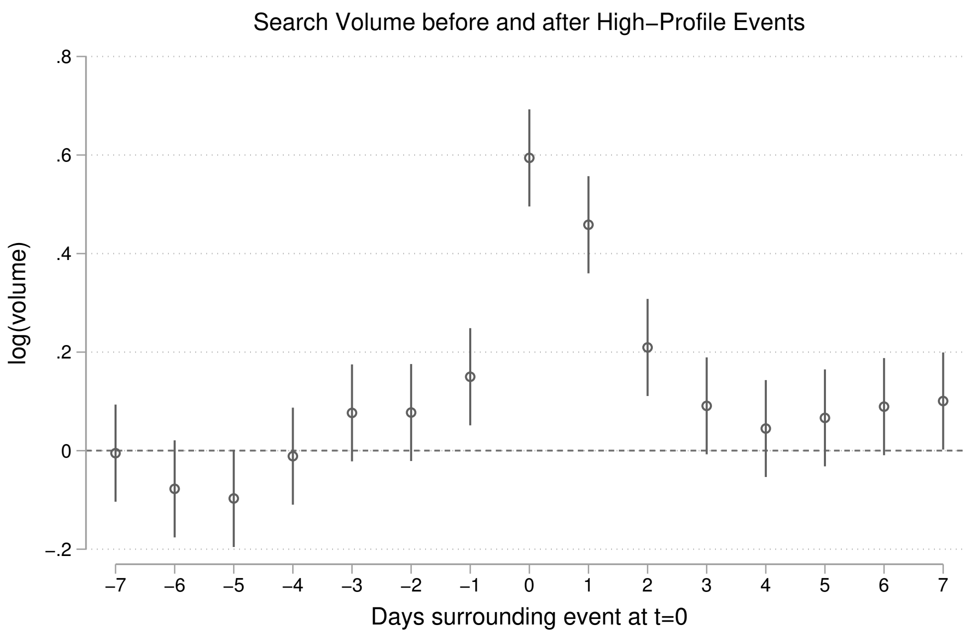

Figure 2. Search volume before and after high profile events, log form.

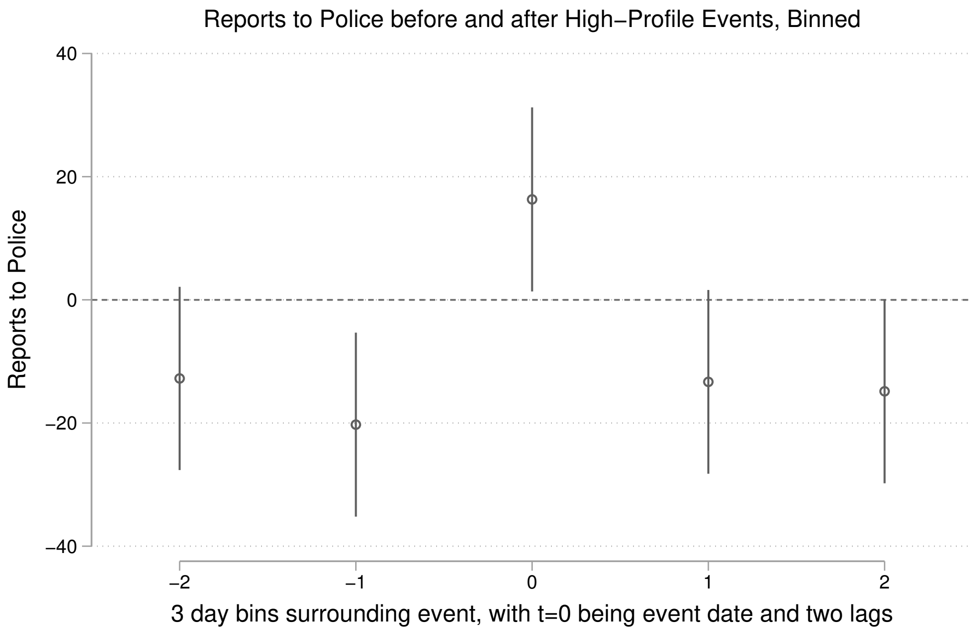

Figure 3. Reports to the FBI before and after High Profile Events, Binned

Table 1—Average daily reports of sexual assaults of different natures

	Reports	% of Total
Sexual Assault	110.5	100
Sexual Assault, victim age 10 to 19	48.3	43.7
Sexual Assault, victim age 20 to 29	23.4	21.2
Sexual Assault, victim age 30 to 39	11.0	10.0
Sexual Assault, victim age 40 to 49	6.5	5.9
Sexual Assault, victim age 50 to 59	2.8	2.5
Sexual Assault, victim age 60 to 69	0.73	0.66
Sexual Assault, white victim	83.2	75.3
Sexual Assault, black victim	21.4	19.4
Sexual Assault, victim of another race	5.9	5.3
Sexual Assault, alcohol involved	11.0	10.0
Sexual Assault, alcohol not involved	99.5	90.0
Sexual Assault, arrest made	19.9	18.0
Sexual Assault, no arrest made	90.6	82.0

Note: This is for the states that report to the NIBRS, which hold about 25% of the national population.

Table 2—Combined results of effect of increases in Google Trend on reports of sexual assault

	(OLS) log_reports	(IV1) log_reports	(IV2) log_reports
log_trend	0.0568***	0.219**	0.140**
	(0.0146)	(0.0777)	(0.0510)
log_trend_10_to_19	0.0611**	0.223*	0.121+
	(0.0190)	(0.0993)	(0.0655)
log_trend_20_to_29	0.0315	0.0814	0.0209
	(0.0196)	(0.0992)	(0.0714)
log_trend_30_to_39	0.0405	0.0794	0.103
	(0.0280)	(0.140)	(0.0952)
log_trend_40_to_49	-0.0324	-0.166	-0.195
	(0.0369)	(0.184)	(0.120)
log_trend_50_to_59	0.0408	0.452*	0.201
	(0.0446)	(0.222)	(0.146)
log_trend_60_to_69	0.0320	0.375*	0.201+
	(0.0468)	(0.182)	(0.112)
log_trend_white	0.0578***	0.228**	0.157**
	(0.0153)	(0.0812)	(0.0535)
log_trend_black	0.0645**	0.175	0.0861
	(0.0233)	(0.118)	(0.0776)
log_trend_other	0.0389	0.324	0.132
	(0.0408)	(0.207)	(0.135)
log_trend_alc	-0.0176	0.125	0.000938
	(0.0530)	(0.180)	(0.138)
log_trend_non_alc	0.0582***	0.225**	0.150**
	(0.0149)	(0.0795)	(0.0520)
N	3288	3289	3289
adj. R^2	0.273	.	0.190

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: N is for the first regression.

Table 3—Effect of increases in state Google Trend on reports of sexual assault

	(OLS)
	log_reports
log_trend	-0.00422
	(0.0289)
N	6676
adj. R^2	-0.300

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4—Effect of increases in Google Trend on percent of reports resulting in arrest

	(OLS)	(IV1)	(IV2)
	percent_arrest	percent_arrest	percent_arrest
log_trend	-0.0234	-0.0203	-0.00684
	(0.0143)	(0.0481)	(0.0343)
N	1461	1461	1461
adj. R^2	0.023	0.013	0.015

Standard errors in parentheses

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

References

Abel, G. G., Becker, J. V., Mittelman, M., Cunningham-Rathner, J., Rouleau, J. L., & Murphy, W. D. (1987, March). Self-Reported Sex Crimes of Nonincarcerated Paraphiliacs. *Journal of Interpersonal Violence*, *2*(1), 3–25.

Allen, W. D. (2007). The Reporting and Underreporting of Rape. *Southern Economic Journal*, *73*(3), 623–641.

Antunes, G., & Hunt, A. (1974, January). Impact of Certainty and Severity of Punishment on Levels of Crime in American States: An Extended Analysis, The. *Journal of Criminal Law and Criminology*, *64*(4), 486.

Bachman, R. (1993, September). Predicting the Reporting of Rape Victimizations: Have Rape Reforms Made a Difference? *Criminal Justice and Behavior*, *20*(3), 254–270.

Bachman, R. (1998, March). The Factors Related to Rape Reporting Behavior and Arrest: New Evidence from the National Crime Victimization Survey. *Criminal Justice and Behavior*, *25*(1), 8–29.

Bachman, R., Paternoster, R., & Ward, S. (1992). The Rationality of Sexual Offending: Testing a Deterrence/Rational Choice Conception of Sexual Assault. *Law & Society Review*, *26*(2), 343–372.

Berger, T., Chen, C., & Frey, C. B. (2018, November). Drivers of Disruption? Estimating the Uber Effect. *European Economic Review*, *110*, 197–210.

Bisschop, P., Kastoryano, S., & van der Klaauw, B. (2017, November). Street Prostitution Zones and Crime. *American Economic Journal: Economic Policy*, *9*(4), 28–63.

Black, M., Basile, K., Breiding, M., Smith, S., Walters, M., Merrick, M., … Stevens, M. (2011). National Intimate Partner and Sexual Violence Survey: 2010 Summary Report.

Borker, G. (2018). *Safety First: Perceived Risk of Street Harassment and Edu-*

cational Choices of Women (Job Market Paper).

Du Mont, J., Miller, K.-L., & Myhr, T. L. (2003, April). The Role of "Real Rape" and "Real Victim" Stereotypes in the Police Reporting Practices of Sexually Assaulted Women. *Violence Against Women*, *9*(4), 466–486.

Fisher, B. S., Cullen, F. T., & Turner, M. G. (2000, December). The Sexual Victimization of College Women. Research Report.

Greenberg, M. S., & Ruback, R. B. (1992). *After the Crime: Victim Decision Making*. Springer US. Retrieved 2019-04-08, from `https://www.springer.com/us/book/9781461364696`

Greenfield, L. A. (1997). *Sex Offenses and Offenders: An Analysis of Data on Rape and Sexual Assault* (Tech. Rep.). American Psychological Association.

Lindo, J. M., Siminski, P., & Swensen, I. D. (2018, January). College Party Culture and Sexual Assault. *American Economic Journal: Applied Economics*, *10*(1), 236–265.

Resnick, H. S., Holmes, M. M., Kilpatrick, D. G., Clum, G., Acierno, R., Best, C. L., & Saunders, B. E. (2000, November). Predictors of post-rape medical care in a national sample of women1. *American Journal of Preventive Medicine*, *19*(4), 214–219.

Sable, M. R., Danis, F., Mauzy, D. L., & Gallagher, S. K. (2006, December). Barriers to reporting sexual assault for women and men: perspectives of college students. *Journal of American college health: J of ACH*, *55*(3), 157–162.

Stack, S. (2003, April). Media coverage as a risk factor in suicide. *Journal of Epidemiology & Community Health*, *57*(4), 238–240.

Wohlfarth, P. (2018, December). Measuring the Impact of Monetary Policy Attention on Global Asset Volatility Using Search Data. *Economics Letters*, *173*, 15–18.

Yu, L., Zhao, Y., Tang, L., & Yang, Z. (2019, January). Online Big Data-Driven Oil Consumption Forecasting with Google Trends. *International Journal of*

Forecasting, *35*(1), 213–223.

Table 4—High Profile Events, collected using Google Related Trends on high-trend days

date	name	big_event
21/07/2009	Roethlisberger	
09/03/2010	Roethlisberger	
30/09/2010	MSU Athletes	
22/11/2010	Notre Dame Suicide	
16/02/2011	Lara Logan	
29/11/2011	Wellesley	
10/01/2012	Joe Philbin son	
01/04/2012	SA Aware Month	
15/05/2012	Prosper TX athlete	
18/10/2012	Amherst Document	
28/12/2012	Case McCoy	
19/01/2013	Michael Crabtree	
01/04/2013	SA Aware Month	
07/05/2013	USAF	
14/11/2013	Jameis Winston	1
01/05/2014	55 colleges sexual assault	
10/09/2014	Jerry Jones	
19/11/2014	Bill Cosby	1
27/02/2015	Scott Walker	1
25/03/2015	Lara Logan	
27/03/2015	Subway	
01/04/2015	Bikram Choudhury	
15/04/2015	Panama City	
21/05/2015	Josh Duggar	
21/09/2015	College Climate paper released	
09/11/2015	Biden Speech	
23/11/2015	Jameis Winston	
02/12/2015	James Deen	
30/12/2015	Bill Cosby	1
07/01/2016	Cologne New Year Assaults	
12/01/2016	David Bowie	
13/02/2016	Peyton Manning	1
29/02/2016	Lady Gaga	1
01/04/2016	SA Aware Month	
13/04/2016	Kobe Bryant	
10/06/2016	Stanford Student	1
08/10/2016	Donald Trump	1