Wells, C. F., Shah, D. V., Pevehouse, J. C. W., Yang, J., Pelled, A., Boehm, F., Lukito, J., Ghosh, S., & Schmidt, J. L. (2016). How Trump Drove Coverage to the Nomination: Hybrid Media Campaigning. *Political Communication*.

Methods Appendix

How Trump Drove Coverage to the Nomination: Hybrid Media Campaigning

CHRIS WELLS,[1] DHAVAN V. SHAH,[1] JON C. W. PEVEHOUSE,[2] JUNGHWAN YANG,[1] AYELLET PELLED,[1] FREDERICK BOEHM,[3] JOSEPHINE LUKITO,[1] SHREENITA GHOSH,[1] AND JESSICA L. SCHMIDT[1]

[1]*School of Journalism and Mass Communication, University of Wisconsin-Madison*
[2]*Department of Political Science, University of Wisconsin-Madison*
[3]*Department of Statistics, University of Wisconsin-Madison*

We developed measures of a variety of aspects of the media and opinion environment surrounding Trump's candidacy for the Republican nomination. The sections below describe how the measures were compiled. Note that our analyses present models of day-by-day measures: newspaper coverage was measured in terms of stories published per day, social media attention was operationalized in terms of retweets of Trump's tweets each day, events were presented as counts of events per day, and so on. The time period covered in our analyses was from June 15th, 2015, the day before Trump announced his candidacy, to May 4th, 2016, the day his last opponent, John Kasich, withdrew from the race. (In some analyses, we divide this period in two, distinguishing the pre-primary period, from June 15, 2015 to January 31, 2016–the day before the Iowa Caucuses–from the primary period, from February 1, 2016 to May 4.)

Newspaper Coverage

Article Retrieval

To identify news articles relevant to Donald Trump's election campaign, we selected four mainstream news sources available from Lexis/Nexis: The Associated Press, New York Times, USA Today, and Washington Post. Separately from the main newspaper sections, we also searched blogs managed by the New York Times and Washington Post. We selected news stories in which "Trump" occurred at least twice in single article or blog post, whether in the article's text, headline, or other Lexis/Nexis field (such as keyword). The search term we used is as follows:

```
trump w/seg trump
```

Because of the frequency with which Trump appeared in Washington Post blog posts, we narrowed the set of Washington Post blog stories to those that mentioned "Trump" twice within 255 words. The search term we used for Washington Post blog is as follows:

```
trump w/255 trump
```

The dates for each search was defined from June 15th, 2015, the day before Trump announced his candidacy, to May 4th, 2016, the day his last opponent, John Kasich, withdrew from the race. The duplicate filter option was set on 'high similarity' to avoid getting duplicated articles.

Source	Number of Articles
Associated Press	3,304
New York Times	2,791
Washington Post	2,780
USA Today	921
New York Times Blog	1,655
Washington Post Blog	7,904

Table 1. Number of News Articles by Source

Validating News Coverage

To verify the news articles we retrieved from Lexis/Nexis are relevant to Donald Trump's presidential campaign, we randomly selected 10% of the articles and conducted a validation check. We coded an article as relevant if the article (a) covered Donald Trump as the main focus of the article, (b) mentioned Donald Trump (but, not as the main focus of the article), or (c) mentioned Trump brand, such as Trump hotel, resort, book, etc. We coded an article as irrelevant if the word "trump" was used in another way (e.g., to play a trump card, to figuratively trump an opponent). Of the articles resulting from our search, our validation showed that over 96% were relevant to Trump (N = 1,908), and over 93% were relevant when excluding blog posts (N = 955).

Event Data

Three Event Categories

We classified Donald Trump's campaign activities into three categories. First, we considered coordinated campaign events in which the candidate interacted with the public, such as campaign rallies and town hall meetings, as *staged public events*. Second, we identified planned media events, such as press conferences and scheduled interviews, as *staged media events*. Third, we classified Trump's common practice of calling in to radio and television shows as *unscheduled media appearances*.

Validating Event Data

These event data were collected using Trumpshow.info, which collects videos of all Trump appearances in public and in media. The dates were validated using Trump's press releases to ensure that there were no missing events or appearances. Additionally, 10% of all staged media events and unscheduled appearances were verified with clips from the media outlets Trump appeared in. Our validation showed that 100% of all staged media events and unscheduled appearances were accounted for.

Twitter Data

Data Collection

We collected a sample of Tweets through the Twitter Streaming API from June 15th, 2015 to May 4th, 2016. Twitter describes this sample as about 1% random sample of the global tweets.

Tweet Retrieval

Donald Trump's tweets Because our sample potentially captures only 1% of Donald Trump's tweets, we retrieved his own tweets from his followers' retweets of his tweets. We compared this retrieved data with the data we directly downloaded from Twitter User Streams and found the two datasets to be identical with respect to all posted tweets.

Retweets of Donald Trump's tweets We found 'RT @realDonaldTrump' in the text of the tweets to identify retweets of Donald Trump's tweets.

Missing data and imputation

Our original Twitter data contained several missing values. In the *Trump tweets* series, there were seven missing values that consisted of one or two missing observations. In those cases, we used simple linear interpolation to fill in the missing data. The other Twitter-based variables, i.e., *RT @realDonaldTrump*, contained one additonal singlular missing value which were also filled with linear interpolation.

Our data also contained three cases where missing data stretched over more than two days. In those instances, we estimated a vector-autoregressive model (VAR) using *Trump tweets*, *RT @realDonaldTrump*, *Trump mentions*, and *Trump keywords* as endogenous variables, and variables measuring Trump's delegate count, the number of rally events, the number of national news headlines containing Trump, and the number of national news stories mentioning Trump twice, as exogenous variables. The VAR contained one lag and was estimated on each full period prior to the block of missing values. A dyanmic forecast was then estimated based on the VAR model to cover the missing days in the time series.

Statistical model building

In addition to the time series regression models reported in the paper, we also estimated VAR models in order to conduct Granger causality tests. Our VAR models' endogenous variables were *RT @realDonaldTrump*, Total number of Trump-related blog posts, total number of print stories mentioning Trump twice, and *Trump tweets*. We specified three exogenous variables which were counts of Trumps 1) unplanned appearances, 2) staged public appearances, 3) staged media appearances. Information criteria indicated a three-lag model was most appropriate.

Table 2 contains the estimates of the Granger causality tests based on the VAR model (which is estimated over both the pre-primary and primary periods). Several interesting results are worth noting. First, Trump's tweeting behavior is exogenous to both retweets and both types of news coverage. Retweets of Trump posts are driven, not surprisingly, by Trump's own Twitter activity. Also not surprisingly, both types of reporting (print and blogs) drive each other. Finally, both blog posts and print stories are driven by retweets of Trump but not Trump's Twitter activity.

Robustness Checks

We conducted a number of robustness checks to verify our results. First, we introduced measures of public opinion into our models. Our polling data came from major polls conducted during the period of analysis. We used only those polls which employed the registered voter screen for respondents. To date the poll for inclusion in our daily time series, we use the last day the poll was reported to have been in the field. We use the value of Trump support in the poll for that day. If multiple polls were present on the same day, we use the mean of the poll values. For our modeling, that value remains the measure of support for Trump until the next poll ends.

The inclusion of this measure of public support for Trump had little bearing on the core results. In the pre-primary period, the measure of public support for Trump was positively related (and statistically significant) to both coverage in print media and in blog posts. In the primary period,

Outcome Variable	Independent Variable	χ^2	p-value
RT of Trump	Blog Posts	6.20	0.102
RT of Trump	Print Stories	4.92	0.178
RT of Trump	Trump tweets	8.66	0.034
Blog Posts	RT of Trump	14.41	0.002
Blog Posts	Print Stories	20.42	0.000
Blog Posts	Trump tweets	4.67	0.197
Print Stories	RT of Trump	17.50	0.001
Print Stories	Blog Posts	37.47	0.000
Print Stories	Trump tweets	3.43	0.330
Trump tweets	RT of Trump	5.32	0.150
Trump tweets	Blog Posts	0.81	0.846
Trump tweets	Print Stories	1.64	0.649

Table 2. Granger Causality Estimates

however, public opinion had little influence on those outcomes. These results are somewhat sensitive to whether one lags the polling data to account for a delay in the release of the survey.

In addition, including an indicator for a primary debate on the day after the debate does not alter our core findings, but is a statistically significant (positive) predictor of both blog posts and print articles concerning Trump in both the pre-primary and primary period.

Finally, it is worth noting that in the time-series regression models, the estimates of ρ are consistently in the 0.41-0.55 range suggesting clear auto-correlation in the data. Tests for non-stationarity (Dickey-Fuller and KPSS tests), however, yield no evidence that unit roots are an issue in these series.