

The Pennsylvania State University

The Graduate School

College of the Liberal Arts

**AN ASSESSMENT OF VARIABLE NEWSPAPER CHARACTERISTICS ON  
THE COVERAGE OF MOVE-ON ANTI-IRAQ WAR VIGILS IN 2005: THE  
RELATIVE IMPACTS OF IRAQ WAR ISSUE ATTENTION, LOCATION  
TYPE, NEWSHOLE SIZE AND PROXIMITY TO A PROTEST**

A Thesis in

Sociology

by

Samuel Michael Sullivan

Copyright 2015 Samuel Michael Sullivan

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Master of Arts

May 2015

The thesis of Samuel Michael Sullivan was reviewed and approved\* by the following:

John D. McCarthy

Distinguished Professor of Sociology

Thesis Adviser

Roger Finke

Distinguished Professor of Sociology and Religious Studies

Director of the Association of Religion Data Archives (ARDA)

Lee Ann Banaszak

Professor of Political Science

Head of the Department of Political Science

John Iceland

Professor of Sociology and Demography

Department Head, Department of Sociology and Criminology

\*Signatures are on file in the Graduate School.

## **Abstract:**

With the rise of protest event analysis (PEA) newspaper data has become a widely used source of event data. Scholars rely on newspapers in order to create datasets of where, when, and what type of protest events take place. Alongside this expansion of the usage of newspaper data, scholars have identified selection bias as a serious concern, specifically citing that the print media cover only a small percentage of the total number of protests that actually take place.

Issue attention (the proportion of a newspaper's total output that relates to a single issue) is an overlooked aspect of the newspaper landscape that we demonstrate creates biases in PEA data. Additionally, using a public forums classification, we show that the type of space within which a protest event takes place has implications for its likelihood of gaining coverage. Events that took place in public and semi-public forums had significantly higher odds of being reported than events that took place in non-public forums. Validating previous results, we also report that the distance between a newspaper and an event, and newspaper circulation size are both positively associated with reporting likelihood.

Uniquely, the current research project relies upon a very large sample of US daily newspapers (N=424) and a sample of protest events that shows very little variation (N=1498); all events were supported by the same organisation (MoveOn.org), took the same form (anti-Iraq War Vigils), and took place at the same time (Summer 2005). Given the dyadic structure of the data, we had a total N of 635,043. We thus utilise a uniquely stable, and uniquely large, platform from which to analyse the dimensions of newspaper selection bias. The dataset was constructed from 1) geocoded reports of where MoveOn supported events were held; 2) reports of these events in US daily newspapers, found through NewsBank archives; and 3) a content analysis of US daily newspapers to determine how much attention they dedicated to the Iraq War.

## Contents

List of Tables .....	v
List of Figures .....	vi
Acknowledgements.....	vii
Introduction .....	1
Background and Hypotheses .....	2
Methods and Data .....	5
Analysis .....	10
Multivariate Results .....	11
Conclusion.....	14
Appendix: A Comparison of the Results of Logistic Regression and Rare Event Logistic Regression.....	15
Bibliography .....	16

## List of Tables

Table 1. Descriptive Statistics .....	7
Table 2. Distribution of the Coverage of MoveOn Protest Events.....	7
Table 3. Distribution of the Public Forum Space Variable .....	10
Table 4. The Effects of Environmental and Newspaper-Specific Variables on the Likelihood of Coverage of MoveOn Protest Events.....	13
Table 5. A Comparison of the Results of Logistic Regression and Rare Event Logistic Regression in Examining the Effects of Environmental and Newspaper-Specific Variables on the Likelihood of Coverage of MoveOn Protest Events.....	15

## List of Figures

Graph 1: A Graph to show the Bivariate Relationship between Coverage and Distance .....	9
---	---

## **Acknowledgements**

The enclosed thesis would have been impossible without the support of Professor John D. McCarthy, Professor Lee Ann Banaszak and Professor Roger Finke. Thank you for all of your illuminating comments and for dedicating so much time to the project.

Additionally, Hyun Woo Kim and Kevin Reuning were indispensable. They were guiding hands throughout the process and I am sincerely grateful.

Finally, my thanks to Vivek Upadhya for all of your proof reading.

## Introduction

The use of print media as a source of event data has increased dramatically in the social sciences, owing to it being continuous over time, rich in detail, geographically diverse, and easily available through large databanks, such as Newsbank and Lexis Nexis. Although undoubtedly of great utility, scholars have long been concerned with issues of media selection bias in regards to the widespread use of such data (Amenta, Caren, Olasky, Stobaugh 2009; Andrews, Caren 2010; Barranco and Wisler 1999; Hocke 1998; Hug and Wisler 1998; McCarthy, McPhail, & Smith 1996; Oliver & Meyer 1999). The problem of selection bias centres on whether the small group of events that receive newspaper coverage is representative of all of the events that actually take place. Previous literature has focused on defining the newspaper-specific and event-specific dimensions of selection bias (see below). However, this paper contends that these analyses have overlooked one key newspaper characteristic and one key event-specific measure: newspaper issue attention and the type of space within which an event takes place.

In examining these effects, we utilised data on MoveOn's anti-Iraq War vigil (2005) campaign to establish an un-mediated event list. MoveOn provided us with a complete list of the vigils (N=1498) that it supported in solidarity with Cindy Sheehan's 2005 anti-Iraq War protest outside of President George W. Bush's Texas ranch. The Iraq War was selected as the issue topic of interest because the war varied in intensity, endured over a long period, and attracted a great deal of media attention. Speaking to this media attention, a large sample of US daily newspapers (N=424) was then searched (via Newsbank) in order to determine if they reported any of these MoveOn vigils. Furthermore, characteristics of the daily newspapers were compiled from secondary sources, allowing us to examine whether previously under-examined newspaper-specific characteristics impact selection bias, specifically the impact of distance and circulation size on the odds of a vigil gaining newspaper coverage. Finally, the MoveOn event list was geocoded, allowing us to examine what effect public forum protection had on the likelihood of vigils being reported.

We report two major findings. Firstly, the more attention that a newspaper dedicated to a single issue-topic (the Iraq War) the higher the likelihood that protest events that fell within that issue category (MoveOn anti-Iraq War vigils) were to gain newspaper coverage. Secondly, we find that vigils that took place in spaces that most forcefully protect free speech (public forums and semi-public forums) saw an increased likelihood of gaining newspaper coverage, compared to events that took place in private spaces (non-public forums). Additional to the substantive findings, this paper offers three key methodological insights. Specifically, we have been able to hold the event variable constant (in terms of both event type and event issue) by examining the effect of one type of event (vigils), facilitated by the same organisation (MoveOn.org), in the same period (Summer 2005), on the same issue topic (Iraq War). This is of critical importance, as previous attempts to observe media biases have run into issues of disaggregating media selection bias from event type, which we have been able to avoid. Secondly, this paper examines selection bias not only through the lens of variation in events (Oliver & Myers 1999; McCarthy, McPhail, & Smith 1996) but, most importantly, with regards to variation in *newspaper characteristics* as well. Finally, by including so many newspapers (N=424) this project also fills a methodological gap, as it overcomes the field's reliance upon small, purposive samples of newspapers, providing a more systematic and representative

analysis of the American newspaper landscape.

## **Background and Hypotheses**

**Issue Attention (Newspaper-Specific):** The selection bias literature has grown dramatically over the last 25 years. However, this growth has been unequal, with the majority of published work that emanates from the social movements world tending to under-represent newspaper-specific characteristics. For this reason, research into newspaper-specific characteristics is still in its infancy, highlighting the importance of studies such as ours. Perhaps the most overlooked aspect of newspaper organisations in the literature (in relation to issues of selection bias or otherwise) is issue attention. Issue attention is a newspaper's pre-existing interest in a particular topic that affects how likely that newspaper is to report on similar topics. Downs (1972) hypothesized that the more a media outlet prioritizes reporting on a particular topic the higher the likelihood of events within the same issue field being reported. Conversely, the more distant an event from the topics in which the media outlet is currently interested, the less likely it is to garner coverage. Essentially, issue attention signals to a reader which stories/topics constitute news (what is newsworthy) through prioritizing particular issues over others. What this means for social movements is that "as social issues move in and out of vogue, the coverage of related social movements ebbs and wanes" (Myers and Caniglia 1999).

Recent studies have afforded some support to Downs's original theory (McCarthy, McPhail, & Smith 1996, Oliver & Meyer 1999, and Koopmans 2004). Writing the only paper that examines issue attention (and not issue attention cycles), McCarthy, McPhail, and Smith report that "when the volume of coverage for an issue is large, pertinent protests, even small ones, are more likely to attract coverage" (1996, P494). However, this conclusion is tentative, as it remains un-validated and McCarthy, McPhail, and Smith relied upon correlations rather than formal statistical analysis, as well as a sample of only 4 newspapers. Given the dearth of subsequent studies and the methodological weaknesses of these findings there is much need to examine issue attention in greater depth and over a greater range of media sources. In addressing this gap in the literature, we hark back to Downs's original logic by hypothesising that:

**H1:** The more extensive a newspaper's previous attention to the War in Iraq the higher the odds that it will report MoveOn anti-Iraq War vigils.

**Circulation Size (Newspaper-Specific):** The logic of the newshole (the total column inches available in any newspaper) dictates that larger newspapers should have more space available than small newspapers and so should be less susceptible to newshole constraints. Previous literature (Andrews & Caren 2010; Barranco & Wisler 1999; Kaniss 1991; Logan and Molotch 1987) has demonstrated that the structure of small, local newspapers' issue attention differs from those of larger newspapers. Typically, results have shown that small, local newspapers tend to focus on local issues to the detriment of national and international stories. However, because defining the boundary between local and national newspapers is somewhat arbitrary, we elected to examine circulation size, a close corollary. Using such logic, the literature holds that newspapers with small circulations tend to have smaller newsholes than larger papers. For this reason, local newspapers have less space to report information that is outside of the core area of interest for its readership, which previous

work has identified as local news. In addition, Meyers and Caniglia report that local events “are more salient to the local audience because they may affect travel in the city, influence local government action and resource distribution, or even provide opportunities for participation in further collective behavior” (2004, P538). Therefore, according to the logic of the newshole, we predict that:

**H2:** Newspapers with larger circulations (and hence, larger newsholes) will have higher odds of reporting MoveOn anti-Iraq War vigils than those with smaller circulations.

**Distance Effects (Newspaper-Specific):** Researchers have shown that as the distance between a newspaper and an event increases the odds of coverage decreases (Andrews & Caren 2010, Mueller 1997, Oliver & Myers 1999). The reasoning behind this is that newspapers have limited human resources, and these diminish as one moves farther from the head office. Reporters are thus clustered most densely within a fairly small area around the office and the number of available reporters falls dramatically with distance.

Perhaps the best known model for examining distance effects is Mueller's (1997) study on the West German, British, and American press. We have followed her design somewhat but rather than use only 6 newspapers, as Mueller does, we assess 424 American daily newspapers and over 1498 vigils (all geocoded), significantly increasing the validity and applicability of the findings to researchers. Additionally, we have incorporated the curvilinear functional form that Meyer and Caniglia (2004) demonstrate is essential to capturing the actual reporting behaviour of the US press, given that the population density is so much higher towards the coasts than in the centre of the country. Based on this literature we predict that:

**H3:** The further a MoveOn anti-Iraq War vigil is from a newspaper's headquarters the lower the odds that it will be covered.

**Spatial Effects (Event-Specific):** In contrast to newspaper-specific characteristics, work has progressed significantly in uncovering event-specific characteristics that significantly predict newspaper coverage. It has been widely reported that larger protest events attract greater media coverage than smaller events (Amenta et al 2009; Barranco & Wisler 1999; McCarthy, McPhail, Smith 1996; McCarthy et al 2008), and that protest events that use dramatic techniques or that encounter violence are more likely to be covered (Amenta et al 2009; Barranco & Wisler 1999; McCarthy, McPhail, Smith 1996; Oliver & Myers 1999). However, these two known event-level dimensions of selection bias were not included in the analysis. Event size could not be calculated reliably because MoveOn did not provide event-size information, and relying on news reports for such data would encompass serious selection effects, as we would be able to gather such information for covered events only. In addition, we have elected not to include a measure of disruptiveness because the activities, and therefore the level of disruption, of the protest events are fairly constant, given that event type is held constant.

Nevertheless, we could still examine one important, and under-analysed, event-specific characteristic; spatial effects. Researchers are beginning to report that the location of an event impacts its likelihood of coverage, both in terms of the absolute distance between an event and a newspaper's offices (Andrews & Caren 2010; Mueller 1997; McCarthy, McPhail, Smith 1996; Meyer and Caniglia 2004; Oliver & Myers 1999) and the type of space within which a protest event takes

place (Oliver and Myers 1999, Rafail forthcoming). Based on this, we argue that the spatio-legal environment within which events get reported is an important, yet largely overlooked, facet of the selection bias literature.

MoveOn vigil locations have been classified according to the public forums model, which is predicated on the ruling that “the existence of a right of access to public property and the standard by which limitations upon such a right must be evaluated differ depending on the character of the property at issue” (Perry Education Assn. 1983). The three-tier public forum model holds that free speech is protected to the greatest degree in public forums, less so in limited public forums, and least of all in non-public forums. Public forums are “places which by long tradition or by government fiat have been devoted to assembly and debate, the rights of the state to limit expressive activity are sharply circumscribed” (Perry Education Assn. 1983). These tend to include spaces such as parks and sidewalks. Limited public forums are state-owned or administered spaces that have been opened up for public interaction and expression, including spaces like schools and libraries. Non-public forum spaces are privately owned spaces or restricted state-owned spaces, like homes, businesses, and military bases.

In one of the few studies to include location as an independent variable, Oliver & Meyers find that events that took place in prominent public and limited public forum locations stood a much greater chance of being reported: “Most public events in Madison tend to occur in well-defined and narrowly limited public spaces, and news coverage focused on those limited areas” (Oliver and Myers 1999, P76/77). However, Rafail (forthcoming) finds that traditional areas of protest, and some of the most prominent public forums, have declined in importance as sites of protest. He reports that the proportion of protest activities taking place in universities and public parks has declined significantly since the 1960’s. He then goes on to demonstrate that privately owned public spaces host a disproportionately large share of protest activities. It is most likely that this changing use of space by social movement organisations (SMOs) can be attributed to a shifting of targets towards corporations and other similar institutions (and away from more public entities), coupled with an increasing privatization of the urban landscape. This migration of protest events from public to private spaces does not pose a significant problem for this research project, however. This is because in the case of MoveOn vigils most events occurred in public forum spaces. According to analysis of the data (Table 3), fully 73.43% of the vigils took place in public forums, with 2.34% taking place in limited public forums, and 24.23% taking place in private forum spaces.

McCarthy & McPhail also find that “a plurality of protests that target some element of the state are located in public forum spaces, and a significant minority of those protests occur in limited forum spaces, most of which are also places where state actors or state institutions reside” (2006, P240). This indicates that protest target and protest location are strongly correlated. MoveOn vigils primarily taking place in public and limited public forums is therefore a result of most governmental buildings being situated in constitutionally protected areas of free speech (public or limited forums). What is less clear, though, is what effect this mismatch between MoveOn vigil locations and the *public’s* use of space has on coverage likelihood. Do reporters inhabit public forums less, as the general public seems to, or do they maintain beats in public forums because protests continue to happen in them? It is hypothesised that when the target is the state, the more constitutionally protected the space that an event takes place in the more likely it is to be reported. This is because

the press is thought to have more access to spaces that are so classified and that because protests targeting state institutions are likely to continue to take place in public forum spaces reporters maintain beats in these areas. Thus, we hypothesise that:

**H4:** The type of public/private space within which a MoveOn anti-Iraq War vigil takes place will affect the odds that it will be covered, with public forum spaces having the highest odds of coverage and non-public forums the lowest.

## Methods and Data

Unlike many previous examinations of media selection bias the dataset that we have constructed is dyadic, with newspaper-event as the unit of analysis (N=635,043). This allows us to link each newspaper with each event, giving us increased insight into which newspapers reported on which events, greatly increasing our sample size (and thus our robustness), as well as the external validity of our findings. Below is a description of how event data and newspaper data was collected, followed by a description of what each variable captures, how each variable was constructed, as well as a description of its distribution.

**Event Data:** MoveOn.org is a non-profit organization that supports liberal causes throughout the US. In 2005, they helped coordinate a wave of vigils that sprang up in the wake of Cindy Sheehan's initial vigil at President Bush's Texas ranch. Sheehan began her campaign after losing a son in Iraq in April 2004, Army Specialist Casey Sheehan. Travelling to Crawford, Sheehan set up camp outside of President Bush's property on August 6<sup>th</sup> 2005. She demanded a second face-to-face meeting with the President (they had met one year previously, in June 2004) as well as an explanation of the motivations for war, given that no weapons of mass destruction had been found (the original justification for invasion). Sheehan's bold demonstration was quickly picked up by the media and garnered a great deal of commentary; positive and negative, local and national.

With such a salient message, and on the back of this media attention, anti-war vigils similar to Sheehan's spread across the country, many of which were coordinated and supported by MoveOn. The organization's documented involvement began on August 10<sup>th</sup> when MoveOn announced that it was "gathering comments via email to place in a two-page newspaper spread in a Sunday edition of the *Waco Tribune* newspaper in support of Sheehan and her efforts. MoveOn gathered more than 250,000 comments, many of which were included in the advertisement" ([http://en.wikipedia.org/wiki/Camp\\_Casey,\\_Crawford,\\_Texas](http://en.wikipedia.org/wiki/Camp_Casey,_Crawford,_Texas)). Through their website, MoveOn subsequently had a hand in the organization of at least 1498 protests across the country that summer. As has been mentioned, this event list was obtained from MoveOn indicating all of the anti-Iraq War vigils with which the organization was involved. This event data contains information on where vigils took place but no information on the happenings at the events (e.g. event size, police presence, duration of the event etc). Only 470 of these events were reported by at least one daily newspaper (427 in detail and 43 ambiguous mentions).

In utilising this data, we have followed the model of employing an independent event-list set forth by other researchers (Barranco & Wisler 1999; Hocke 1996; McCarthy, McPhail, & Smith 1996; McCarthy et al. 2008), allowing us to get an un-mediated count of the number of events that took

place. In addition, this independent dataset is the key to maintaining event type and issue. Exploiting MoveOn's anti-Iraq War vigil event-list means that all of the events in our data-set were organised in some capacity by the same organisation, at the same time, and as part of the same campaign. Thus, we have a uniquely unchanging baseline against which to assess newspapers' reporting rates. However, Hug & Wisler note a limitation of our approach: "non-systematic sources such as organizational records or countercultural press may inflate measures of protest activity for one movement" (1998, P 142). Since we have no way to uncover the true number of events that took place, we cannot speak to the "missingness" of our event list, which has to be borne in mind when reading our results.

**Newspaper Data:** All US daily newspapers available via Newsbank were included in our analysis (N=424). We have thus included newspapers from all 50 states, including Alaska and Hawaii. In addition to this geographic variation, there is a great deal of difference in newspaper size, with newspapers ranging from the large and long-established (e.g. New York Times) to the small and recently founded (e.g. Bluffton Today). Nevertheless, the total number of newspapers in the sample is some way short of the 1457 daily newspapers in print in 2005 (Editor's and Publisher's Yearbook, 2005). It is very difficult to correct for this validity problem, however, because we could not get access to more newspapers as a result of Newsbank's own sampling limitations (their failure to carry all of the US daily newspapers). This missing data reduces the analysable sample by around 18% but, as it is very likely missing at random it does not introduce any systematic bias (see sections below for more detail). In addition, we do not reduce the validity of our results by a great deal either because, by circulation, we have 50.988% of the total readership serviced by the newspapers in our sample  $[(27,852,807/54,626,138)*100]$ , with circulation numbers correct as of 30 September 2004.

Newspaper data was primarily used to construct the issue attention variable for *all newspapers* (see below). However, newspaper data was also used to collect descriptive information (participant numbers, counter demonstrations, celebrity speakers etc.) on *reported events*. If multiple newspapers reported the same event then data was triangulated before being coded. However, those events that went unreported contain no descriptive information other than that provided by MoveOn (location).

#### **Variables:**

Table 1 (below) shows summary statistics for all of the variables included in models 1-4. The statistics reported herein are the summary statistics for the sample that was actually analysed and not the summary statistics for the full sample (N=858,958). The analyzable sample is smaller than the full sample because, as mentioned above, some cases were removed because of missingness and data concerns. 18% of the issue attention data is missing due to NewsBank not maintaining information on certain newspapers, resulting in an N of 635,043.

**Table 1. Descriptive Statistics**

	Mean	Standard Deviation	Number of Observations	Minimum	Maximum
Covered	.0009385	.031	635041	0	1
Circulation	65736.84	95795.33	635041	3789	715052
Circulation (logged)	10.389	1.167	635041	8.24	13.48
Distance	1242.263	822.575	635041	0	5184.806
Distance <sup>2</sup>	2219845	2762944	635041	0	26882213.26
Public Forum	.734	.442	635041	0	1
Semi-Public Forum	.023	.151	635041	0	1
Non-Public Forum	.242	.428	635041	0	1
Issue Attention 1					
Year	1.8	1.019	635041	0	9.078

**Dependent Variable:** Reporting of MoveOn Vigils (Binary)

**Table 2. Distribution of the Coverage of MoveOn Protest Events**

	Frequency	Percent	Cumulative Percent
<i>Before Listwise Deletion</i>			
Covered	830	0.1	0.1
Not Covered	858128	99.9	100
Total	858958	100	100
<i>After Listwise Deletion</i>			
Covered	596	0.09	0.09
Not Covered	634445	99.91	100
Total	635041	100	100

Note: Missing data and data that is paired to the Issue Attention variable, which has the most missing data of all of the variables, has been listwise deleted for the final analysis

The coverage variable (the DV) captures whether or not a MoveOn anti-Iraq War vigil was covered by at least one daily newspaper. The dependent variable has been constructed as a newspaper-event variable, whereby it captures whether a specific newspaper reported on a specific vigil. Therefore, the complete matrix contains dyads of each newspaper by each vigil. By constructing our data in this way we get an accurate picture of exactly which newspapers reported on which vigils and can more easily tie independent newspaper and event variables to the dependent variable. This variable has a very small mean and standard deviation because only 0.09% of the dyads (594 newspaper-events) within our sample were reported. Table 2 shows the effect of reducing the complete sample to the analyzable sample (through listwise deletion) on the distribution of the dependent variable. The table shows that the proportion of events that were covered remains

constant after listwise deletion, meaning that the distribution of the dependent variable was not altered as a result of the sample size being limited. The proportion of covered events remains constant at about 0.1% of total events.

### **Independent Variable 1: Issue Attention Cycle (Continuous)**

The Issue Attention variable was constructed by searching Newsbank for the total number of articles published by each available US daily newspaper (N=424). Three forms of newspaper output were included in these searches, including standard articles, editorials and letters to the editor, and photo captions. A second search was then run on this same population of newspapers using the keyword "Iraq War." The latter was then divided by the former and turned into a percentage (see formula below). A larger percentage means that a greater proportion of that newspaper's total output was dedicated to the Iraq War. This, in turn, means that this newspaper has a more intense Iraq War issue attention. Searches were confined to three time periods (6 months, 9 months, and 1 year prior to the initial Sheehan protest on August 6<sup>th</sup> 2005) in order to get a reading of the consistency of Iraq War issue attention for each newspaper, as well as to test if the size of the window for measuring issue attention matters. Given that these three time periods turned out to be very highly correlated and that there were no obvious discrepancies between them, only issue attention data from 1 year prior to the Sheehan protest was included in the final analysis.

Issue Attention = (Total number of "Iraq War" articles/ Total number of articles published in time period)\*100

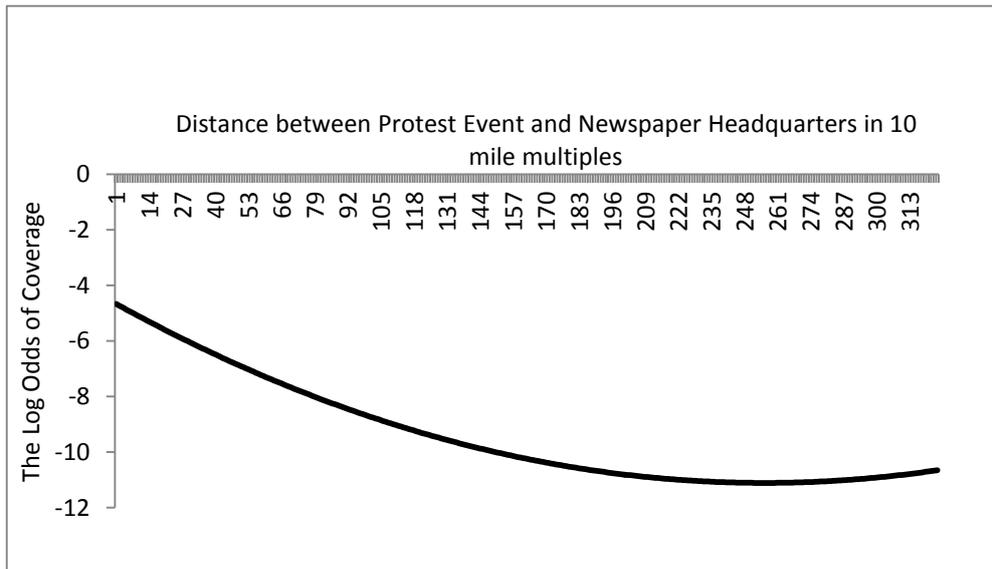
The issue attention variable ranges from 0 to 9.078%, meaning that some newspapers have absolutely no coverage of the Iraq War, whereas others dedicate up to 9% of their total output to the Iraq War. The mean level of attention, though, is 1.8% of total output, with a fairly large standard deviation of 1.019. From these results we can conclude that there is a great deal of variation in terms of how much attention newspapers devote to the Iraq War but that most newspapers devote less than 3% of total output to the issue.

### **Independent Variable 2: Distance (Continuous)**

Vigils were geocoded based on the textual location descriptions provided by MoveOn. A small number of vigils had poor descriptions and so the Python code employed for geocoding the majority of vigils could not determine an exact location. This sub-sample of vigils was isolated and locations were found using Google Maps as far as possible. Four vigils could not be located even using this method and so they were removed from the sample. Using the longitude/latitude geolocations, the distance between *each* vigil (N=1498) and *each* newspaper (N=424) was calculated, creating distance dyads. The below formula, which takes the curvature of the earth into account, was used to compute the values of these dyads.

=ACOS(COS(RADIANS(90- $\phi$ E11)) \*COS(RADIANS(90-ABA $\phi$ 2)) +SIN(RADIANS(90- $\phi$ E11))  
\*SIN(RADIANS(90-ABA $\phi$ 2)) \*COS(RADIANS( $\phi$ F11-ABA $\phi$ 3))) \*3958.756

**Graph 1: A Graph to show the Bivariate Relationship between Coverage and Distance**



In line with Meyer and Caniglia’s earlier finding (2004), distance was found to have a curvilinear bivariate distribution with coverage (see above). The bivariate distribution shows that as distance increases the odds of an event being covered by any newspaper is reduced. At around 2500 miles, however, there is a slight uptick in the odds of coverage. This is almost certainly down to population density effects, whereby events in large coastal cities (e.g. New York, DC, San Francisco, and LA) seem to be more newsworthy than events that take place in the middle of the country. This is not surprising given the vast gulf in population density between the coastal US and the centre of the country, as well as the increased density of publishing networks between the coasts compared to between the coasts and the centre of the country.

**Independent Variable 3: Public Forums Classification (Nominal-Categorical)**

Using the geocoded vigil locations provided by MoveOn, combined with Google Map (Street View) searches to visualize the locations, we made a determination of which type of public forum each vigil location constitutes. Each event has been coded as taking place in a public forum (1), limited public forum (2), non-public forum (3), or is classified as missing (4).

Table 3 (below) shows how the distribution of the Public Forum variable (the only categorical variable) is effected by the aforementioned listwise deletion. Much like coverage, the Public Forum dummy variables’ aggregate distributions remain fairly consistent after listwise deletion. Table 3 shows that in the analyzable sample the Public Forum accounts for 73.43% of cases, whereas before listwise deletion it accounted for 67.36%. The Semi-Public Forum accounts for 2.34% of the analyzable sample, and 2.14% of the total sample. Finally, the Non-Public Forum accounts for 24.23% of the analyzable sample and 22.23% of the total sample. The primary reason for the small differences between the analyzable and total sample is that the 8.27% of the total sample that was missing has been listwise deleted, along with data that matched the missing data on the Issue Attention variable. When viewed in unison with the largely unaltered distributions of the other variables (see Tables 2 and 3 above), the maintenance of the distributions of these variables strongly

implies that the values missing on the Issue Attention variable are missing at random, which means that the reduction in the size of our sample has not affected its representativeness.

**Table 3. Distribution of the Public Forum Space Variable**

	Frequency	Percent	Cumulative Percent
<i>Before Listwise Deletion</i>			
Public Forum	578600	67.36	67.36
Semi-Public Forum	18410	2.14	69.5
Non-Public forum	190938	22.23	91.73
Missing	71010	8.27	100
Total	858958	100	100
<i>After Listwise Deletion</i>			
Public Forum	466334	73.43	73.43
Semi-Public Forum	14837	2.34	75.77
Non-Public forum	153872	24.23	100
Total	635043	100	100

Note: Missing data and data that is paired to the Issue Attention variable, which has the most missing data of all of the variables, has been listwise deleted for the final analysis

**Independent Variable 4: Circulation Size (Continuous)**

Data on the circulation size of each newspaper was collected from Editor’s and Publisher’s Yearbook 2005. The variable was logged in order to counter a slight right skew. Nevertheless, the transformed and untransformed summary statistics were included in Table 1 in order to aid interpretation, as well as to provide a representation of what the untransformed variable looks like. As can be seen, newspaper circulation size ranges from 3789 to 715,052, with a mean of just over 65,000. The untransformed variable also has a very large standard deviation of over 95,000. Once transformed, however, the standard deviation becomes much smaller (as a ratio of the sample size), speaking to the improved fit of this functional form with the data.

**Analysis**

Given the vast discrepancy in the number of non-reports versus reports (see Table 2) the applicability of regular logistic regression was a cause for concern. Logistic regression uses maximum likelihood as its estimation technique, which is known to be biased when dealing with small-samples (Potter 2005). Therefore, at the outset, it was unclear whether the small sample of covered events (approximately 0.09%% of the total sample) was large enough to allow the maximum likelihood function to perform at its optimum level. This was especially worrisome given that in some cases the absolute size of the discrepancy between the number of ones and zeros is as important as the relative difference.

There are a number of ways to counter the aforementioned small sample bias, primary among which is to make the events less rare. This can be done by altering the definition of "event" or altering the

time-frame, amongst other measures. None of these strategies was applicable to this project, however, because our event type and window were necessarily limited by the research question and by data availability. A second method, proposed by King and Zeng (1999) concerns sampling on the dependent variable and limiting the zero events to a small, randomly selected sample, rather than the entire universe of non-events. Their argument holds that by limiting the number of zeros researchers lose very little information but increase the balance of their overall sample (a better ratio of 1's to 0's). Again, this strategy is not pertinent to the project at hand, as the zeros mean more in our project than they do for the purposes intended by King and Zeng; In this project, all dyads stand a reasonable chance of reporting (to be a 1) whereas in the study of war, King and Zeng's intended purpose for this methodology, not all dyads stand a reasonable chance of fighting (being a 1). The authors give the example that Burkina Faso and St Lucia are highly unlikely to ever go to war (King and Zeng 2001, P694).

For our purposes, then, the ultimate solution to the potential problem of small sample bias lay in the use of rare event logistic regression (RELogit). What can be read most clearly from King and Zeng's results is that the greatest bias is introduced when there is a small sample size and a small number of ones (with the effect decreasing as both parameters increase). However, in seriously unbalanced samples "no sample size is large enough to evade finite sample problems if the events are sufficiently rare" (2001 P153). The question, then, was whether our events were common enough not to cause finite sample problems. Only comparison of the logistic regression and RELogit results would inform the answer, requiring both forms of logistic regression to be used. The results of the comparison (see Table 5 in appendix) showed that there was almost no difference between the results of the logistic regression and rare event logistic regression and so only the results of the logistic regression are reported herein.

## **Multivariate Results**

The stepwise analytical approach that was followed resulted in four models of increasing complexity (Table 4). Model one shows that as logged circulation increases by 1 unit, the odds of that newspaper reporting a MoveOn anti-Iraq War vigil increased significantly by 1.484 ( $p < .01$ ). The Pseudo  $R^2$  for this model was .0140, meaning that 1.4% of the total variation is explained by circulation size alone.

We then added distance to the specification, resulting in model 2. The results show that distance has a curvilinear effect on coverage, as the lower-order term has a significant coefficient of 0.995, meaning that as distance increases by 1 mile the odds of an event being covered are reduced by 0.005 ( $p < .01$ ). However, as distance reaches its upper values the odds of coverage increase significantly. The distance<sup>2</sup> variable captures the higher order effect, whereby a 1 mile increase in distance results in the odds of coverage significantly increasing by 1.000001 ( $p < .01$ ). Table 4 shows 1.000 as the coefficient for distance<sup>2</sup> because it has been rounded to three decimal places. The standardized coefficient (standardized on both X and Y variables) shows that distance has a much larger effect relative to circulation size. At lower values, as distance increases by one standard deviation coverage *decreases* by 1.546 standard deviations, but at higher values as distance<sup>2</sup> increases by 1 standard deviation coverage *increases* by 1.019 standard deviations. The addition of the distance variables increased the Pseudo  $R^2$  by exactly 14%, by far the largest single increase in

explained variance, to a Pseudo R<sup>2</sup> of .1540. A likelihood ratio test was run in order to test whether the addition of the distance variables had an impact on the model's predictive power. The likelihood ratio test gave a significant result (Prob>chi2 = .000), meaning that distance has a significant impact on the model.

Model 3 shows the results of the addition of the Public Forum location dummy variables. The coefficients for circulation and distance remain very similar and significant and so will not be discussed here. An event that took place in a Public Forum increased the odds of being covered by a daily newspaper by 1.964 ( $p < .01$ ) compared to events that took place in a Non-Public Forum (omitted category). Interestingly, events that took place in Semi-Public Forums had an even larger coefficient, increasing the odds of coverage by 2.415 ( $p < .01$ ) compared to events that took place in a Non-Public Forum. The relative effects of these dummy variables tell a different story, however. Events that took place in a Public Forum saw a .114 standard deviation increase in the odds of being covered compared to events that took place in Non-Public Forums. Events that took place in Semi-Public Forums, on the other hand, only saw only a .051 standard deviation increase in the odds of being covered compared to events that took place in Non-Public Forums. We re-ran the analyses with different omitted categories and the substantive findings remained consistent, bolstering the robustness of these findings. The addition of the Public Forum dummy variables to the model specification added only a small amount to the Pseudo R<sup>2</sup> value. Model 3 explains 15.81% of the total variation present in the sample, an increase of .0041 (0.41%) over model 2. Nevertheless, as in model 2, a likelihood ratio test was run and the results show that the addition of the Public Forum dummy variables significantly increase the predictive power of the model (Prob>chi2 = .000).

The final specification, model 4, includes the addition of the Issue Attention variable. None of the aforementioned substantive findings changed after this addition and so I will focus only on the Issue Attention results here. As a newspaper's issue attention increases by 1%, the odds of that newspaper reporting a MoveOn anti-Iraq War vigil increase by 1.198 ( $p < .01$ ). When I examined the relative effect of issue attention on the odds of coverage it became clear that issue attention has an effect about half the size of circulation: a one standard deviation increase in issue attention results in a .071 standard deviation increase in the odds of coverage. The Pseudo R<sup>2</sup> of model 4 is .1610, a .0029 (.29%) increase over model 3. Testing that the addition of issue attention to the model specification had an impact, the results of the likelihood ratio test show that the inclusion of the issue attention variable had a significant impact on the model's predictive power (Prob>chi2 = .000).

In addition, we ran Wald tests on model 4 for each of the predictor variables as a robustness check to the likelihood ratio tests. All of the variables produced significant results (Prob>chi2 = .000), verifying that each addition to the model significantly impacted its predictive power. Finally, a variance inflation factor (VIF) test was run in order to examine how related the predictor variables were. The mean VIF was 3.26, which is well under the generally accepted cutoff of 5. This indicates that the predictor variables are not highly correlated, increasing the robustness of the findings yet further.

**Table 4. The Effects of Environmental and Newspaper-Specific Variables on the Likelihood of Coverage of MoveOn Protest Events**

	Model 1		Model 2		Model 3		Model 4	
	Coefficient (Odds Ratio)	Standardised Coefficient	Coefficient (Odds Ratio)	Standardised Coefficient	Coefficient (Odds Ratio)	Standardised Coefficient	Coefficient (Odds Ratio)	Standardised Coefficient
Circulation (logged)	1.484*** (.051)	-	1.419*** (.049)	.157	1.418*** (.049)	.156	1.363*** (.049)	.138
Distance			0.995*** (.000)	-1.546	0.995*** (.000)	-1.535	0.995*** (.000)	-1.527
Distance <sup>2</sup>			1.000*** (.000)	1.019	1.000*** (.000)	1.012	1.000*** (.000)	1.008
Public Forum					1.964*** (.235)	.114	1.961*** (.234)	.114
Semi-Public Forum					2.415*** (.631)	.051	2.399*** (.627)	.051
Issue Attention (1 Year)							1.198*** (.038)	.071
Constant	0.0000139 (.000)		0.0002477 (.000)		0.0001429 (.000)		0.0001501 (.000)	
Number of Obs	635043		635043		635043		635043	
Pseudo R <sup>2</sup>	.0140		.1540		.1581		.1610	

\*= $p < .10$  \*\*= $p < .05$  \*\*\*= $p < .01$  (two tailed tests)

## Conclusion

This paper concludes with two novel contributions to the literature (H1 and H4 confirmed) and two validations of known findings (H2 and H3 confirmed), reached through analysis of a particularly large and representative dataset. Firstly, our findings demonstrate that audience preferences and newspaper-specific characteristics interact to define the salience of an issue. Specifically, our models' validation of the positive relationship between issue attention and coverage likelihood demonstrate that the audience's appetite and/or a newspaper's own preference for an issue help to shape the landscape for reporting similar news, confirming hypothesis 1. We report that as issue attention increases the likelihood of related protest events gaining newspaper coverage also increases. In relative terms, issue attention is about half as impactful as circulation size on garnering newspaper coverage. That issue attention is significantly related to the odds of coverage, and that it has such a large effect, is therefore of great use to scholars utilising protest event analysis, as it is an important, but previously undetected, dimension of selection bias that must be accounted for.

Secondly, we have shown that location impacts protest event coverage. We demonstrate that public and semi-public forum spaces are correlated with an increased likelihood of event reporting, confirming hypothesis 4. However, it remains an open question as to whether journalists maintain beats in public and semi-public forums because protests targeting state institutions and issues continue to take place there, or whether protesters target these forums partially because reporters are already in the vicinity for other reasons (e.g. covering court cases or state politics). Future researchers would do well to untangle this empirical puzzle.

Thirdly, with an unprecedentedly large sample, our findings validate the literature's conclusions that increasing circulation size and distance between a protest event and a newspaper are strong, positive predictors of the likelihood of a protest event gaining coverage, confirming hypotheses 2 and 3. Our findings also throw weight behind Myer and Caniglia's result that, in the US at least, distance and coverage share a curvilinear relationship, with events taking place in coastal areas being slightly more likely to be covered by newspapers of the opposite coast than distant events in the heartland.

In order to enhance our understanding of issue attention, future research should concentrate on examining whether issue attention functions in the same manner for protest events of a different nature (e.g. marches\divestment etc.), given that this article was limited (by design) in its scope of event types. In addition, researchers should examine whether issue topic impacts the selection odds of particular events versus other similar ones. Researchers should focus specifically on whether less politicised issue-topics display similar issue attention patterns as those found herein (for a highly political issue-topic). This would go some way to determining whether our results are unique to the Iraq War issue or whether they also extend to less salient issue topics.

## Appendix: A Comparison of the Results of Logistic Regression and Rare Event Logistic Regression

**Table 5. A Comparison of the Results of Logistic Regression and Rare Event Logistic Regression in Examining the Effects of Environmental and Newspaper-Specific Variables on the Likelihood of Coverage of MoveOn Protest Events**

	Model 1		Model 2		Model 3		Model 4	
	Logistic Regression	Rare Event Logistic Regression						
Circulation (logged)	.395*** (.034)	.395*** (.033)	.350*** (.035)	.350*** (.033)	.349*** (.035)	.349*** (.033)	.310*** (.036)	.309*** (.035)
Distance			-.005*** (.000)	-.005*** (.000)	-.005*** (.000)	-.005*** (.000)	-.005*** (.000)	-.005*** (.000)
Distance <sup>2</sup>			.000*** (.000)	.000*** (.000)	.000*** (.000)	.000*** (.000)	.000*** (.000)	.000*** (.000)
Public Forum					.675*** (.119)	.670*** (.119)	.674*** (.119)	.669*** (.119)
Semi-Public Forum					.882*** (.261)	.903*** (.261)	.875*** (.261)	.896*** (.261)
Issue Attention ( 1 Year)							.181*** (.031)	.183*** (.023)
Constant	-11.186 (.376)	-11.184 (.367)	-8.303 (.387)	-8.303 (.360)	-8.853 (.401)	-8.847 (.377)	-8.804 (.404)	-8.798 (.384)
Number of Obs	635043	635043	635043	635043	635043	635043	635043	635043

\*= $p < .10$  \*\*= $p < .05$  \*\*\*= $p < .01$  (two tailed tests)

Note: All coefficients are Log Odds

## Bibliography

Amenta, E. and Caren, N. and Olasky, S.J. and Stobaugh, J.E. 2009. "All the Movements Fit to Print: Who, What, When, Where, and Why SMO Families Appeared in the New York Times in the Twentieth Century." *American Sociological Review* 74: P636–56.

Andrews, K. T. and Caren, N. 2010. "Making the News: Movement Organizations, Media Attention, and the Public Agenda." *American Sociological Review* 75(6): P841-866

Barranco, J. & Wisler, D. 1999. "Validity and Systematicity of Newspaper Data in Event Analysis." *European Sociological Review* 15(3): P301-322

Baumgartner, F. R. and Jones, B.D. 1993. *Agendas and Instability in American Politics*. Chicago, IL: University of Chicago Press.

Berry, J. M. 1999. *The New Liberalism: The Rising Power of Citizen Groups*. Washington D.C.: Brookings Institution Press

Bogart, L. 1984. "The Public's Use and Perception of Newspapers." *Public Opinion Quarterly* 48(4): P709-719.

Danzger, H. M. 1975. "Validating Conflict Data." *American Sociological Review* 40(5): P 570-584

Downs, A. 1972. "Up and Down with Ecology – The issue Attention Cycle." *Public Interest* 28: P38-50

Earl, J. and Martin, A. and McCarthy, J.D. and Soule, S.A. 2004. "The Use of Newspaper Data in the Study of Collective Action." *Annual Review of Sociology* 30: P65-80

Editor & Publisher. 2005. *International Yearbook: The Encyclopedia of the Newspaper Industry 2005*. New York, NY: Editor and Publisher

Edwards III, G.C. and Wood, B.D. 1999. "Who Influences Whom? The President, Congress, and the Media." *American Political Science Review* 93: P327–44.

Firth, D. 1993. "Bias Reduction of Maximum Likelihood Estimates." *Biometrika* 80(1): P27-38

Grosswindhager, S. 2009. "Using Penalized Logistic Regression Models for Predicting the Effects of Advertising Material." Available at: [http://publik.tuwien.ac.at/files/PubDat\\_179921.pdf](http://publik.tuwien.ac.at/files/PubDat_179921.pdf) (09/12/2014)

Hazlitt, W. 1917. *Selected Essays, 1778-1830*. Edited by George Sampson. Cambridge: Cambridge University Press

Hilgartner, S. and Bosk, C.L. 1988. "The Rise and Fall of Social Problems: A Public Arenas Model." *American Journal of Sociology* 94: P53–76.

Hocke, P. 1998. "Determining the Selection Bias in Local and National Newspaper Reports on Protest

Events." In: *Acts of Dissent: New Developments of the Study of Protest*, Dieter Rucht, Ruud Koopmans, and Friedhelm Neidhardt (eds.). Lanham, Boulder, New York, Oxford: Rowman & Littlefield 1999: P131-163

Hug, S. and Wisler, D. 1998. "Correcting for Selection Bias in Social Movement Research." *Mobilization* 3(2): P141-161

Johnson, M. A. 1997. "Predicting News Flow from Mexico" *Journalism and Mass Communication Quarterly* 76(2): P315-330

Kaniss, P. 1991. *Making Local News*. Chicago, IL: University of Chicago Press.

King, G. and Zeng, L. 1999. "Logistic Regression in Rare Events Data." Department of Government, Harvard University.

King, G. and Zeng, L. 2001. "Explaining Rare Events in International Relations." *International Organization* 55(3): P693-715

Koopmans, R. 2004. "Movements and Media: Selection Processes and Evolutionary Dynamics in the Public Sphere." *Theory and Society* 33: P367-391

Logan, J. R. and Molotch, H. L. 1987. *Urban Fortunes*. Berkeley, CA: University of California Press

McCarthy, J.D. and McPhail, C. and Smith, J. 1996. "Images of Protest: Dimensions of Selection Bias in Media Coverage of Washington Demonstrations, 1982 and 1991." *American Sociological Review* 61(3): P478-499

McCarthy, J.D. & McPhail, C. 2006. "Places of Protest: The Public Forum in Principle and Practice." *Mobilization: An International Quarterly* 11(2): 229-247

McCarthy, J.D. and Titarenko, L. and McPhail, C. and Rafail, P.S. and Augustyn, B. 2008. "Assessing Stability in the Patterns of Selection Bias in Newspaper Coverage of Protest during the Transition from Communism in Belarus." *Mobilization: The International Quarterly* 13(2): P127-146

Myers, D. J. & Caniglia, B. S. 2004. "All the Rioting That's Fit to Print: Selection Effects in National Newspaper Coverage of Civil Disorders, 1968-1969" *American Sociological Review* 69(4): P519-543

Mueller, C. 1997. "International Press Coverage of East German Protest Events, 1989." *American Sociological Review* 62(5): P820-832

Oliver, P.E. and Myers, D.J. 1999. "How Events Enter the Public Sphere: Conflict, Location, and Sponsorship in Local Newspaper Coverage of Public Events." *American Journal of Sociology* 105(1): P38-87

Perry Education Assn., Appellant v. Perry Local Educators' Assn., et al: 460 U.S. 37, Feb 1983

Potter, D.M. 2005. "A Permutation Test for Inference in Logistic Regression with Small-and Moderate-Sized Data Sets." *Statistics in Medicine* 24(5): P693-708.

Rafail, P. 2015. "Shifting the Ground Under Our Feet': The Distribution of Protest Spaces in New York, 1960-2006." Unpublished Manuscript.

Schoenfeld, A.C. and Meier, R.F. and Griffin, R.J. 1979. "Constructing a Social Problem: The Press and the Environment." *Social Problems* 27: P38–61.

Smith, J. and McCarthy, J.D. and McPhail, C. and Augustyn, B. 2001. "From Protest to Agenda Building: Description Bias in Media Coverage of Protest Events in Washington, D.C." *Social Forces* 79(4): P1397-1423

Wilkes, R. and Ricard, D. 2007. "How Does Newspaper Coverage of Collective Action Vary? Protest by Indigenous People in Canada." *The Social Science Journal* 44: P231–251

2014. "Camp Casey, Crawford Texas." Wikipedia , July 17<sup>th</sup>. Retrieved September 12<sup>th</sup>2014 ([http://en.wikipedia.org/wiki/Camp\\_Casey,\\_Crawford,\\_Texas](http://en.wikipedia.org/wiki/Camp_Casey,_Crawford,_Texas))