The Impact of #MeToo on Sexual Criminality

Germain Gauthier

CREST - Ecole Polytechnique

September 23, 2020

Sex Crimes in the US

- Sexual criminality ranges from misdemeanors such as sexual harassment to extreme felonies such as rape.
- Major public health issue, with potentially more pervasive consequences in the workplace.
- The US Common Law penalizes many forms of sexual violence.
  - Since 1800s: rape as a felony
  - Since 1980s: sexual harassment as sex discrimination
  - The scope of sexual crimes has been broadened multiple times (e.g. marital rape in 1993).
- And yet sexual crimes are still widely prevalent.
  - $\approx$ 298,000 victims of rape and sexual assault (NCVS, 2016)
  - 6758 complaints for sexual harassment (EEOC, 2016)

Imperfect Monitoring, Shaky Empirics

- Law enforcement agencies imperfectly monitor sex crimes.
- Survey evidence suggests many crimes go unreported to the police.
  - $\approx$ 20-40% of sexual assaults are reported to the police (NCVS)
- Consequences for public policy research
  - $\rightarrow$ Researchers and public officials work with a selected sample of crimes.
  - $\rightarrow$ Complicates impact evaluations of interventions aimed at fighting crime.

The Me Too Movement

- October 2017: In the wake of the Weinstein affair, Alyssa Milano tweets #MeToo.
- Over the next months, millions of women protest against sexual harassment and sexual assault on social media.
- The movement explicitly aimed at:
  - Empowering victims (#MeToo)
  - Deterring offenders (#TimesUp)

How successful was Me Too at changing victim and offender behaviors?

#MeToo in the US

This Paper

- **Data**
  - Incident-level police data for five US cities (2003-2020)
- **Methodology**
  - Clarify econometric issues related to police data
  - Propose a novel empirical strategy to disentangle the crime rate from reporting behaviors of victims
- **Empirics**
  - Event-study on victim and offender behaviors for sexual felonies

Preview of Results

- **Methodology**
  - Strong assumptions in many applications:
    1. Treatment only impacts crime rates or reporting behaviors.
    2. No lagged reporting
  - Based on variations in lagged reports, we can infer variations in reporting behaviors.
- **Empirics**
  - Evidence of an increase in reporting behaviors
  - Evidence of a deterrent effect

Related Literature

- **Sexual Violence**
  - Basu (2003); Bhatnagar et al. (2019); Lee & Suen (2019); Levy & Mattsson (2019);
- **Econometrics of crime**
  - Coleman and Moynihan (1996); Durlauf, Navarro & Rivers (2010); Aizer(2010); Stephens-Davidowitz (2013); Bellego & Drouard (2019)
- **Duration models**
  - Van den Berg (2001); Abbring & Van den Berg (2001); Dörre & Emura (2019)
- **Crime deterrence**
  - Drago et al. (2009); Doleac (2019)
- **Social norms**
  - Benabou & Tirole (2011); Young (2015); Bursztyn et al. (2019)

Plan

Data

- Incident-level police records for five US cities:
  New York, Los Angeles, Austin, Nashville, Cincinnati
- Unbalanced panel from 2003 to 2020
- Some relevant variables: socio-demographic characteristics of victims and offenders, granular crime categories.
- Importantly, both the date of the incident and the date of its report to the police are recorded.

→ Let us first understand the specifics of police data and their implications.

A Simple Econometric Setup

- A researcher wishes to assess the impact of a treatment D_t on effective crimes C_t, but only observes reported crimes R_t:

$$R_t = \beta_0 + \beta_1 D_t + \varepsilon_t \tag{1}$$

- But a share of crimes is never reported to the police. Denote r_t the crime reporting rate.

- Bellego and Drouard (2019) suggest:

$$r_t \cdot C_t = \beta_0 + \beta_1 D_t + \varepsilon_t \tag{2}$$

$\rightarrow$ If r and C are correlated to D, conclusions will be unclear.

Common Assumptions and Work-arounds

"All we possess of statistics of crime and misdemeanors would have no utility at all if we did not tacitly assume that there is a nearly invariable relationship between offenses known and adjudicated and the total unknown sum of offenses committed."
(Adolphe Quêtelet)

- Commonly made assumption: either r or C is orthogonal to D.
- Some studies work with proxy variables to infer variations in r or C (e.g. Google trends, emergency records, other crimes).
- Some studies work with victimization surveys, but they are also subject to well-documented biases.

▸ NCVS Estimates

Lagged Reporting

- No approach takes into account the existence of lagged reports.
- Yet a sizable share of crimes is reported with a lag relative to the date of the incident.
- In my data: 66% of sexual crimes and 15% of non-sexual crimes are lagged reports.

→ An additional problem or a handy solution?

Lagged Reporting

- Denote τ_1 and τ_2 respectively the first and last calendar date of data collection. At each period k, victims choose to report the crime to the police or to abstain with probability $P(k|t)$.
- If t is the date of the incident (e.g. NIBRS):

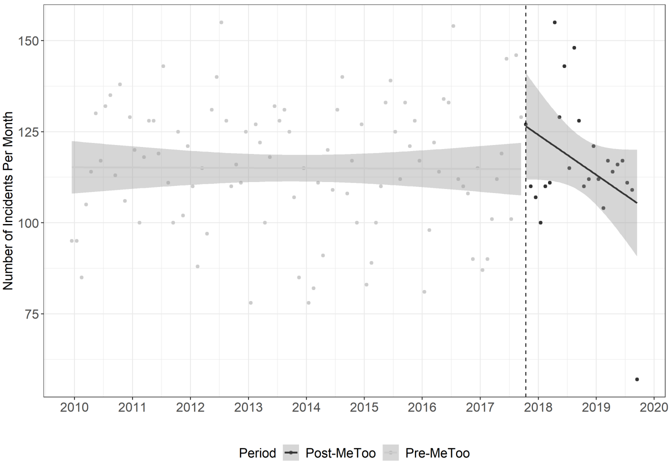
$$\sum_{k=t-\tau_1}^{\tau_2-t} R_{t,k} = \sum_{k=t-\tau_1}^{\tau_2-t} P(k|t)\, C_t = \beta_0 + \beta_1 D_t + \varepsilon_t \tag{3}$$

- If t is the date of the report (e.g. UCR):

$$\sum_{j\in[\tau_1,\tau_2]} R_{j,k} 1_{t=j+k} = \beta_0 + \beta_1 D_t + \varepsilon_t \tag{4}$$

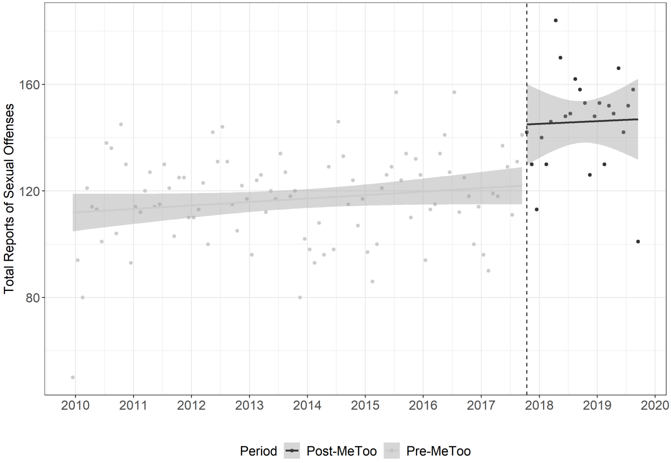
An Example

Figure: Sexual Crimes - Date of the Incident (NYC)



An Example

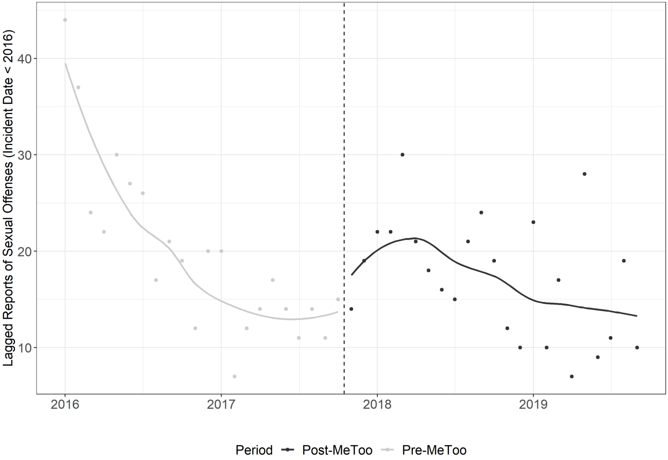
Figure: Sexual Crimes - Date of the Report (NYC)



Plan

Preliminary Evidence and Intuition

Figure: Aggregate Hazard of Sexual Crime Reports (NYC)



▸ Placebos

Modeling MeToo's Impact

- Probability of reporting in k a crime committed in t conditional on not having reported it before:

$$h(k \mid t) = \frac{R_{t,k}}{C_t - \sum_{j=1}^{k-1} R_{t,j}} \tag{5}$$

- It is natural to think of the MeToo outbreak as a shock which shifts all probabilities such that:

$$h(k \mid t, MeToo) = (1 + \Delta) h(k \mid t) \tag{6}$$

- But C_t is (1) unobserved and (2) potentially affected by MeToo (e.g. composition effects)...

Police data as survival analysis

A Simple Solution

Method

1. For a specified p, guess $F(p \mid t) = \mu$

   Example: $p = 30$, $\mu = \frac{1}{2}$
   50 crimes reported in less than 30 days for period t
   $\rightarrow$ 100 crimes committed in total in period t

2. Focus on crimes which occurred before MeToo.

Assumptions

- Reporting behaviors were stable in the pre-treatment period.
- Some degree of proportional hazards is required to infer $F(p \mid t, MeToo)$.

Baseline Specification

- I model the hazard such that:

$$h(k \mid t, X, Z(k)) = h_0(k) \cdot exp(\beta_t + \gamma' X + \tau' Z(k)) \tag{7}$$

- k = time-to-report to the police
- t = date of the incident
- X = time-invariant victim characteristics
- Z(k) = vector of time-varying covariates (i.e. metoo(k))
- h_0 = unspecified baseline hazard (Cox, 1972)

Lessons from Monte Carlo Simulations

Plan

Main Results

Table: Me Too Effect on Reporting Behaviors

	Dependent variable: Time-to-Report to the Police							
	ATE	PH Test	Persistence	Austin	Nashville	Cincinnati	Los Angeles	New York
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Me Too	0.403***	0.396***	0.282***	0.026	0.157	0.760**	0.385***	0.514***
	(0.069)	(0.066)	(0.061)	(0.215)	(0.397)	(0.336)	(0.098)	(0.125)
Me Too (3 months)			0.236***					
			(0.083)					
Me Too (6 months)			−0.029					
			(0.073)					
Me Too * Old Crime		0.051						
		(0.079)						
Period Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City Strata	Yes	Yes	Yes	No	No	No	No	No
Start Date	2016	2016	2010	2016	2016	2016	2016	2016
End Date	2019	2019	2019	2019	2019	2019	2019	2019
Observations	22,759	99,082	118,703	4,001	1,559	1,414	11,104	4,681

Note: Cox regression results. Estimates are displayed on the log-scale.
*p<0.1; **p<0.05; ***p<0.01

▸ Reporting Patterns ▸ Old or Recent Crimes? ▸ Discrete-time Specification ▸ Placebo Dates

Main Results

- Average treatment effect of 0.4 $\rightarrow$ Hazard ratio of $1 + \Delta = 1.5$
- Interpretation: "At any given period k, the probability of reporting a crime to the police conditional on not having reported it before increases by 50% after Me Too."
- No evidence of a stronger effect for past sexual crimes.
- The effect increases in the first months and then remains stable for the following year.
- Strong treatment heterogeneity between cities (true differences or data quality issues?)

Back of the Envelope Calculations

- Given the Cox model, we have

$$S(k \mid t, MeToo) = S(k \mid t)^{(1+\Delta)} \tag{8}$$

- If $\lim_{k\to\infty} S(k \mid t) = 50\%$, then $\lim_{k\to\infty} S(k \mid t, MeToo) \approx 35\%$

$\rightarrow$ 15% additional victims would file a complaint to the police.

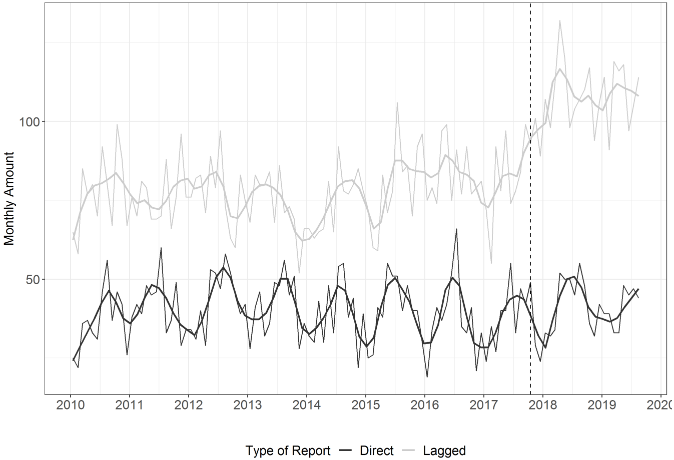
- If $\lim_{k\to\infty} S(k \mid t) = 90\%$, then $\lim_{k\to\infty} S(k \mid t, MeToo) \approx 85\%$

$\rightarrow$ 5% additional victims would file a complaint to the police.

Plan

Once More, Reported Crime Is Likely Misleading.

Figure: Direct vs. Lagged Reports (NYC)



Studying ~~Reported~~ Effective Crime

- Recall that:

$$R_{t,1} = h(1 \mid t, MeToo)C_t = h_0(1 \mid t)(1 + \Delta)C_t \tag{9}$$

- I rescale all direct police reports at a constant reporting rate:

$$\tilde{R}_{t,1} = \frac{R_{t,1}}{(1 + \Delta)} = h_0(1)C_t \tag{10}$$

- I can now study variations in effective crime.

Event-Study Specification

$$log(\tilde{R}_{t,c}) = \alpha + \beta_t Month_t + \theta_c City + \delta_c \cdot City \cdot t + \gamma MeToo_t + \varepsilon_t \quad (11)$$

- $\tilde{R}_{t,c}$ = sexual crimes in period t at a constant reporting rate $h_0(1)$
- $MeToo_t$ = dummy for the treatment period
- δ_m = month fixed effects
- δ_c = city fixed effects
- I control for differential linear time trends in crime reports per city.
- ε_t = error term

Main Results

Table: Me Too Effect on Reported vs. Effective Sexual Crimes

	Dependent variable: Monthly Sexual Crime Statistics (in logs)					
	Direct	Direct	Total	Total	Corrected	Corrected
	(1)	(2)	(3)	(4)	(5)	(6)
Me Too	-0.043	-0.088	-0.009	-0.118**	-0.446***	-0.133
	(0.044)	(0.094)	(0.029)	(0.059)	(0.047)	(0.097)
Me Too * Cincinnati		0.049		0.234***		-0.792***
		(0.134)		(0.085)		(0.138)
Me Too * Los Angeles		0.143		0.095		-0.205
		(0.133)		(0.084)		(0.137)
Me Too * Nashville		-0.202		-0.200**		-0.390**
		(0.149)		(0.094)		(0.153)
Me Too * New York		0.151		0.315***		-0.218
		(0.133)		(0.084)		(0.137)
Month Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
City Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Linear Time Trend	Yes	Yes	Yes	Yes	Yes	Yes
Observations	546	546	545	545	546	546
R^2	0.879	0.880	0.946	0.950	0.879	0.887
Adjusted R^2	0.874	0.875	0.944	0.948	0.874	0.882
Residual Std. Error	0.294	0.293	0.191	0.185	0.311	0.302
F Statistic	180.939***	153.194***	439.466***	394.589***	181.454***	163.427***

Note: Results from an event study for direct reports of incidents which occured between January 2010 and Sept. 2019. *p<0.1; **p<0.05; ***p<0.01

Plan

How Should We Interpret These Results?

- Law as *'a system of rules that are created through formal institutions to regulate behavior'*.
- But law enforcement agencies imperfectly monitor illegal behaviors... Some remain unpunished and persist over time.
- Social protest movements may try to breach the status quo through norms-based interventions (Benabou & Tirole, 2011)
- Attempt to enforce a new social norm (i.e. define 'normal' behavior)

→ Empirical evidence that norms-based interventions may be successful at shifting social norms.

Some Limitations

- The analysis of the deterrent effect is an extrapolation and thus heavily depends on the duration model's assumptions, notably proportional hazards.
- I cannot monitor whether people changed their definition of a sexual crime over time.
- If some plaintiffs report inappropriate yet legal behaviors of men as felonies, this would lead to an upward bias in victim reporting behaviors...
- And consequently to an upward bias on the Me Too movement's deterrent effect.

Plan

Key Take-Aways

- Researchers should be cautious when working with police data, as under-reporting and lagged reporting may lead to spurious correlations.
- Lagged reports offer an intuitive identification strategy to separate victim and offender behaviors from police records.
- The MeToo movement has led to many controversies on its supposed flaws and merits...
- I provide empirical evidence that it likely:
  - Increased victim reporting of sexual crimes
  - Had a deterrent effect on sexual offenders

The Impact of #MeToo on Sexual Criminality

Germain Gauthier

CREST - Ecole Polytechnique

September 23, 2020

#MeToo in the United States

Figure: #MeToo Tweets in the United States

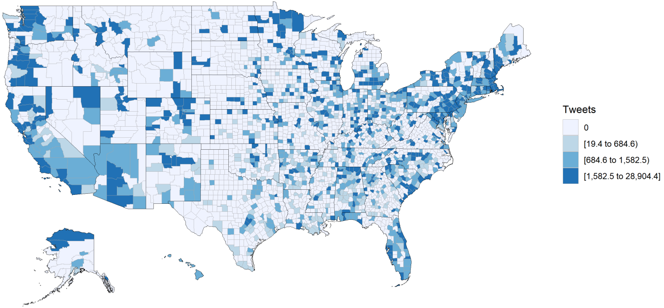
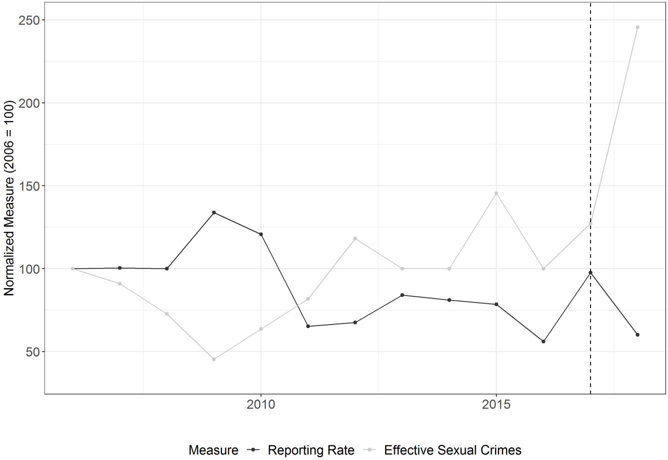


Figure: *

Source: Author's own calculations. The number of tweets is weighted by the inverse of the twitter penetration rate per county.

NCVS - Estimates

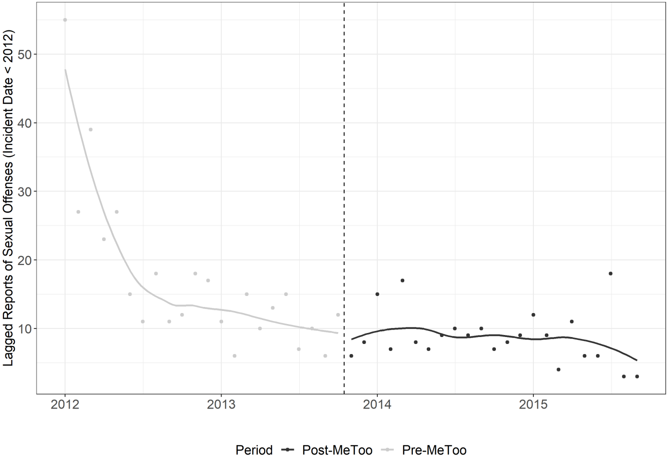
Figure: Variations in Crime and Reporting Rates (NCVS)



Go Back

Pseudo-Hazard (Placebo Date)

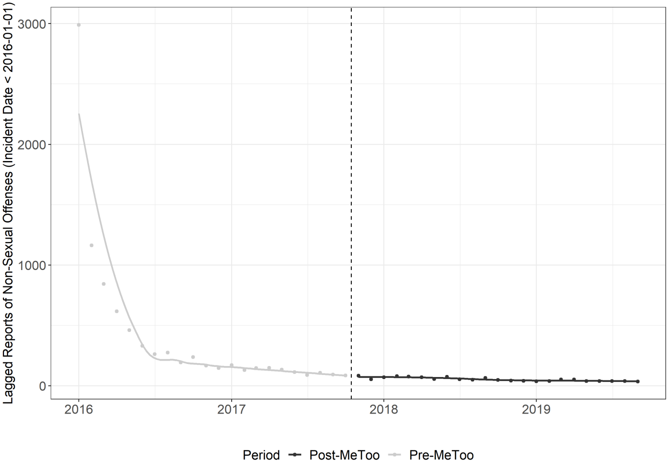
Figure: Pseudo Hazard of Sexual Crime Reports - Placebo Date (NYC)



Go Back

Aggregate Hazard (Non-sexual Crimes)

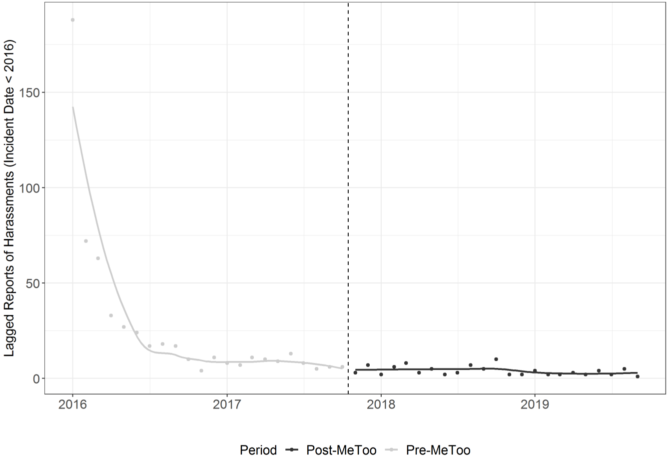
Figure: Aggregate Hazard of Non Sexual Crime Reports (NYC)



Go Back

Aggregate Hazard Male Harassments

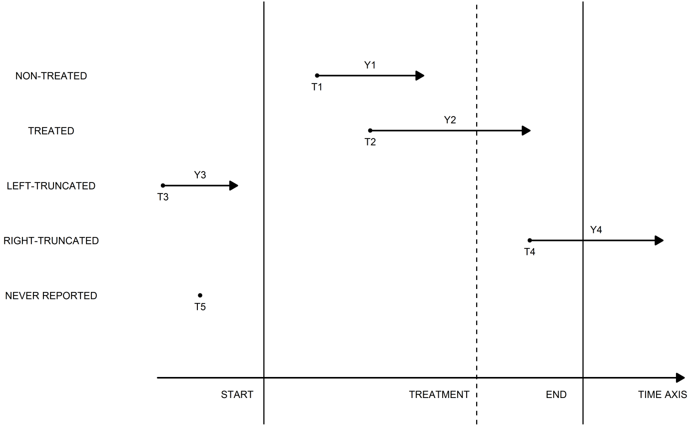
Figure: Aggregate Hazard of Male Harassments (NYC)



Go Back

Police Data as Survival Analysis

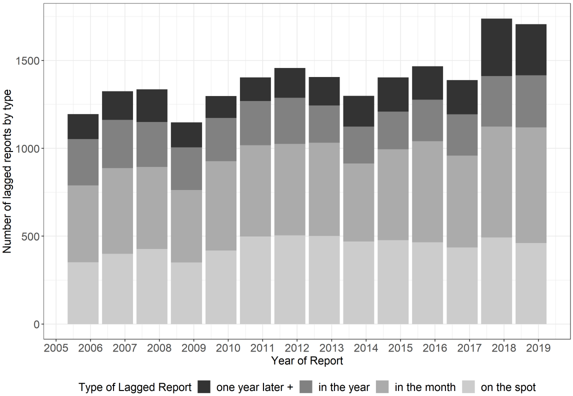
Figure: Aggregate Hazard of Sexual Crime Reports (NYC)



Go Back

Old or Recent Crimes?

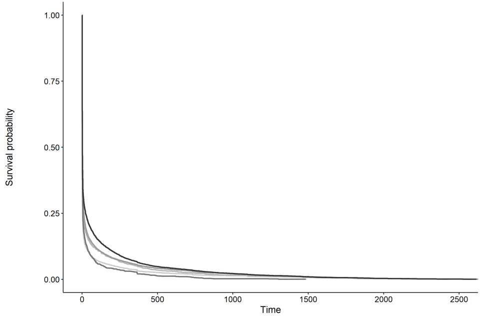
Figure: Sexual Crime Reports By Duration and Year (NYC)



Go Back

Survival Curve Estimates - Cities

Figure: Sexual Crime Reports By City



Number at risk

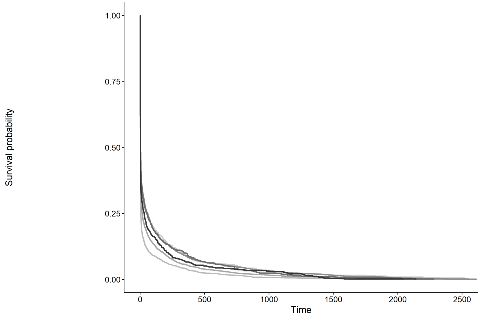
City	0	500	1000	1500	2000	2500
austin	6096	157	85	39	17	4
cincinnati	2376	83	37	20	6	0
los_angeles	18011	756	378	177	58	13
nashville	1742	23	4	0	0	0
new_york	10183	493	223	94	33	9

Time

Go Back

Survival Curve Estimates - Victim Race

Figure: Sexual Crime Reports By Victim Race (NYC)



Number at risk

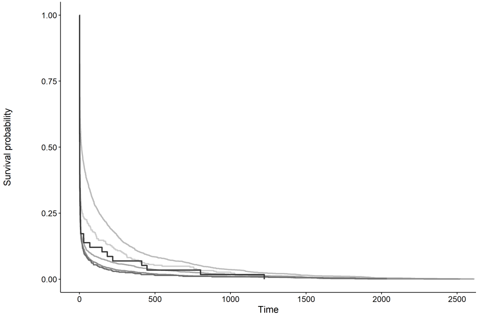
Victim Race	0	500	1000	1500	2000	2500
UNKNOWN	571	37	19	9	5	3
WHITE	1704	42	13	6	1	0
BLACK	3712	145	59	23	5	2
WHITE HISPANIC	2962	196	98	47	18	4
BLACK HISPANIC	659	44	16	7	3	0
ASIAN / PACIFIC ISLANDER	575	29	18	2	1	0

Time

Go Back

Survival Curve Estimates - Victim Age

Figure: Sexual Crime Reports By Victim Age (NYC)



Number at risk

Victim Age	0	500	1000	1500	2000	2500
UNKNOWN	306	16	8	2	1	0
<18	3716	315	135	55	18	5
18-24	2498	91	44	19	10	3
25-44	3034	60	30	15	3	1
45-64	571	9	5	3	1	0
65+	58	2	1	0	0	0

Time

Monte Carlo Simulations

- I simulate the data-generating process:
  - $t \in [1, 100]$ & $k \in [1, 10]$
  - h(k) $= 0.05 +$ log(2) MeToo(k) $+ \mu_k$ with $\mu_k \sim N(0, 0.01)$
  - $C_t = 1000 + \varepsilon_t$ with $\varepsilon_t \sim N(0, 100)$
- I find that:
  1. The approach successfully recovers the variation in victim reporting behaviors.
  2. The approach does not suffer from right-truncation bias.
  3. Results may be sensitive to the initial guess for the dark figure of crime. Overshooting leads to an upward bias. Conversely, underestimating the dark figure leads to a downward bias. The size of the bias is an empirical matter.

Go Back

Robustness to Alternative Specifications

- I model the hazard such that:

$$g(h(k \mid t, X, Z(k)) = h_0(k) + \beta_t + \gamma' X + \tau' Z(k) \tag{12}$$

- $g(\cdot)$ is a link function (Gompertz)
- k = time-to-report to the police
- t = date of the incident (monthly)
- X = time-invariant victim characteristics
- Z(k) = vector of time-varying covariates (i.e. metoo(k))

Go Back

Robustness to Alternative Specifications

Table: Robustness - Me Too Effect on Reporting Behaviors (Gompertz Model)

	Dependent variable: Time-to-Report to the Police					
	ATE	Austin	Nashville	Cincinnati	Los Angeles	New York
	(1)	(2)	(3)	(4)	(5)	(6)
Me Too	0.439***	0.079	0.168	0.769**	0.429***	0.546***
	(0.068)	(0.211)	(0.391)	(0.326)	(0.096)	(0.122)
Period Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
City Strata	Yes	Yes	Yes	Yes	Yes	Yes
City Strata * Period Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	193,438	34,673	12,460	11,938	92,091	38,179
Log Likelihood	-20,196.740	-3,027.170	-1,116.406	-1,098.262	-10,027.900	-4,915.128
Akaike Inf. Crit.	40,783.490	6,158.341	2,324.813	2,298.524	20,159.790	9,940.256

Note: Results from an event study for incidents which occurred after January 2016 and were reported before December 2019, 31. Estimates are displayed on the log-scale. *p<0.1; **p<0.05; ***p<0.01

Go Back

Does the Value of μ Matter?

Table: Me Too Effect on Reporting Behaviors - Robustness

	Dependent variable: Time-to-Report to the Police				
	$\mu = 2$	$\mu = 4$	$\mu = 6$	$\mu = 8$	$\mu = 10$
	(1)	(2)	(3)	(4)	(5)
Me Too	0.403***	0.410***	0.412***	0.413***	0.413***
	(0.069)	(0.070)	(0.070)	(0.070)	(0.070)
Period Fixed Effects	Yes	Yes	Yes	Yes	Yes
City Strata	Yes	Yes	Yes	Yes	Yes
Observations	22,759	54,651	86,543	118,435	150,327
Wald Test (df = 20)	53.690***	49.710***	49.160***	48.950***	48.850***
LR Test (df = 20)	51.481***	49.005***	48.736***	48.647***	48.605***
Score (Logrank) Test (df = 20)	52.523***	49.971***	49.689***	49.594***	49.549***

Note: Results from a Cox regression for incidents which occurred after January 2016 and were reported before December 2019, 31. Estimates are displayed on the log-scale. *p<0.1; **p<0.05; ***p<0.01

Go Back

Were Reporting Patterns Stable in the Pre-Treatment Period?

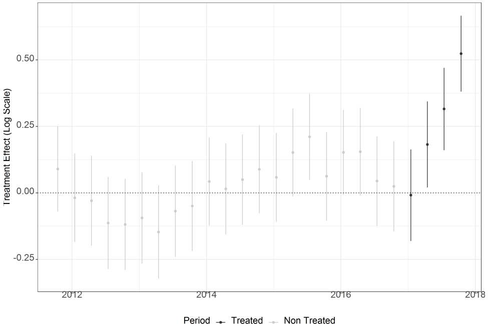


Figure: Treatment Effect for Placebo Dates (NYC)

Go Back