Methodological supplement to "Mental Illness, the Media, and the Moral Politics of Mass Violence: The Role of Race in Mass Shootings Coverage"

Data collection

Random sampling principles must be followed to ensure the validity of results when performing quantitative analysis on documentary data. Firm criteria for inclusion and exclusion are necessary to ensure that the data are meaningful and measurements consistent (Hodson 1999). Thus, we constructed our sample by first identifying every news document that offered any motive or explanation for the shooting and for which the race of the perpetrator could be ascertained, roughly 5,000 documents. These are the key dependent and explanatory variables in our quantitative analyses, and so any article that did not fit these criteria would be treated as missing data and omitted from our final models.

We determined our sample size by calculating the margin of error for the sample such that we can be at least 95% confident that the sample reflects the universe of identified documents. The sample size was determined using $n = \frac{p(1-p)}{({.05}/{1.96})^2}$, where .05 reflects the margin of error, and *p* is the probability that an observation will deviate from the population mean. This is an algebraic rearrangement of $ME = 1.96\sqrt{\frac{p(1-p)}{n}}$, where ME is the margin of error, which we set to .05 to ensure 95% confidence in the sample reflecting the population. 1.96 is the *z*-score for the distribution of the confidence interval on which ME is based. Although this approach is typically employed in survey data collection, Singleton and Straits (2005) make the case that this type of sample design is appropriate for all kinds of quantitative research (pp. 140 – 141). Since *p* was unknown prior to collection, we followed conventional collection protocol and set *p* = .5. This is recommended when the expected probability of an event is unknown prior to data collection because it yields conservative estimates (Singleton and Straits 2005). This yielded an n of 385 articles. Hodson (1999) recommends including at least 15 cases per explanatory variable when collecting documentary data for statistical analysis. With 12 explanatory covariates, an n of 385 well exceeds the minimum necessary number of articles for statistical analysis (12 * 15 = 180).

Nevertheless, our data are nested because multiple articles cover the same shooting. If left unaddressed, nested data may misrepresent the amount of variation between documents. Thus, we increased the sample size by another 20% to 440 desired articles. Prior research informed the size of the increase to ensure that our sample is representative of an array of news coverage and provided sufficient statistical power for analyses. Samples of a few hundred to a thousand are typical in mixed-methods and quantitative analyses of documents (Carlson 2016; Dixon and Martin 2012; Hodson 1999; Roscigno et al. 2015; Savelsberg and Nyseth-Brehm 2015). Such samples are sufficiently small to render both in-depth coding and variable construction feasible, while also being large enough to offer statistical power.

Next, we employed systematic random sampling to collect a manageable dataset of documents covering mass shootings from the global list of roughly 5,000 articles. We created an interval size, k, and then selected every k^{th} document in the global list of mass shootings coverage between 2013 and 2015. The interval k was established by dividing the number of total cases by n (Singleton and Straits 2005). In our study, k = 11 (5,000/440).

To correct for potential selection bias—a common problem in analyzing news data where news document samples over-represent widely covered events (Earl et al. 2004; Martin 2005)—we collected a maximum of 15 documents per shooting. The choice of 15 documents per shooting was guided by prior studies on news data, where sampling design is constructed to ensure that there is substantial representation of events that receive relatively little coverage (Dixon and Martin 2012; Martin, McCarthy, and McPhail 2009). For example, in an analysis of strikes, Dixon and Martin (2012) constrain their sample such that roughly 40% of their data cover strikes which received little media attention.

Inter-rater reliability

Inter-rater reliability is a common concern in documentary analysis. Hodson (1999) recommends a few strategies to ensure that coding is consistent and that the resulting measurements are valid. First, variables should be mostly descriptive in nature. Thus, our covariates focus primarily on characteristics of the shooting and explicit references to the shooter's mental health. Second, Hodson (1999) recommends using at least two coders to examine 10% of the total sample to ensure that there is reliability between raters. We evaluated inter-rater reliability with three coders examining 40% of the total sample to ensure high reliability. All three coders evaluated a subsample of documents routinely throughout coding to ensure that codes remained consistent. Building in such reality checks to documentary analysis is a recommended strategy to ensure construct validity (Firebaugh 2008; Hodson 1999). This approach is also consistent with recent literature performing quantitative analysis of documents (Carlson 2016; Dixon and Martin 2012; Roscigno et al. 2015; Savelsberg and Nyseth-Brehm 2015). Consistent with prior studies (Hodson 1999; Roscigno et al. 2015), we measure inter-rater reliability as the percent agreement between coders divided by the total number of coding decisions. Three-coder reliability was high at 91%; above 80% is considered sufficient for statistical analysis (Hodson 1999).

Sensitivity analyses

There are two additional issues that warrant further attention in our statistical analyses. The first is nesting structure. Nested data violate the regression assumptions of independent and identically distributed observations with no autocorrelation in the errors. Failure to correct for this can result in biased regression results.

There are two common approaches to correcting for nesting. The simplest is to cluster the standard errors according to the nesting unit. This strategy is typically appropriate when nesting structures are not strong. The second strategy is to use multilevel models, which are more efficient than regression models with clustered standard errors, but they may overestimate contextual influence when there are only a few cases per cluster (Raudenbush and Bryk 2002). Thus, clustering standard errors is appropriate in our sample because the nesting structure is weak, with roughly 1.9 documents per shooting.

Nevertheless, we estimate a multilevel model as well as a logistic regression with clustered standard errors to highlight the robustness of the results. Results from both estimation strategies are reported in Table 1. While there are differences in point estimates between the two models, the substantive interpretation of results is consistent across modeling strategies.

[Table 1 here]

The second statistical issue which warrants attention is low cell counts. Logistic regression requires at least 5 entries in all cross-tabulations to yield meaningful estimates (King and Zeng 2001). As denoted in Figure 2 in the main text, we have only 6 instances of Black shooters who are framed as mentally ill. Such low cell counts can overestimate odds ratios in logistic regression.

There are two analytic approaches for correcting bias in logistic regression coefficients when data have low cell counts. The first is King and Zeng's (2001) rare events logistic regression; the second is Firth's (1993) bias-reduced penalized likelihood logistic regression. Rare events logistic regression corrects for bias in postestimation. King and Zeng (2001) show that the bias in a logistic regression coefficient can be estimated using weighted least squares regression. Thus, the rare events approach first estimates a logistic regression and then subtracts the bias in parameter vector to yield bias-corrected regression estimates.

Firth's (1993) strategy recognizes that rare events do not affect logistic regression directly, but rather introduce bias by problematizing maximum likelihood estimation (MLE). MLE finds the parameter set that maximizes the probability of observing the response vector. When there are a high number of zeros for a cross tabulation—as in the case of low cell counts—disproportionate weight is assigned to the variance and covariance estimates for those coefficients. Firth's logistic regression handles low cell counts by penalizing the likelihood function directly rather than removing bias in postestimation. Penalized likelihood estimation finds the parameter set that maximizes the product of the conventional likelihood and a saddle-shaped diffuse prior (Jeffrey's prior). The non-informative diffuse prior offsets the influence of low cell counts in the MLE, yielding valid coefficient estimates. Since penalized likelihood estimation restricts the influence of the data by weighting it towards a noninformative prior,[1] coefficients in Firth's regression are typically conservative estimates of the true population parameters.

Table 2 presents results for rare events logistic regression, Firth's logistic regression, and the logistic regression in the main text. Consistent with the expectation, Firth's regression provides smaller covariate effects than the bias-adjusted or conventional logistic regressions. Nevertheless, as in Table 1, the substantive interpretation of the results does not change across estimation strategies. Thus, sensitivity analyses suggest that the quantitative results specified in the main text are relatively unaffected by low cell counts.

[Table 2 here]

References

Carlson, Jennifer. 2016. "Moral Panic, Moral Breach: Bernhard Goetz, George Zimmerman, and Racialized News Reporting in Contested Cases of Self-Defense." *Social Problems* 63(1):1-20

Dixon, Marc, and Andrew W. Martin. 2012. "We Can't Win This on Our Own: Unions, Firms, and Mobilization of External Allies in Labor Disputes." *American Sociological Review* 77(6):946-69.

Earl, Jennifer, Andrew Martin, John D. McCarthy, and Sarah A. Soule. 2004. "The Use of Newspaper Data in the Study of Collective Action." *Annual Review of Sociology* 30:65-80.

[1] This is conceptually comparable to shrinkage towards the prior when using Bayesian methods to analyze small samples, though there is nothing Bayesian about Firth's regression.

Firebaugh, Glen. 2008. *Seven Rules for Social Research.* Princeton, NJ: Princeton University Press.

Firth, David. 1993. “Bias Reduction of Maximum Likelihood Estimates.” *Biometrika* 80(1): 27-38.

Hodson, Randy. 1999. *Analyzing Documentary Accounts.* Thousand Oaks, CA: Sage Publishing.

King, Gary, and Lagnche Zeng. 2001. “Logistic Regression in Rare Events Data.” *Political Analysis* 9(2):137-63.

Martin, Andrew W. 2005. “Addressing the Selection Bias in Media Coverage of Strikes: A Comparison of Mainstream and Specialty Print Media.” Pp. 141-78 in *Research in Social Movements, Conflict, and Change* Vol. 26, edited by P. G. Coy. Emerald Group Publishing.

Martin, Andrew W., John D. McCarthy, and Clark McPhail. 2009. “Why Targets Matter: Toward a More Inclusive Model of Collective Violence.” *American Sociological Review* 74(3):821-41.

Newman, Katherine S., Cybelle Fox, David J. Harding, Jal Mehta, and Wendy Roth. 2004. *Rampage: The Social Roots of School Shootings.* Cambridge, MA: Basic Books.

Raudenbush, Stephen W., and Anthony S. Bryk. 2002. *Hierarhical Linear Models: Applications and Data Analysis Methods.* Thousand Oaks, CA: Sage Publications.

Roscigno, Vincent J., Julia Miller Cantzler, Salvator J. Restifo, and Joshua Guetzkow. 2015. “Legitimation, State Repression, and the Sioux Massacre at Wounded Knee.” *Mobilization: An International Quarterly* 20(1):17-40.

Savelsberg, Joachim, and Hollie Nyseth-Brehm. 2015. "Representing Human Rights Violations in Darfur: Global Justice, National Distinctions." *American Journal of Sociology* 121(2):564-603.

Singleton, Royce A., and Bruce C. Straights. *Approaches to Social Research, 4th edition.* Oxford, UK: Oxford University Press.

Table 1. Comparison of multilevel model and logistic regression with clustered standard errors reported in the main text.

Independent Variables	**Multilevel**	**Main-text**
	Coefficient (SE)	**Coefficient (SE)**
Race (vs. Black)		
Whites	4.45** (1.46)	2.91** (1.18)
Latinos	2.93* (1.45)	2.52* (1.09)
Other	5.30** (1.84)	3.53* (1.42)
Number of victims	.209 (.120)	.086 (.115)
Number killed	.326 (.222)	.193 (.123)
Setting (vs. private)		
Business	1.50 (1.47)	1.70 (.950)
Public	2.87* (1.47)	1.59* (.772)
Other	4.40* (2.11)	1.53 (.959)
Gang violence	1.27 (1.08)	.763 (.819)
Murder-suicide	4.50** (1.54)	2.61** (.795)
Age of shooter	1.04 (.019)	.035 (.019)
Sex of victims (vs. women only)		
Men	-1.44 (1.64)	.307 (.944)
Men and Women	-.904 (1.16)	.873 (.910)
Unknown	2.46 (2.20)	3.07** (1.08)
Children shot	.361 (1.19)	.498 (.617)
Domestic-related	-.074 (1.19)	-1.01 (.845)
News Source (vs. local)		
Regional	-.118 (.897)	.156 (.486)
National	.963 (.862)	.974 (.486)
Year (vs. 2013)		
2014	.419 (.139)	.543(.665)
2015	-.874 (1.45)	.018 (.622)
Constant	-14.52*** (2.97)	-9.21*** (1.82)
Variance component	10.47	-
Log-likelihood	-78.4	-98.12

AIC	200.76	238.24
Model χ^2	-	176.86***

*p<.05 **p<.01 ***p<.001

Table 2. Comparison of rare events logistic regression, Firth's logistic regression, and logistic regression reported in the main text.

Independent Variables	**Rare events**	**Firth**	**Main-text**
	Coefficient (SE)	**Coefficient (SE)**	**Coefficient (SE)**
Race (vs. Black)			
Whites	2.53** (1.40)	1.40* (.653)	2.91** (1.18)
Latinos	2.21** (.820)	.573* (.256)	2.52* (1.09)
Other	3.09** (.869)	2.85 (1.72)	3.53* (1.42)
Number of victims	.081 (.073)	.126 (.068)	.086 (.115)
Number killed	.154 (.090)	.187 (.154)	.193 (.123)
Setting (vs. private)			
Business	1.43* (.645)	.607 (.442)	1.70 (.950)
Public	1.37* (.617)	.775 (.460)	1.59* (.772)
Other	1.32 (.834)	3.69 (4.46)	1.53 (.959)
Gang violence	.697 (.661)	-.275 (.369)	.763 (.819)
Murder-suicide	2.23*** (.536)	4.32* (2.17)	2.61** (.795)
Age of shooter	.029 (.017)	.040 (.024)	.035 (.019)
Sex of victims (vs. women only)			
Men	.195 (.753)	-.422 (.488)	.307 (.944)
Men and Women	.683 (.609)	-.146 (.457)	.873 (.910)
Unknown	2.61** (1.01)	.681 (.937)	3.07** (1.08)
Children shot	.444 (.552)	.374 (.507)	.498 (.617)
Domestic-related	-.861 (.614)	-.331 (.531)	-1.01 (.845)
News Source (vs. local)			
Regional	.157 (.517)	.017 (.438)	.156 (.486)
National	.858 (.510)	1.32 (.837)	.974 (.486)
Year (vs. 2013)			
2014	.440 (.561)	.551 (.423)	.543(.665)
2015	.006 (.533)	.136 (.413)	.018 (.622)
Constant	-7.90*** (1.39)	-1.03 (.828)	-9.21*** (1.82)
Model χ^2	176.86***	131.99***	176.86***
AIC	238.24	200.70	238.24

**p*<.05 ***p*<.01 ****p*<.001