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

By Harry Elworthy*

Draft: March 27, 2019

Your abstract here.

Files available at github.com/harryelworthy/Thesis

* Duke University, 837 Clarendon St Durham NC, harryelworthy@gmail.com.

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). At Duke, this figure is estimated at closer to 40%, as well as 10% of men (Fox, 2017). Despite this, very few assaults are reported to police for reasons including self-blame, guilt, fear of the perpetrator or fear of not being believed (Du Mont, Miller, & Myhr, 2003).

The Federal Government has in the past made an effort to increase reporting of sexual assaults, primarily at college campuses.

The benefits of reporting are many:

A number of different measures have been taken by many schools to address campus sexual assault, including significant federal reform via Title IX reform in 2011, which affected almost every school in the US. These changes aimed to make reporting easier for an assault victim, thus increasing reports and hopefully, in equilibrium, decreasing assault.

The issue of non-reporting has been especially salient over the past years, as first Harvey Weinstein, then Supreme Court nominee - now Justice - Brett Kavanaugh made national headlines being accused of sexual assaults that were not reported to police at the time they were committed . President Trump tweeted in response to the Kavanaugh claims: "I have no doubt that, if the attack on Dr. Ford was as bad as she says, charges would have been immediately filed with local Law Enforcement Authorities by either her or her loving parents" (2018). As above, there are many reasons why an individual may not report: social pressures, abusive relationships, and fear of not being believed, for example. More importantly, however, this tweet illustrates a source of motivation for this paper: had Dr. Ford reported when the crime had been committed, her testimony now would be more impervious to detractors. There are many reasons that increased reporting would be a desirable outcome - this is just one of them.

> Not completely sure this the case for Weinstein

> Need more on why reporting is desirable

The metoo movement that the Weinstein allegations started 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? Several other questions follow: if they are encouraged, what is the magnitude in the increase in reports? How long does the increase in reporting last for? How does local reporting affect reporting vs. national reporting? And - most importantly - does this coverage affect the behavior of potential perpetrators?

In this paper, I explore these questions using incident-level FBI data of crime reports from 1991 to 2016, along with data from Google Trends. I also conduct a number of event studies using Title IX investigations of universities and a novel dataset of high-profile sexual assault allegations.

A bit more summary here of data and methodology

I.   Background

The background section of (Lindo, Siminski, & Swensen, 2018) has a good background on sexual assault reporting, as do a number of my cited papers. I need to make my own ASAP! I've deleted all previously written stuff as my paper has pivoted away from it enough to make it pretty irrelevant. Here's what I'd like to talk about in my background section when I write it:

- One or two sentences about the multitude of work on prostitution/porn/sexual harrassment etc. by economists in the past two decades (many of the pieces in my Lit Review are like this)

- Background on history of economic papers on sexual assault reporting decision (Allen (2007) needs to be much more at the fore).

- Papers from other disciplines on the reporting decision - more research needs to be done here. Psychology/sociology. Would be good for talking about

cost of reporting later.

- Short history of high profile sexual assault cases. Run through most important ones from my list of events since 2008, plus some discussion of pre-2008.

- Some on metoo? Although not directly relevant to my investigation as such

- Much more on why reporting is so important - in itself and in what we hope to see in its outcomes, i.e. less assaults

I'd also like to include something like Figure 1 below. Note that this is not a great graph right now, as it shows a time when the number of police stations reporting these numbers was increasing. I need to make it per-capita, which will just take a re-jigging of my code that runs on the server.
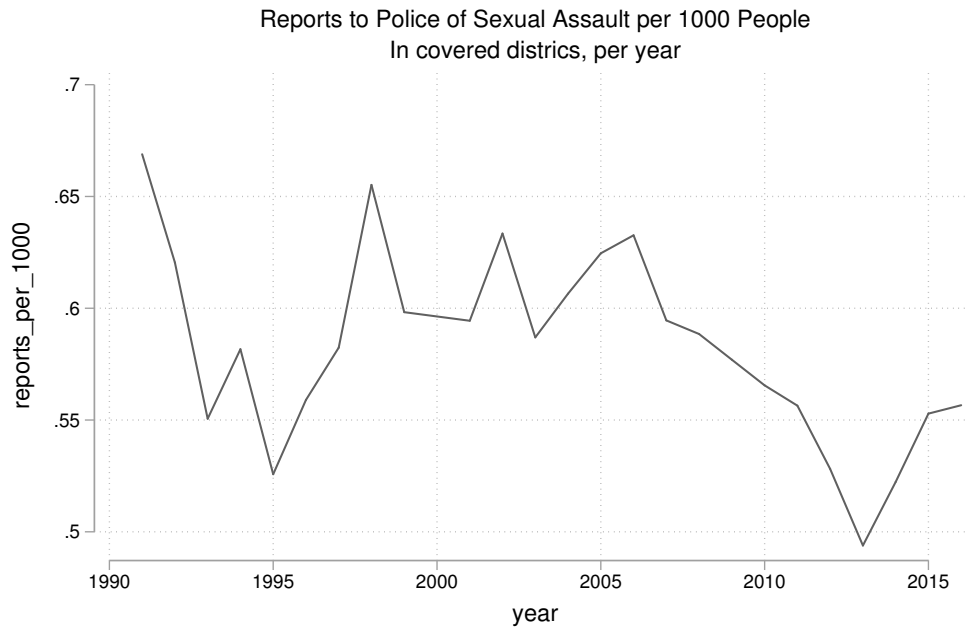

Figure 1. Reports to Police per Person, 1991 to 2015

This is obviously a big area of work for my paper. My literature review is also still (I know I know) not organized at all around my paper, and is instead just a list of summaries. This is also a big area of work. Because I've been shifting my research question around, some of these are now irrelevant while others have become much more important. This week my big task will be synthesizing the intro, my background and the lit review into one coherent package.

II. Literature Review

Since Becker outlined his economic model of crime, illicit activities have continued to have a place in the economic literature. Sexual assault has received a share of this attention, although perhaps less so than other crimes. One reason for this deficit is the difficulty of gathering accurate data on sexual assault. Crime is underreported in general, sexual assault especially (Kilpatrick, Resnick, Ruggiero, Conoscenti, & McCauley, 2007) (Fisher et al., 2000). Recently, however, several economics papers have focused on sexual assault and harassment.

Allen (2007) investigates the factors that influence an individual's decision to report a rape to law enforcement using survey data from The National Sample of Rape Victims, completed in 1985 and released in 2000, and finds that victims will be more likely to report sexual assault given more 'social support and ancillary evidence associated with the crime.' This paper is important to this investigation as it shows that the decision to report is influenced by factors that may be affected by policy, and thus supports the notion that policy to ease the reporting process could be beneficial. It is also some of the only research done on what affects the individual's decision to report. Given that I will be diving deeper into reasons this decision may have been made more often after 2011, this seems especially important to my investigation.

Yung (2015) investigates the idea that universities undercount reports in order to save face. Comparing report numbers from years before and after an audit by

the OCR, the paper estimates a consistent 40% uptick in reports by universities in the year of an audit, followed by a return to preexisting trends the year directly after an audit. This is relevant for my paper, as this undercounting could affect the accuracy of the CSS data I am using, although the homogeneity found in undercounting indicates that it should not lead to bad estimates.

Lindo, Siminski, and Swensen (2018) looks at the effect of partying culture on reports of sexual assault. Specifically, using the plausible exogeneity of Division 1 football games, the paper estimates the effects of increased partying that comes with such events on reports of rape to law enforcement. The authors find a 28% increase in rape reports associated with game days. Estimates are higher when the opponent is a rival, when the game is a Home rather than Away game, and when the game is televised. This paper is useful as it is a recent, high profile economics paper on the causes of sexual assault on college campuses, and for its use of NBIRS data, which will allow my paper to make use of this data much more easily.

Lindo, Marcotte, Palmer, and Swensen (2018) considers the effects of a Title IX investigation on a universities outcomes such as enrolment, applications, degree completion and donations. Their estimates show significant upticks in both enrolment and applications following an investigation for both men and women, with no evidence of effects on degree completion or donations. As part of their analysis, they use Google Trends data as a proxy for public awareness of investigations, concluding that the investigations are indeed in the public spotlight, even while federal policy on sexual assault may not be. The paper has an in-depth background of the 2011 Title IX changes that is very useful for my paper, as well as being closely related in subject.

Would be interesting to do this and see if the size of a school's increase in Reports has any effect on the size of the increase in enrollment

III. Data Summary

I have four main data sources for this project:

- National Incident-Based Reporting System (NIBRS) Data

Need to include sources for these, make this not a bullet point list, explain them a bit more in detail (I use on-campus reports etc.) but will wait to

- – Individual reports of crime to police stations. 1991 to 2016.

- – About 40% of population covered (some police stations dont report) and this number has increased since 1991 as more stations have begun reporting.

- – Timestamped, both report and incident datetime, lots of auxiliary information i.e. about the victim in question

- – Because is by incident, can be collapsed to any specification: Nationally Daily, State-by-Week, etc.

- Google Trends data

  - – Daily and weekly trends for sexual assault 2008 to 2018 [1]

  - – National and statewide trends.

  - – Relative trends out of 100, scaled to 2008 numbers, so some numbers later on are higher.

  - – Merge daily with police data for most of my investigation. Thus have reports of sexual assault grouped by either report or incident date together with Google Trends, daily, from 2008 to 2016.

- Related Events with Prominent News Coverage

  - – This is very recent, I need to go over it with my advisor etc, but I think it is useful.

  - – I've created a dataset of 35 big-headline sexual assault events from 2008 until 2016, along with the dates that they were first in the news.

  - – To do this I used Google's Related Topics tool. This tool shows for a given time period what related searches were to a given search.

> Can probably change this to 2004 as not using News, need to rerun

---

[1] 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. I test both for salience, and "rape" is not responsive to Title IX cases while "sexual assault" is. May try to add both back in.

- – I looked at related search terms to 'sexual assault' each day that the trend for 'sexual assault' was above 70% of its 6 month maximum. For events that had coverage for multiple days, I only included the first day. If there was more than a month between coverage I counted these as separate events.

  – The 35 events I found are shown in Appendix 1. I also categorize them into allegations and 'big allegations,' which are events that held the google trend above 75 for more than 3 days in a row.

  – I am sure that I've missed some events as my process could have been better, but each of these events is definitely a high profile sexual assault event. For an event study, it would definitely be better to get more events, but my results should be relatively good estimators as is (just with large standard errors from low n)

- Campus Safety and Security (CSS) Data

  – Collected and distributed by the Department of Education, 2005 to 2016

  – Reports of Sexual Assault by year by university, for all schools that receive financial aid (7663 schools that span the full time period)

  – Not granular at all - because of an internal standards change in 2013 for how to count sexual assault reports, have to include all assaults, including non-forcible ones/statutory assault/etc.

  – Can tie in a lot of auxiliary data by school ID from other sources, such as funding, SAT scores, enrollment by race/sex, etc.

- Title IX Cases

  – Opened when a student believes they were mistreated by a school's reporting system

  – Only began after 2011

- Data for each start/end date by school

- Used in Lindo, Marcotte, et al. (2018) to test effects on enrollment/applications/etc. They find increase in these factors, not decrease, even for women. They also find that case opening has sizable impact on google trends for [school name] rape so somewhat salient. I hope to expand this salience check.

- I am basing my panel data models off of theirs'

This section needs to be taken out of bullet form once I've settled on my final list of figures and tables. I'd also like to have a table of summary statistics here.

IV. Methodology

In the first half of this section I'll discuss sexual assault, both the crime and the reporting of it, from an economic standpoint. This is outlined below. Second half, I outline the regression equations I'll be using.

- Discussion of why someone might not report, fueled by Allen (2007) (where we see that a social safety net helps ease the reporting decision among other things) and Du Mont et al. (2003) as well as any other papers I find about this

- Pull that discussion into a more formal discussion of the costs and uncertainty that one faces in reporting, and how coverage of sexual assault might affect that in one way or another: by lowering or raising expected cost of social stigma, by inspiring and perhaps increasing expected benefit, the idea that reporting can help reduce sexual assault. More here, need to come up with as exhaustive a list as I can, as this is important.

- Then spend some time discussing a similar thing but for potential perpetators - expected cost of assault. Obviously even more than the other one this is a behaviour that is tough to rationalize, but it's not wild to think (and

one would seriously hope) that at the margin these people can be influenced one way or another

- Talk about the two in tandem, and again why hopefully reporting affects the second, and thus is important. If possible, this link would be great to try to estimate, but very difficult given the nature of the data.

In this paper, I'll run a number of time-series, panel-data and event-study type regressions. The general form of these regression equations is outlined below. My time-series regressions are at the daily level, and are of the form:

$$y_t = \beta_0 + \sum_{b=-7}^{7} \delta_b x_{t+b} + \gamma_t + \varepsilon_t$$

Where y_t is the outcome variable in question; x_{t+b} is the independent variable in question along with a set of leads and lags, and γt is a vector including day-of-week, week-of-year and year fixed-effects. These fixed effects should take care of most seasonality in the data.

My panel data regressions are of the form:

$$y_{i,t} = \beta_0 + \sum_{b=-7}^{7} \delta_b x_{i,t+b} + \alpha_i + \gamma_t + \varepsilon_{i,t}$$

Where $y_{i,t}$ is the outcome variable in question; $x_{i,t+b}$ is the independent variable in question along with a set of leads and lags, α_i is a fixed effect at the level of the panel, usually by state or by school, and γt is a vector including year fixed-effects, as well as day-of-week and week-of-year fixed effects if the data is at the daily level.

My event study regressions are of the form:

$$y_t = \beta_0 + \sum_{b=-7}^{7} \delta_b x_{t+b} + \gamma_t + \varepsilon_t$$

Where y_t is the outcome variable in question; x_{t+b} is a dummy for the event in

> I need to decide on a good reason for how many leads and lags I will be including, then put this reasoning here. Advisor also said that in final form, i.e. for tables rather than graphing, I should drop leads from this.

question along with a set of leads and lags, and γt is a vector including day-of-week, week-of-year and year fixed-effects.

V. Results

In general, I may drop uninteresting things from this section as I decide exactly on the narrative on my paper. I'll also include more tables in the final paper (and probably less graphs) but graphs are just nicer to glance through.

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 2

> I think this could maybe be split into primary and secondary results, or 'evidence from schools' or something, but need to work out narrative better first

Reports to Police vs. Trends, Daily
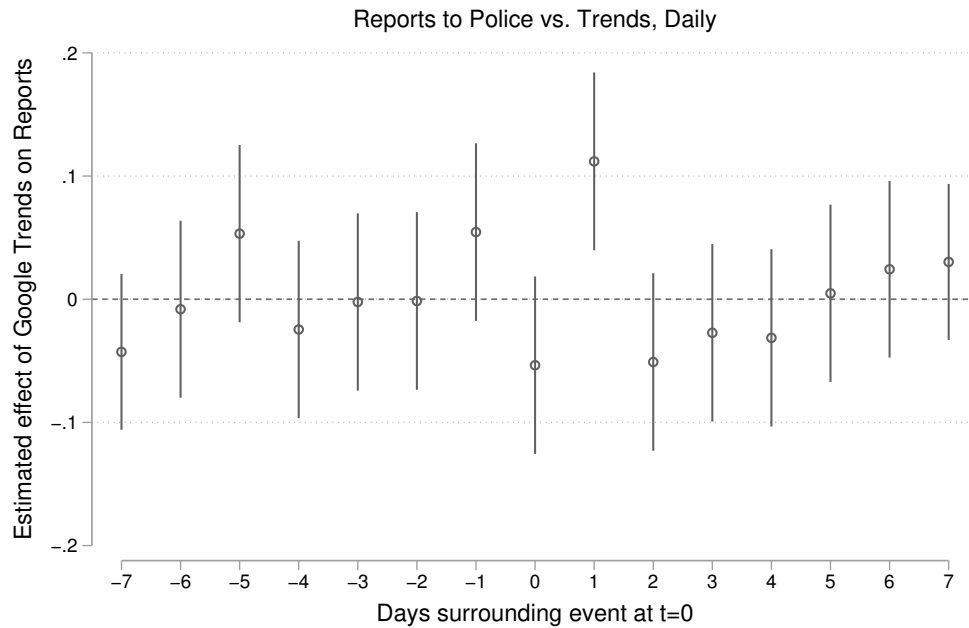
FIGURE 2. TIME-SERIES REGRESSION OF FBI REPORTS OF SEXUAL ASSAULT ON GOOGLE TRENDS FOR 'SEXUAL ASSAULT'

There is a clear, statistically significant effect on the first lead variable. I discuss

> I think I worked out that I had miscoded this, and that it was on the first lag variable. Need to go over and fix. Also fix x axis.

interpretations of the effect magnitude below.

This increase in reporting is driven by white reporters under 20 and to a lesser extent under 30, reporting incidents that did not involve alcohol, as shown in Figure 3, Figure 4, and Figure 5:
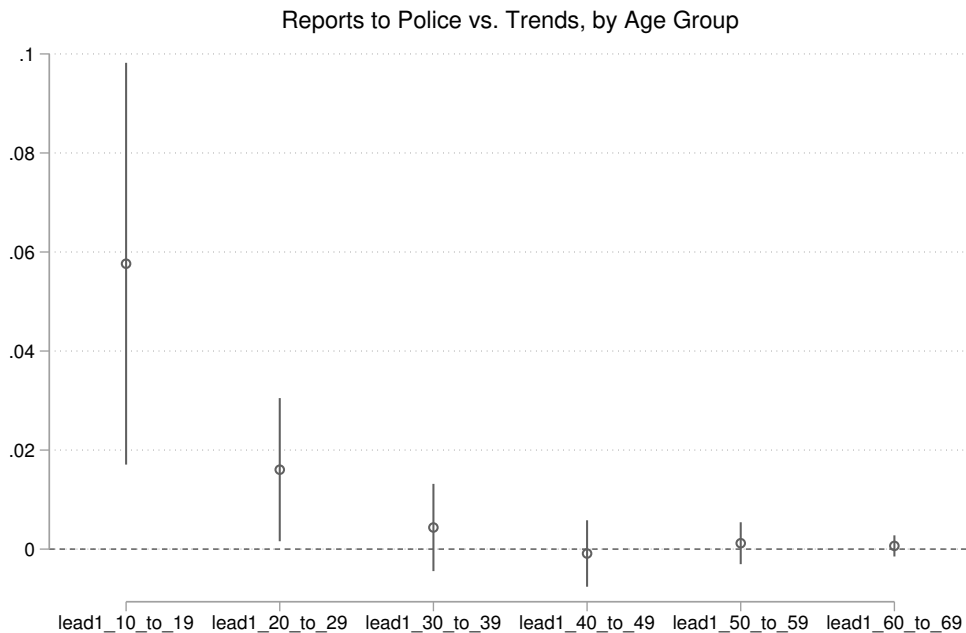
Don't actually show below yet. Probably use the Bill Cosby case as it's within sample, maybe also with a smaller case, as well as all 35 of my events. Show their graphs, eyeball approximate estimates, for the full 35 event sample calculate average effect.

FIGURE 3. ESTIMATES OF EFFECTS OF GOOGLE TRENDS ON FBI REPORTS OF SEXUAL ASSAULT BY AGE GROUP

Here I'd look at how state by state variation in trends impacts reporting across states, but that regression has been giving me grief and is not presentable. I think it's going to be a small or 0 effect.

I now turn to an event study of 35 high-profile sexual-assault related stories from 2008 to 2016. To check that these events were indeed high-profile, I run an event-study regression of Google Trends on each of these event dates, results shown in Figure 6:

As can be seen, these events have sizable effects on the google trend for sexual

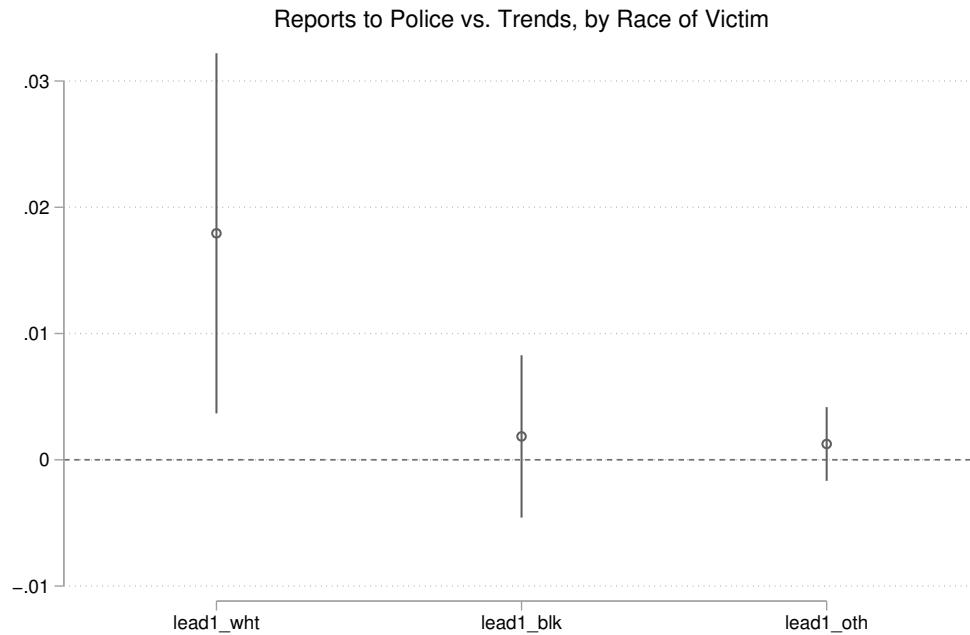Reports to Police vs. Trends, by Race of Victim

FIGURE 4. ESTIMATES OF EFFECTS OF GOOGLE TRENDS ON FBI REPORTS OF SEXUAL ASSAULT BY RACE

assault that last for about 3 days. Thus they look to be good examples of random positive shocks to the google trend.

Should explain this more

I now look at the effects these events have on reports to the FBI. The results of a simple event-study regression are shown in Figure 7:

There looks to be a possible effect, but it's not showing as significant on any individual day. To try to fix this, I bin the results in 3 day groups (from the 3 days that they impact google trends) in an effort to reduce standard errors. Results are shown in Figure 8:

There is still not a significant coefficient (p value is 0.15). I need to think about this and maybe bin more. Regardless, we have what looks to be a positive coefficient but SE's that are too large to confirm it. More events (higher n) would help with this.

I also estimate the effects of just events related to sexual assault allegations in

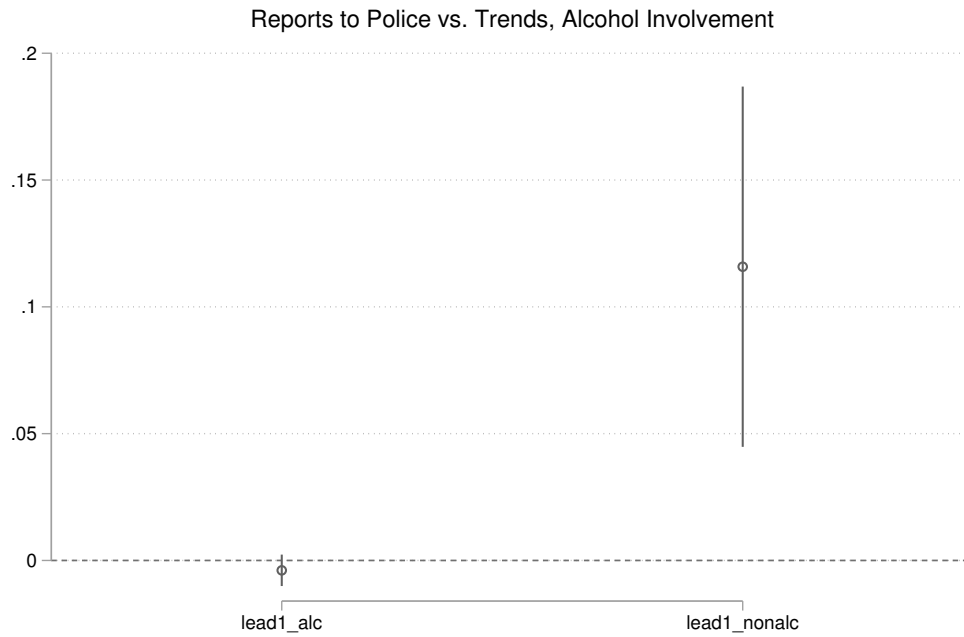Reports to Police vs. Trends, Alcohol Involvement

FIGURE 5. ESTIMATES OF EFFECTS OF GOOGLE TRENDS ON FBI REPORTS OF SEXUAL ASSAULT BY ALCOHOL INVOLVEMENT

Figure 9, and the effects of 'big allegations' in Figure 10. Both of these figures are in 3-day bins to get consistent effect estimates.

We see also see no significant individual days after allegations. After 'big allegations,' we see a large, statistically significant increase in reports 2 days after the event. A spike in 30 reports after each of these high profile cases is huge.
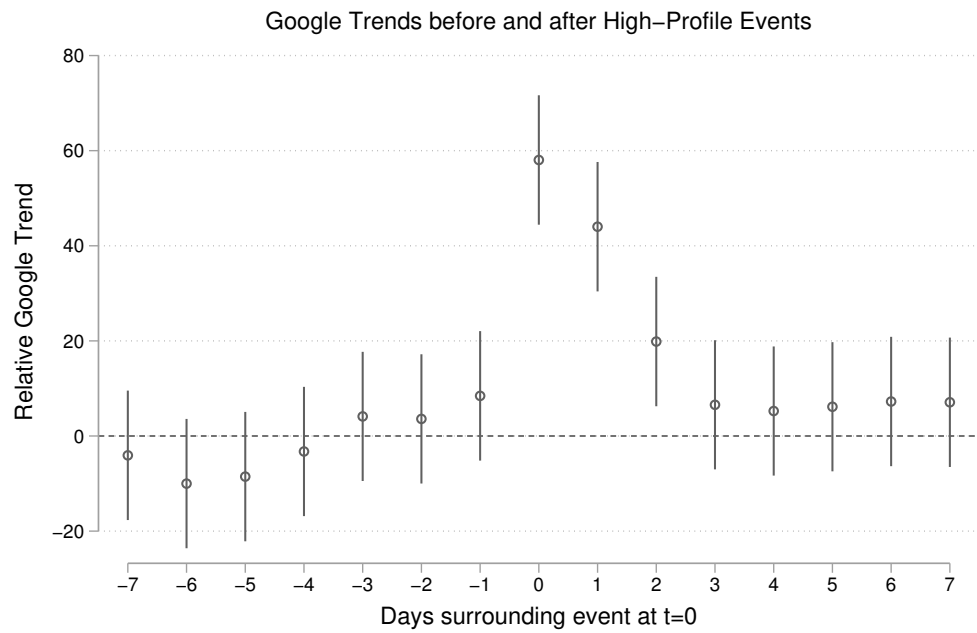
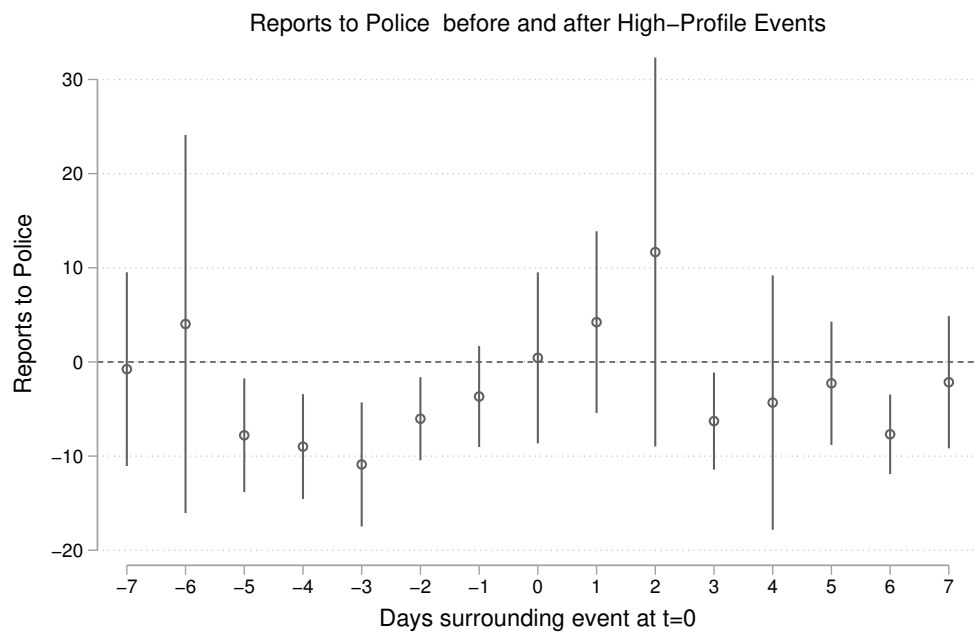FIGURE 6. GOOGLE TRENDS BEFORE AND AFTER HIGH PROFILE EVENTS

Reports to Police before and after High−Profile Events

FIGURE 7. REPORTS TO THE FBI BEFORE AND AFTER HIGH PROFILE EVENTS

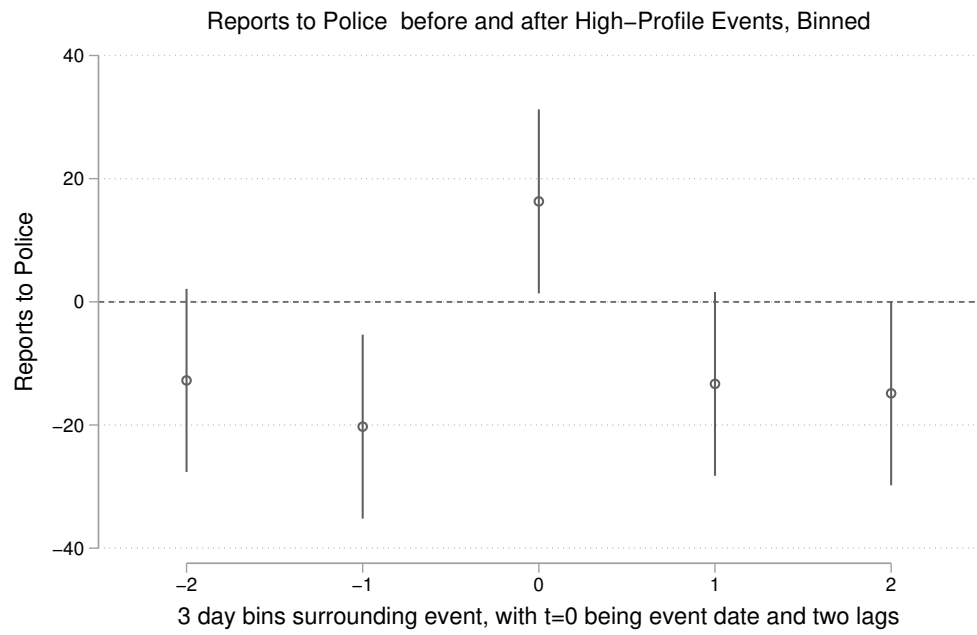Reports to Police before and after High−Profile Events, Binned

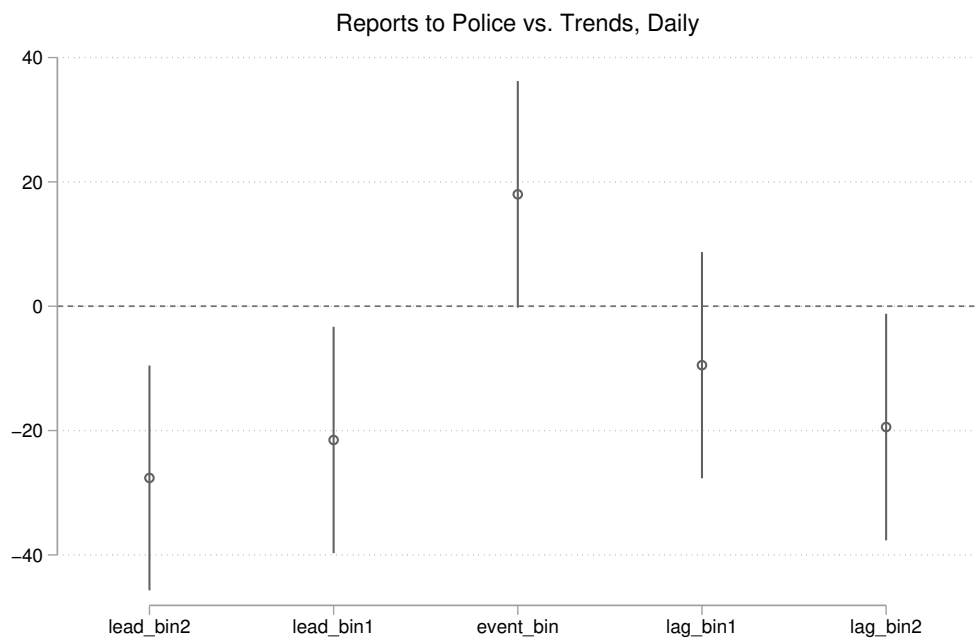FIGURE 8. REPORTS TO THE FBI BEFORE AND AFTER HIGH PROFILE EVENTS, BINNED

FIGURE 9. REPORTS TO THE FBI BEFORE AND AFTER ALLEGATIONS OF SEXUAL ASSAULT
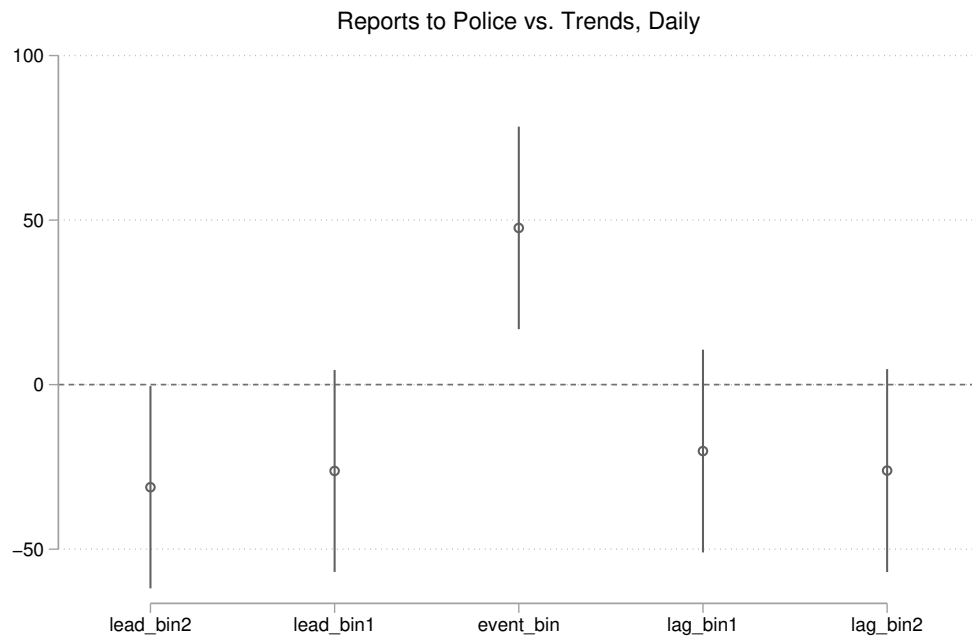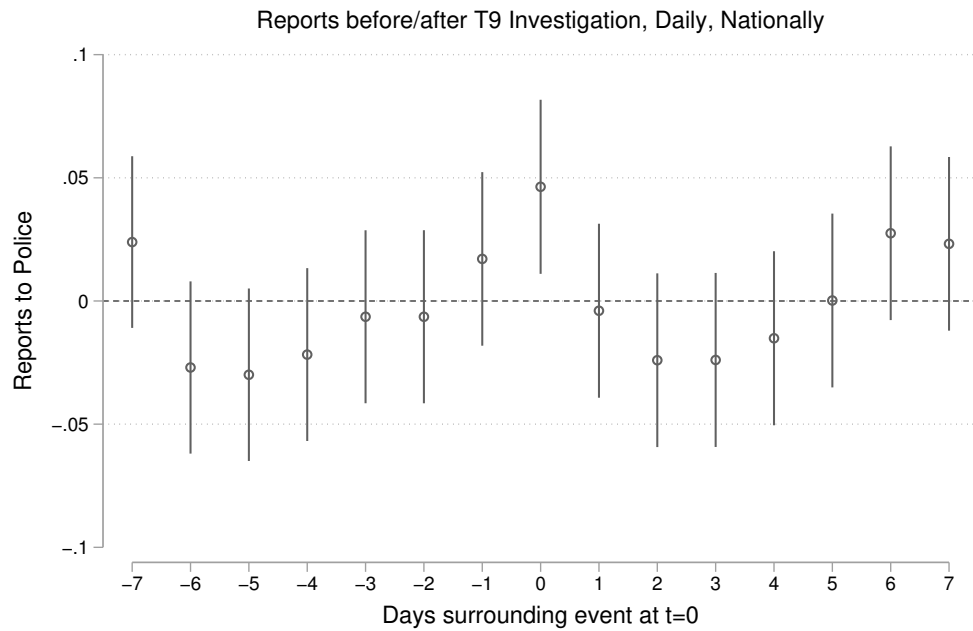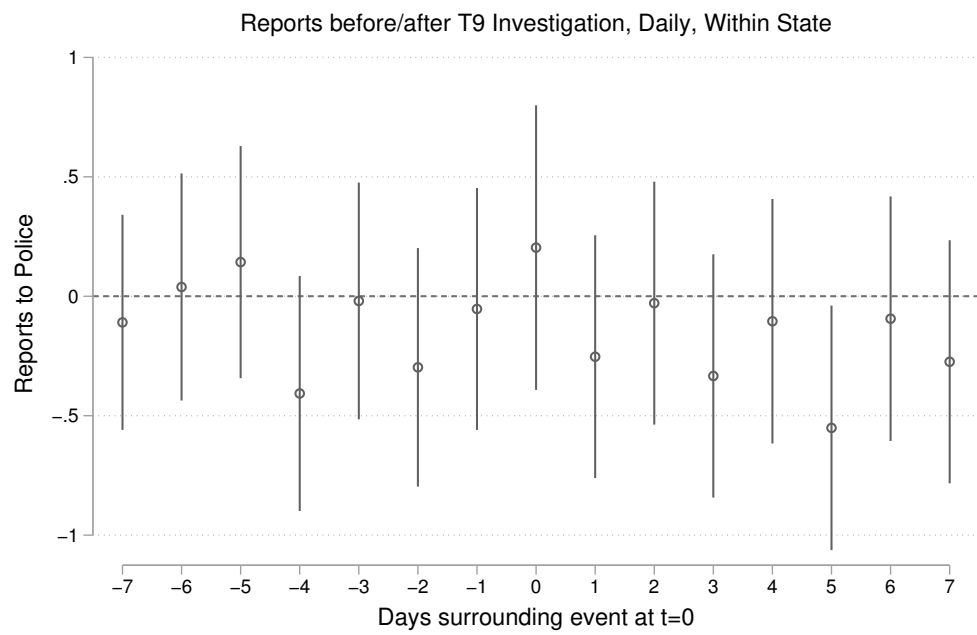
FIGURE 10. REPORTS TO THE FBI BEFORE AND AFTER 'BIG ALLEGATIONS'

VI.   Extra figures and results

The main extra result that needs to get added in is that IV analysis using the events gives a result for the first estimate (of reports on google trends) that is very close to the estimate of the time series, with p¡0.05. Need to talk with PB about what exactly this means and how to integrate it. Probably lose a lot of the graphs above and have a couple of tables with lots of results.

Below I look at national and state-level daily trends before/after T9 cases get opened (in those states). Not sure how to interpret this - the cases don't influence national trends and yet have a single significant coefficient, which is more than the 35 events I have can say (although perhaps it is significant because there are 300 instead of 35 events). The national trend also looks weirdly like it doesn't have DOW fixed effects (though it does).

Reports before/after T9 Investigation, Daily, Nationally

Reports before/after T9 Investigation, Daily, Within State

Below I look at incident dates of incidents reported within a month after a high profile event, compared to the same for a sample of placebo dates. Trying to work out how to interpret this, and add error bars.
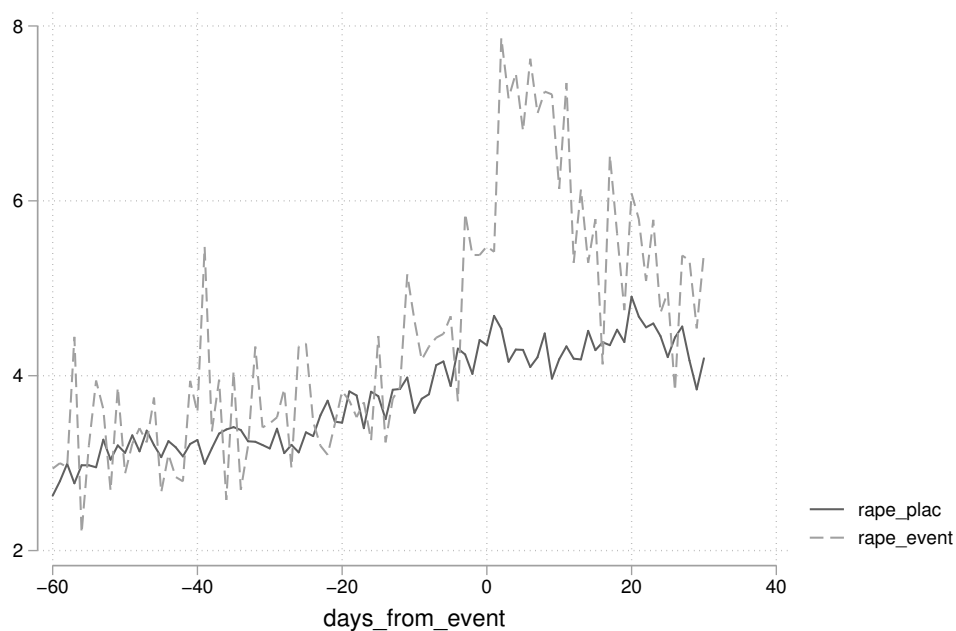
Bring up school reporting again? Don't think so. If I did my narrative would go something like this:

- Maybe we also see effects for much more local cases

- Potential evidencefor this: school reporting after title IX cases

- Show huge increase in reports at schools after title IX cases get opened there, explain that the cases are very salient as shown in Lindo, Marcotte, et al. (2018)

- However, looking closer this doesn't seem to hold up especially well

- Figure showing that effect is only on the first case, not subsequent ones

- Figures showing no increase in nearby schools or police stations (still need to rerun this, especially daily police) - maybe there is a small increase

If not, should have somewhere that school data was not feasible for x reasons (?)

VII. Discussion

TBD

VIII.   To Discuss

- IDT analysis graph - hopefully. Otherwise methodology of it. Probably just qualitative. I'm found of using it as a way to discuss possible next steps.

- Talk about local/national T9 results with P Bayer. Bad news for my other results - other mechanisms? Although not the same type of event

- Talk about methodology on trend. Log? Original Trend?

- Best tables? Just one with overall effect, subgroups, columns are naive, instrument, log, log instrument? or something?

- Tables/Figures draft approx sequence

- Should make numbers national equivalent in tables?

- Should I run with year interactions/something to check changing effect? Not that many years?

- IRF? Not sure I see what would be different than the graphs I have already

- Intro/Background/how to break up? How long intro?

- Summary statistics? What would be interesting?

IX. Next Steps

- Run state fixed effects with weights by population (once decided) as well as state cases

- Finish idt_analysis. Graph

- Do same for 50/100/200 random generated time segments depending on time the above takes, compare shapes

- Check results on population1 - use if not ridiculous

- Calculate average event effect, equivalent to the US. Note the assumption in homogeneous effect.

- Find any more

- Write a paragraph on future work. More on perpetrator behavior. More on whether increased reports are true new reports or just earlier reports.

- Put in methodology/somewhere about how this may be people reporting more quickly, not new reports. Assume some are new at least.

- Include state panel data with date fixed effects somewhere - null

- Look for more high-profile cases for event study. Go back before 2008, re-go-over already done dates

References

Allen, W. D. (2007). The Reporting and Underreporting of Rape. *Southern Economic Journal*, *73*(3), 623–641. Retrieved from `https://www.jstor.org/stable/20111915` doi: 10.2307/20111915

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. Retrieved 2019-03-18, from `https://ncvc.dspacedirect.org/handle/20.500.11990/250`

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. Retrieved 2019-03-18, from `https://doi.org/10.1177/1077801202250960` doi: 10.1177/1077801202250960

Fisher, B. S., Cullen, F. T., & Turner, M. G. (2000, December). The Sexual Victimization of College Women. Research Report. Retrieved 2018-09-08, from `https://eric.ed.gov/?id=ED449712`

Fox, K. (2017, March). University survey reveals 'startling' sexual misconduct statistics. *The Chronicle*.

Kilpatrick, D. G., Resnick, H. S., Ruggiero, K. J., Conoscenti, L. M., & McCauley, J. (2007, February). *Drug-facilitated, Incapacitated, and Forcible Rape: A National Study* (Tech. Rep.). US Department of Justice Office of Justice Programs United States of America. Retrieved from `https://www.ncjrs.gov/App/Publications/abstract.aspx?ID=240972`

Lindo, J. M., Marcotte, D. E., Palmer, J. E., & Swensen, I. D. (2018, July). *Any Press is Good Press? The Unanticipated Effects of Title IX Investigations on University Outcomes* (Working Paper No. 24852). National Bureau of Economic Research. Retrieved from `http://www.nber.org/papers/w24852` doi: 10.3386/w24852

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. Retrieved 2018-09-08, from `https://www.aeaweb.org/articles?id=10.1257/app.20160031` doi: 10.1257/app.20160031

Trump, D. J. (2018, September). *I have no doubt that, if the attack on Dr. Ford was as bad as she says, charges would have been immediately filed with local Law Enforcement Authorities by either her or her loving parents. I ask that she bring those filings forward so that we can learn date, time, and place!* [Tweet]. Retrieved from

Yung, C. R. (2015). Concealing campus sexual assault: An empirical examination. *Psychology, Public Policy, and Law*, *21*(1), 1–9. Retrieved 2018-09-18, from `https://www.apa.org/pubs/journals/releases/law-0000037.pdf` doi: 10.1037/law0000037

Table 1—High Profile Events, collected using Google Related Trends on high-trend days

name	date	allegation	big_allegations
Roethlisberger	18099	1	.
Roethlisberger second	18330	1	.
MSU Athletes	18535	1	.
Notre Dame Suicide after reporting SA	18588	.	.
Lara Logan assaulted by mob in Egypt	18674	.	.
Police investigate SA in Wellesley	18960	.	.
Joe Philbin son	19002	1	.
Prosper TX athlete	19128	1	.
Amherst Document	19284	.	.
Case McCoy	19355	1	.
Michael Crabtree	19377	1	.
USAF SA case	19485	.	.
Jameis Winston	19676	1	1
55 colleges sexual assault	19844	.	.
Jerry Jones/Ron Washington	19976	1	.
Bill Cosby	20046	1	1
Scott Walker	20146	1	1
Lara Logan hospitalized again	20172	.	.
Subway (first fogle stuff?)	20174	1	.
Bikram Choudhury Yoga	20179	1	.
Panama City Rape	20193	.	.
Josh Duggar	20229	1	.
College Climate paper released	20352	.	.
Biden talk	20401	.	.
Jameis Winston again (?)	20415	1	.
James Deen	20424	1	.
Bill Cosby again	20452	1	1
Cologne New Year Assaults	20460	.	.
David Bowie	20465	1	.
Peyton Manning	20497	1	1
Lady Gaga	20513	1	1
Kobe Bryant	20557	1	.
Stanford Student	20615	1	1
Trump/Bill Clinton	20735	1	1
Casey Affleck	20877	1	.

States included in FBI reporting