Methodological Appendix I – XII for "Whose Lives Matter? Mass Shootings and Social Media Discourses of Sympathy and Policy, 2012-2014." Yini Zhang, Dhavan Shah, Jordan Foley, Aman Abhishek, Josephine Lukito, Jiyoun Suk, Sang Jung Kim, Zhongkai Sun, Jon Pevehouse, and Christine Garlough. *Journal of Computer-Mediated Communication*.

## Appendix I

We collected the event data from three distinct sources. The MSA is collected based on online news media sources. The GVA database is based on a combination of online news sources, police media outlets, and police blotters. The USA Today database is based on the Supplementary Homicide Reports (SHR) from the FBI. While no individual event dataset claims to be exhaustive, they represent three diverse levels of source selection (news media, local police reports, and FBI-reported data) and each have their own form of source validation.

When coding for race, coders first reviewed event source data for a police report or news organization that explicitly identified the race of the victim. Phenotypical attributes were only used in combination this kind of contextual information. If there is any ambiguity, coders deferred to marking as race unknown. In rare cases where multiple shooters were present (N = 3), both shooters were of the same race.

## Appendix II

Two steps were taken to collect the relevant data. First, general search strings—“gun,” “shooter,” “shooting,” “firearm,” “second amendment,” “2nd amendment”, “nra”—were used to capture relevant content comprehensively. The search strings yielded 75,271,571 tweets. The second step reduced noise in the harvested dataset. Two coders were assigned to produce an exclusion list containing words and phrases marking irrelevant tweets. The list of exclusion words was generated by applying Latent Dirichlet Allocation (LDA), a form of topic modeling, to classifying tweets into 100 topics. Two graduate student coders each went through 50 topics, evaluating documents and terms with the goal to further reduce irrelevant tweets by adding words/phrases to the exclusion list. All the tweets containing one of the exclusion words or phrases were deleted; foreign language tweets were also removed through the exclusion list.

| **Exclusion List (words/phrase used to exclude noise from the collected Twitter data)** |
|---|
| bangun, gunna, bingung, guna, video, que, lagunya, camera, film, photo, movie, smoking, smokes, algun, song , gundy, gunzo, begun, topgun, top gun, laguna, flu shot, moscow, canada, rcmp, moncton, indonesia, toronto, vancouver, britain, lancaster, manchester, london, australia, france, paris, french, pakistan, karachi, afghanistan, iraq, baghdad, yemen, syria, isis , egypt, bahrain, qatar, saudi, turkey, turkish, malala, taliban, charliehebdo, charlihebdo, ukrain, kenya, nairobi, sudan, africa , nigeria, borno , bomb, bird , tiger , india , delhi, idf , gaza , israel, palestin, director, tony scott, arsenal, nuclear, germany, berlin, dutch, venezuela, uae, walking dead, talking dead, walkingdead, talkingdead, russian, nemtsov, tulsa, robbery, zombie, walkers, wii , kiev, montolivo, missile, meyiwa, segund, segunod, palestinian, anzhi, copenhagen, charlie, hebdo, music, singer, latore, alguns, chikungunya, screenshot, walker, haram , boko , ninguno, kabul, pregunta, abeokuta, malaysian, dungun, gratata, benghazi, laden, drone, ebola, gunter, shottar, russia, khalifa, soviet, wwii, birth control, horse, kashmir, gundam, xbox, tayo, malaysia, riot, jordan, rubber, lagos, canadian, homie, check twitter, bright, sydney, ankara, beckham, free kick, shot me, istanbul, sex , mug shot, mugshot, police shooting, ferguson, cop shooting, deadass, coffee, stoppoliceviolence, gungtang, already killed me, tamerlan, tsarnaev, rideau, vote rigging, pull my trigger, knife, freeze, chris brown, chrisbrown, mike brown, mikebrown, ontario, shmurda, trayvon, zimmerman, gunpoint, jim crow laws, tamir rice, anggun, gung, take a shot, brazil, brasil, german, jihad, tunisia, tunis, libya, segun, korea, peshawar, milan, gunung, somali, islamist, libyan, melbourne, deadline, pergunta, zedi, feruzi, abuja, jamaica, japan, denmark, mali, benue, singapore, afghan, columbia, switzerland, marseille, tripoli, rio, burundi, ottawa, sweden, china, macedonia, belfast, swiss, swedish, iran, abuja, gunplay, just killed a man, waking up beside you , ambulance i think my friend is dead, his estranged wife and another woman are dead, mexico mayor-elect, shotgun, lol, military, battle, police brutality, nigga, nerf, buckingham palace, deadshot, kansas city chiefs, columbian drug barons, screenshot , screenshooting, ebonyi, nigga, nerf, buckingham palace, deadshot, columbian drug barons, gaza, call of duty, long-shot, longshot, head shot, sharpshooter, monyashooter, goodshooter, shootout, ladiesshooting, troubleshooting, sharpshooter, teamshooter |

## Appendix III

| **Codebook for Relevance Classifier** |
|---|
| **Relevance: label relevant=1; irrelevant=0** |

Relevant tweets include “factual” statements and personal commentaries about:

1) Any domestic/U.S. shooting (so long it’s not about terrorist attack like Boston Marathon Bombing; we focus on domestic gun violence) including breaking news and emotional reactions. E.g., “Suspect in shooting near Auburn University that killed 3, turns himself in to... http://t.co/3XzWSXIl”
2) General gun-related violence, including general commentary about hearing gunshots, such as “I hear 16 shots fired”
3) Gun rights
4) Gun policy
5) NRA
6) Remember that we are looking for relevant tweets about domestic gun violence for projects that come along later

Irrelevant tweets include:

1) Non-U.S. shootings
2) Terrorist attacks that do not involve mass shootings.
3) Police shooting and #BLM, any shooting involving police
4) International events or foreign language tweets
5) Tweets that are too general like

## Appendix IV

**Codebook for Discourse Classifier**

**Three categories: use rights, control, thoughts, and 0 to label**

General rules:

1) Always consider the hashtag, but it's not the sole base for making your judgment
2) If the content of the tweet is IN DIRECT CONFLICT or IN DIRECT OPPOSITION with the use of a hashtag that is traditionally control/rights, the context of the tweet takes precedence.
3) In situations where the content would be neutral, code according to the hashtag
4) If there are a mix of hashtags and the content of the tweet is ambiguous, code 0.
5) Everyday gun violence should be relevant content for consideration (a tweet doesn't necessarily have to be about a mass shooting).
6) Exclude police brutality or police shootings. The shootings of police, however, should be relevant.

Defending second amendment rights [rights]:

1) NOTICE THAT "PRO-GUN" TWEETS ARE NOT NECESSARILY "RIGHTS"
2) NOTICE THAT SOME TWEETS CONCERN GUN RIGHTS IN GENERAL, LIKE ARGUING AGAINST GUN RIGHTS, BUT HERE WE ARE CODING FOR TWEETS THAT EXPLICITLY SUPPORT OR DEFEND GUN RIGHTS.
3) Tweets that specifically and explicitly support second amendment rights, constitution, freedom and liberty.
4) It could be about "news" as well as opinions on gun rights, constitutional rights, freedom and liberty.
5) Any "news" type of tweets that indicates positive sentiment toward gun rights should be included. E.g. "Senator Rand Paul expresses support for second amendment rights."
6) Look for major hashtags under this category: These include hashtags such as 2a, 2nd, 2ndamendment, billofrights, constitution, donttreadonme, gunright, gunrights, iamforgunrights, right2defend, rights, righttobeararms, protect2a, secondamendment, freedom, liberty, selfdefense, shallnotbeinfringed, wethepeople.
7) Don't extrapolate here. A tweet like "Guns don't kill people, abortions do," although indirectly expresses support for gun rights should not be included. By the same token, "Proud gun owner, #NRA" will not be included.

Call for gun control policy [control]:

1) NOTICE THAT "ANTI-GUN" TWEETS ARE NOT NECESSARILY "CONTROL"
2) NOTICE THAT THERE WILL BE TWEETS ABOUT LEGISLATION IN GENERAL, BUT WE ARE CODING FOR TWEETS THAT SUPPORT GUN POLICY THAT RESTRICTS GUN ACCESS.
3) Tweets that call for gun control policy: more strict gun regulation and restricted access to guns.
4) This category should include any tweet that concerns "news" or opinions on particular gun control policy or calls for legislative action with regard to gun control policy.

5) Any news type of tweets that disseminate restrictive gun control discourse should be included. E.g., “Stars React to Colorado Shooting, Push for Gun Control Laws - Amidst the sadness and... http://t.co/OdRmw0MO #DarkKnightRises #MileyCyrus”
6) But tweets like “RT @floodthedrummer: Rendell, Nutter demands vote on Gun Legislation. http://t.co/veLnc2ulvw @ALBDAMN @MrJAlabaster @normbond @HMCTwit @ ...” are not clear on what the gun law is should be excluded.
7) Tweets that call for “stopping gun violence” without specific mention of gun control measure should not be included.
8) Tweets that are descriptive, not prescriptive, (“Why school shootings don’t lead to tighter gun control in the US http://t.co/A7736ua9 http://t.co/4obllwVS”) should not be included.
9) Major hashtags include: backgroundcheck backgroundchecks fixdvgunlaws guncontrol guncontrolnow gunlaw gunlaws gunreform gunregistry control demandaction demandaplan momsde momsdemand momsdemandaction nowisthetime Universal...checks wedemandavote whatwillittake
10) Don’t extrapolate. A tweet like “Anti-Gun Chicago Legislator Arrested At Airport -- With Gun... http://t.co/AcxTRAUH” should not be included. Though it hints at the hypocrisy of the legislator, it boils down to a person, not a policy. By the same token “RT @News24lHOT: Virginia USA Arlington » http://t.co/621F5K76vD #JamesBrady 878 James S. Brady, Symbol of Fight for Gun Control, Dies at 7…” is irrelevant.

Thoughts and prayers [thoughts]:

1) About condolences and sadness.
2) Major hashtags / keywords under this category include: pray, prayer, prayerfor…, prayers, prayersfor…, prayfor…, prayforthe…, praying, rip, pray for, heartbreaking, tragic, tragedy, godbless, sosad.
3) Don’t not include tweets that only contain other sentiments like disgust and anger.

All other [0]:

1) any other tweets

**Appendix V**

**Three Groups of Hashtags (selected from hashtags appearing 200 times+ in our data)**

| Thoughts and Prayers | Gun Control | Gun Rights |
|---|---|---|
| godbless | backgroundcheck | 2 |
| godblessamerica | backgroundchecks | 2nd |
| heartbreaking | bradycampaign | 2nda |
| pray | control | 2ndamend |
| prayer | demandaction | 2ndamendment |
| prayerfornewtown | demandaplan | 2ndammendment |
| prayers | doitforgabby | beararms |
| prayersforconnecticut | endfacebookgunshows | billofrights |
| prayersfornewton | endgunviolence | constitution |
| prayersfornewtown | fixdvgunlaws | donttreadonme |
| prayf | gunban | freedom |
| prayfo | gunc | gunright |
| prayfor | gunco | gunrights |
| prayforaurora | guncon | iamforgunrights |
| prayforboston | guncont | liberty |
| prayforclaire | guncontr | newnjgunlaw |
| prayforcolorado | guncontro | newnjgunlaws |
| prayforconnecticut | guncontrol | nj |
| prayforct | guncontrolnow | nj2 |
| prayfordc | gunfreezone | nj2a |
| prayforfsu | gunfreezones | nj2ar |
| prayforne | gunreform | nj2as |
| prayforneworleans | gunse | protect2a |
| prayfornewto | gunsen | right2defend |
| prayfornewton | gunsense | rights |
| prayfornewtown | gunviolence | righttobeararms |
| prayforpilchuck | gunvote | secondamendment |
| prayforpurdue | indygunsense | selfdefense |
| prayforsikhs | momsde | shallnotbeinfringed |
| prayforspu | momsdemand | thelibertyamendments |
| prayfortheparents | momsdemandaction | tyranny |
| praying | momstakethehill | wearethepeople |
| restinpeace | momsvote | wethepeople |
| rip | noguns | |
| sad | nogunsallowed | |
| tragedy | nomoreguns | |
| tragic | nowisthetime | |
| | pagunsense | |
| | protectchildrennotguns | |
| | stoptheviolence | |
| | theydeserveavote | |
| | timetoact | |
| | universalbackgroundchecks | |
| | votegunsense | |
| | wedemandavote | |

# Appendix VI

**Correlation Matrix: Supervised Machine Learning Approach and Hashtag-based Approach**

| | Thoughts and Prayers (hashtags) | Gun Control (hashtags) | Gun Rights (hashtags) |
|---|---|---|---|
| Thoughts and Prayers (tweets) | **.94** | .59 | -.02 |
| Gun Control (tweets) | .43 | **.86** | .13 |
| Gun Rights (tweets) | .16 | .41 | **.89** |

## Appendix VII

To further examine whether the high correlations resulted from the supervised ML classifiers using the grouped hashtags to infer tweet content, we identified all tweets containing a certain type of hashtags, and then examined the overlap between them and all tweets classified by ML algorithms. As can be seen in the table below, the overlap between the tweets derived from the two approaches is small. For example, the common “thoughts and prayers” tweets based on the two approaches make up only 26% and 9% respectively of the total number of tweets based on the hashtags and ML classifiers. Although in one case 60% of gun rights tweets classified by ML appeared in the gun rights tweets classified by hashtags, only 24% of gun rights tweets classified by hashtags were picked up by the ML classifier. Therefore, we are confident that the high correlations between hashtag and ML measures are robust.

**Overlap between Tweets Containing Hashtags and Tweets Classified by Supervised Machine Learning**

| discourse category | total number of hashtag tweets | total number of ml tweets | number of shared tweets | proportion of shared tweets in total hashtag tweets | proportion of shared tweets in total ml tweets |
|---|---|---|---|---|---|
| Thoughts and Prayers | 50,728 | 146,337 | 13,387 | .26 | .09 |
| Gun Control | 226,302 | 172,837 | 26,822 | .12 | .16 |
| Gun Rights | 238,050 | 94,612 | 56,791 | .24 | .60 |

## Appendix VIII

**Correlations between Independent Variables**

| | # of victims | # of children killed | # of African Americans killed | Shooter Race (1=white) | Public shooting | School shooting |
|---|---|---|---|---|---|---|
| # of victims | 1 | | | | | |
| # of children killed | .239 | 1 | | | | |
| # of African Americans killed | -.111 | -.106 | 1 | | | |
| Shooter Race (1=white) | .168 | .144 | -.588 | 1 | | |
| Public shooting | .394 | -.012 | -.155 | .092 | 1 | |
| School shooting | .179 | .421 | -.127 | -.086 | 0.307 | 1 |

*Note.* The low correlations between the the independent variables alleviate collinearity concerns.

## Appendix IX

In our final data, 26 rows/days have NAs across all six social media variables, which suggest loss of Twitter data stream on those days. Table 1 was based on data using a simple data imputation method—recoding all those NAs as zero. We also applied more sophisticated data imputation techniques, including linear interpolation (for 1-3 day gaps) and forecasting (for one longer 7-day gap) to replace NAs with estimated values. The results are presented below. Compared with Table 1 presented in the article, the significance of coefficients is identical, with only slight changes to the values of coefficients and standard errors.

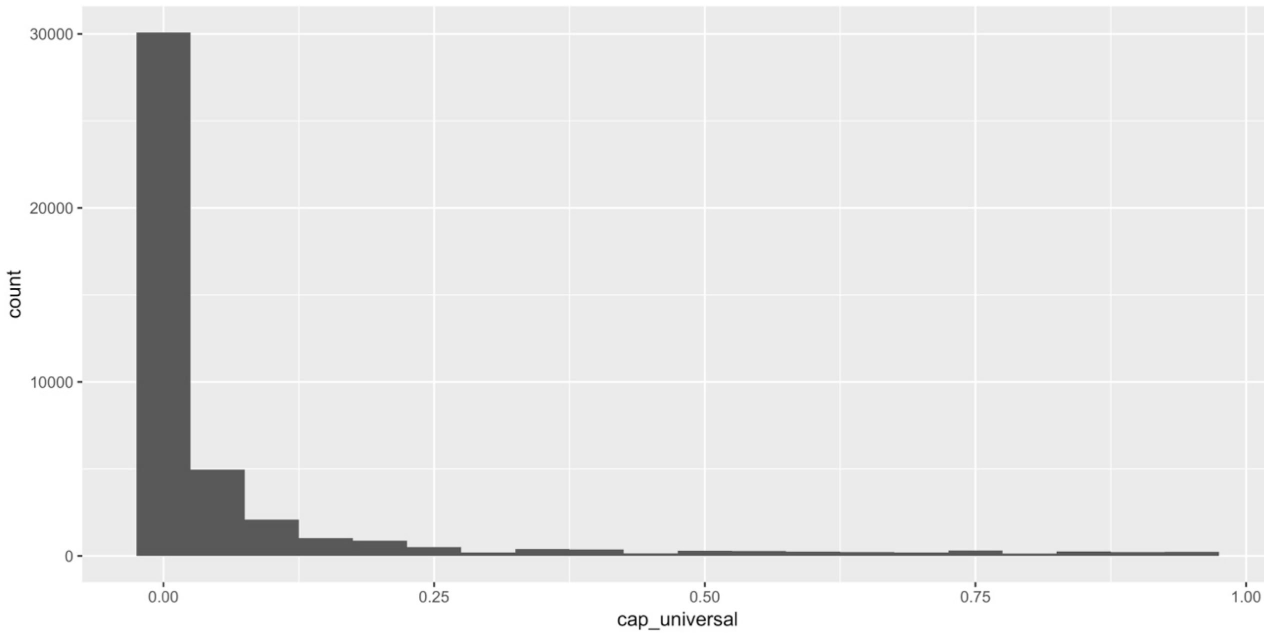
**Time Series Regression Models Predicting the Volume of Tweets and Hashtags (with missing values in social media variables imputed through linear interpolation and forecasting)**

| | Supervised machine learning approach (tweets) | | | Hashtag-based approach (hashtags) | | |
|---|---|---|---|---|---|---|
| | Thoughts and Prayers | Gun Control | Gun Rights | Thoughts and Prayers | Gun Control | Gun Rights |
| Number of victims | 344.145*** (13.285) | 8.272* (4.135) | 1.123 (.762) | 63.960*** (6.888) | .728 (3.390) | .589 (1.901) |
| Number of children killed | 1096.959*** (45.184) | 130.739*** (14.068) | 13.968*** (2.589) | 459.564*** (23.428) | 201.086*** (11.532) | 9.973 (6.459) |
| Number of African Americans killed | -789.074*** (58.229) | -47.239** (18.111) | -7.298* (3.351) | -238.900*** (30.182) | -71.776*** (14.848) | -19.107* (8.360) |
| Shooter Race (1 = white) | -2983.555*** (220.068) | 222.121** (68.466) | -31.443* (12.650) | -929.545*** (114.078) | -312.747*** (56.127) | -63.573* (31.564) |
| Public shooting | 364.582 (286.219) | 61.487 (89.150) | 34.263* (16.370) | 350.361* (148.429) | 251.636** (73.080) | 109.978** (40.849) |
| School shooting | -191.041 (521.385) | 132.236 (162.328) | -5.578 (29.873) | 103.695 (270.342) | 134.110 (133.069) | -46.985 (74.542) |
| Constant | 17.842 (44.849) | 149.193*** (31.307) | -.946 (1.252) | 14.955 (27.264) | 225.349*** (23.530) | .327 (3.144) |
| *Adjusted $R^2$* | .695 | .132 | .051 | .462 | .324 | .008 |

*Note.* Standard errors are in parentheses.
* $p < .05$, ** $p < .01$, *** $p < .001$ (two-tailed).

## Appendix X

The figure below plotted the distribution of bot probabilities from all sampled users. As is demonstrated, .25 seems to be a reasonable cut-off points as the overwhelming majority of users have a bot probability lower than .25.




*Distribution of Bot Probability (cap_universal) for all users with numeric Botometer result*

## Appendix XI

**Distribution of Types of Users with Tweets Originating from Each Type in Parentheses**

| | Supervised machine learning approach | | | Hashtag-based approach | | |
|---|---|---|---|---|---|---|
| | Thoughts and Prayers | Gun Control | Gun Rights | Thoughts and Prayers | Gun Control | Gun Rights |
| bot likely | 5% (5%) | 9% (10%) | 8% (14%) | 4% (13%) | 6% (9%) | 7% (13%) |
| bot unlikely | 62% (62%) | 68% (70%) | 69% (48%) | 60% (59%) | 73% (66%) | 68% (59%) |
| not authorized | 16% (16%) | 13% (12%) | 11% (17%) | 18% (11%) | 10% (13%) | 11% (11%) |
| not exist | 18% (17%) | 10% (8%) | 13% (21%) | 19% (17%) | 11% (11%) | 13% (17%) |
| Total | 100% (~100%) | 100% (100%) | 100% (100%) | 100% (100%) | 100% (100%) | 100% (100%) |

*Note*. Results are based on Botometer API. We label users with bot probability equal to and higher than .25 as "bot likely" and those otherwise as "bot unlikely." When a user does not authorize public access to its tweets, an error term "not authorized" is returned. Similarly, when its page is suspended or deleted, an error term "Sorry, that page does not exist." is returned. Percentage points without parentheses represents the percentage of users within each discourse, whereas percentage points with parentheses refers to the percentage of tweets originating from those users.

## Appendix XII

**Correlation Between Time Series of Sample Tweets from Bot Unlikely Users and Time Series of Total Tweets/Hashtags Used in Modeling**

| | Thoughts and Prayers | Gun Control | Gun Rights |
|---|---|---|---|
| Supervised machine learning approach | .99 | .96 | .91 |
| Hashtag-based approach | .99 | .89 | .95 |

*Note.* Each number represents the correlation between sample tweets time series produced by bot unlikely users only and complete tweets time series used in our modeling, for a certain discourse produced by ML or hashtag-based approaches. For example, for the "thoughts and prayers" discourse identified by ML approach, the correlation between the time series of the sample tweets produced by bot unlikely users and the time series of total "thoughts and prayers" tweets is .99. Correlation was calculated based only on common days that have none-zero observations (the time series from sample tweets by bot unlikely users only has missing observations on certain days since it is only a 10% sample).