Studying Media Events in the European Social Surveys across Research Designs, Countries, Time, Issues, and Outcomes

William Pollock*
Jason Barabas*
Jennifer Jerit*
Martijn Schoonvelde**
Susan Banducci**
Daniel Stevens**

September 30, 2014

Abstract

Scholars often study isolated media effects in single country using one method at one moment in time. We seek to generalize the research in this area by examining hundreds of press-worthy events across dozens of countries at various points in time with an array of techniques and outcome measures. In particular, we merge a database containing thousands of national events with five waves of the European Social Survey to conduct analyses across countries and individuals as well as within countries and subjects. The results suggest that there is an impressive degree of heterogeneity when it comes to how citizens react to political developments. In particular, some events generate significant opinion changes when groups of individuals who are "treated" are compared to "control" cases. However, other events produce modest or even null findings with methods that employ different counterfactuals. Thus, findings of both strong and weak media effects that scholars have uncovered over the years could be a function of methodological choices as well as context-specific factors such as institutional arrangements, media-systems, eras, or event characteristics. Data limitations also make some research designs possible while they preclude others. We conclude with advice for others who wish to study political events in this manner as well as discussion of media effects, broadly construed.

* Department of Political Science, Stony Brook University, Stony Brook, NY 11794, USA
** Politics Department, University of Exeter, Exeter, Devon, United Kingdom, EX4 4RJ

Note: The authors of this paper wish to acknowledge the receipt of generous research support from the Economic and Social Research Council (ESRC). John Berry Ryan provided valuable comments as did panelists at the 2014 London Media Effects Research workshop. David Martin and Matt Harris provided helpful research assistance. Please address correspondence to William Pollock at william.pollock "at" stonybrook.edu.

Numerous media effects studies exist. Some employ statistical analyses of cross-sectional datasets to arrive at their conclusions (e.g., Dalton, Beck, and Huckfeldt 1998; Druckman and Parkin 2005; Kahn and Kenney 2002). Others use panel studies (Ladd and Lenz 2009; Patterson and McClure1976) or coverage variations in natural settings (Lassen 2005; Prior 2007; Finserass and Listhaug 2013). Experimental approaches are also growing in popularity (Iyengar, Peters, and Kinder 1982; Berinsky and Kinder 2006; Neuman, Just, and Crigler 1992). Yet another class of studies employs hybrid designs, such as comparing individuals to themselves within a given survey as if they were panelists (Barabas and Jerit 2009) or comparisons of survey experiments to actual media events in the natural world (Barabas and Jerit 2010). On top of all of this, the domains of inquiry vary too—both geographically and temporarily—with some focusing on one country over time (e.g., Soroka 2006; Kellstedt 2000; Stevens and Banducci 2013) and others comparing across countries (e.g., Soroka et al. 2012; Fraile 2013; Iyengar et al 2010; 2009).

Given the methodological heterogeneity underlying the study of media effects, perhaps it is not surprising that the findings are also quite mixed. There has been an evolution over the years from minimal effects (Klapper 1960; McGuire 1986) to massive effects (Bartels 1993; Zaller 1996). Yet, some wonder whether a new era of minimal effects may be upon us (Bennett and Iyengar 2008). Complicating matters further, it could also be that media effects exist but that most surveys lack features needed to reveal the effects. For instance, media effects might be hard to detect without sufficiently detailed measures of exposure (e.g., Barabas and Jerit 2010; Dilliplane, Goldman, and Mutz 2013; Druckman and Parkin 2005), but some critics believe media exposure measures are deeply flawed (Prior 2009) while others argue that exposure analyses may prove to be futile because many media studies lack statistical power (Zaller 2002).

In this paper, we adopt a broad view of media effects research in an attempt to pit various designs against each other using various types of data in a diverse set of countries over many years. While we identified some noteworthy patterns, on balance we find that casting a wide net tends to yield very little in the way of statistically significant media effects. However, the lack of significance stems most notably from data issues related to the number of observations, the timing of the inquiry, and (most importantly) the design choices that lead to alternative counterfactuals. In the end, some subtle but important ways in which media effects data are collected and analyzed may help scholars better document their existence.

Media Effects Heterogeneity due to Designs, Data, and Context

In an ideal world, at least from the vantage point of a media effects researcher, news stories of varying levels of importance and on various topics would be randomly assigned to a diverse set of citizens. In such a world, we would also see randomly distributed variation across types of media outlets, types of stories, and temporal eras. To complete the vision of this scholarly utopia, data to evaluate the effects would be plentiful and of high quality.

Unfortunately, the real world departs from this ideal in several ways. Decisions about what appears in the news are often left to journalist and their employers (Dunaway 2008), although sometimes everyday people have input as “citizen journalists” in some public journalism schools of thought (Rosen 2001). It is also the case that data availability constrains analytical latitude—with inadequate measures of individual-level characteristics as well as the information environment they inhabit. That is, studies frequently lack variables thought to be important in media effects research such as media exposure (e.g., Jerit, Barabas, and Bolsen 2006) or they study media effects without including information on the nature of the media-inspiring event.

Perhaps for these reasons, single-country analyses abound. They are helpful because they allow analysts to go into depth when it comes to what they can say about the effects of any particular event. For example, Zaller and Hunt (1995) studied Ross Perot's paranoia during the 1992 campaign. Lenz and Ladd (2009) focus on an editorial endorsement change during the British election. Stevens et al. (2011) focus on newspaper endorsements of Prime Minister Tony Blair in the mid-2000s. But these studies often focus single media events at a singular moment in time in a particular country.[1] Occasionally scholars conduct comparisons of a few countries within specific years (Curran et al. 2009; Iyengar et al. 2010), sometimes with a focus on well publicized events (e.g., Finseraas and Listhug 2013; Finseraas, Jackobsson, and Kotsadam 2011), but studies with additional countries spanning multiple years are exceptions rather than the rule (cf. Schoonvelde 2013). In this study, we adopt a broad view, calculating and comparing media effects across a range of methodological, system, and issue characteristics.

Methodological Factors

Early on during empirical projects on media effects, and typically at the outset, scholars select a research design or designs. A common choice is the comparison of those who report being exposed to the media to others who report less or no exposure within a given country (e.g., Eveland et al. 2008; Hutchings 2001; Stevens and Karp 2012). Given its popularity, this design is the baseline against which we wish to compare other possible choices (we call this the "media

[1] Events are used to study the effects of media on public opinion. For instance, Smetko et al. 2003 used the June 1997 Amsterdam Summit, known as "Eurotop" in the Dutch press, to determine how it altered attitudes toward the European Union. They found that attentiveness strongly determined whether or not opinions changed. In contrast, Statham and Tumber failed to find linkages between events related to gay rights in Ireland and public support for allowing gay men and lesbians "…to live their own life as they wish" (Statham and Tumber 2013, 749-51). However, that same study suggests there could be a link between opinion movement on European unification in Ireland and negative evaluations detected in media claims, but the analysis was cast as speculative (p. 749).

exposure" design or model). Assuming the data are observational—and hence, already collected—then a few alternatives to this basic design exist.

One alternative is to compare treated individuals to their untreated selves in a technique known as within-survey/within-subjects (WS/WS) comparisons (see Barabas and Jerit 2009).[2] This technique pushes the logic of counterfactual inference (e.g., Morgan and Winship 2007) to its logical end by comparing individuals to themselves rather than engaging in comparisons with other survey respondents via "controls." The idea is to identify questions on the same topic, one of which receives media coverage and the other which receives little or no coverage ("the baseline"). By looking at the differences in the outcome measure when there is coverage versus when there is less or none, researchers can identify media effects that control for all individual-level characteristics, measured or not.

Designs like WS/WS are attractive because media exposure measures are not needed—i.e., we see differences in outcomes for individuals at any given level of media exposure, whatever they choose, no matter how accurately they report it in a survey, etc. However, comparable outcome items needed in a WS/WS analysis are often not available in cross-sectional surveys. For this reason, researchers might be tempted to explore other design variants, such as differences-in-differences (DID) approaches (see Wooldridge 2013, Chapter 13). In DID models, researchers have data before and after some key event along with some way (e.g., geography, media exposure, etc.) to differentiate those who are exposed to the relevant messages as well as those who are not. Even if the two groups start off with baseline differences in the dependent variable before the event, assuming they both react similarly, then researchers can take the

[2] For earlier within-subjects panel designs on media effects, see Lazarsfeld and Fisk 1938 or Lazarsfeld, Berelson, and Gaudet 1944.

difference in the changes between the two groups as an estimate of the effect (e.g., Barabas and Jerit 2010; Fair, Malhotra, and Shapiro 2012; Keele, Malhotra, and McCubbins 2013).

Each design choice has subtle but important ramifications. For example, identifying media effects—from a statistical standpoint—depends on having sufficient statistical power to reject null hypotheses of no effect. Zaller (2002) demonstrated that detecting significant media effects of even 5 or 10 percentage points is often very difficult without thousands of observations—far more than most studies or designs often permit. On top of this, if survey respondents are harder to reach now than in the past (e.g., Keeter el al. 2006; 2007; National Research Council 2013), it could be that more recent studies are smaller or conducted differently than in the past. Thus, *when* the survey is conducted may matter as much as how large it is or what designs are used. We will consider all of these factors simultaneously and in relation to other possible determinants of media effects discussed next.

Country-level Factors

Aside from the designs employed, countries vary on many dimensions in ways that might accentuate or diminish media effects. For instance, some countries have relatively free and open media systems with journalists in control of producing and distributing their own content (e.g., Hallin and Mancini 2004). The shear availability of media—in both quantity and quality—might mean greater media effects. In other countries, however, government authorities have a greater role in the media system. Thus, media system freedom could relate to media effects. Scholars have found strong positive effects of media system freedom on political knowledge (Schoonvelde 2013), but freer systems may have so many information access points that the media could be irrelevant. In other words, new media and alternative sources of information (e.g., political discussion) may make the media system characteristics less important than previously thought. Even people who are unexposed to traditional media outlets may learn about important news events, perhaps ushering in a new minimal effects era.

Countries also vary in their style of government and electoral rules. Some are representative democracies while others employ parliamentary systems. Also, some nations require everyone to vote—presumably increasing the likelihood that public affairs are covered and followed by the populace—while other governments allow people to check out of politics. Especially when coupled with high levels of choice regarding what to watch, citizens in "post-broadcast" democracies (Prior 2007) can tune out politics, which could diminish the impact of the mass media.

Beyond these factors, countries differ for reasons either relating to—or in spite of—their institutional configurations. For example, some are wealthy and others are not. Wealth is often a marker of other differences related to education or socio-economic status. Race, gender, and immigration all conspire to produce different political dynamics. Still, most of these factors would be associated with wealth per capita, making this an important catch-all variable in comparisons across countries. In some ways, the number of ways countries could vary is limitless; no study could ever hope to control for all relevant differences. Thus, designs like those discussed earlier are one way to contend with the possibility of spuriousness or selection in non-experimental settings.

Issue Factors

Aside from methodological or country-level factors, the topics being studied might have differential effects. Economic considerations often predominate in elections (e.g., Hetherington 1996). Scandals are often pivotal too (Miller 2010), but sometimes natural catastrophes are just as devastating as man-made ones (Maestas et al. 2008; Gomez and Wilson 2008). Still other distinctions revolve around whether the event in question is an election and whether protest movements like the "Occupy Wall Street" campaign exist to galvanize citizens, making people extra sensitive (or, paradoxically, perhaps less sensitive) to political communications.[3] So, we analyze effects across different issue areas, but we do so with the recognition that issue saliency could cut in different ways empirically, generating strong effects because many or most people consider the topic important (e.g., Krosnick 1990) or weakened effects because people have already been exposed and have no further to move (e.g., Druckman and Leeper 2012).

Heterogeneous Media Effects

Given the scope of our analyses—an attempt to study the effects of hundreds of press-worthy events across more than a decade in dozens of countries—we are purposely vague regarding our expectations. We suspect that certain factors, such as methodological choices will be important when it comes to identifying significant media effects. However, there always exists the possibility that combining the various factors will obscure our ability to identify effects that are real. Likewise, powerful events at one time point may dissipate in another.

Thus, if anything, we expect heterogeneity. The notion of heterogeneity speaks to broader concerns about forms of validity. Increasingly, scholars have focused on internal validity (i.e., "causality"), and design choices weigh heavily upon it. However, "…internal validity is not the sine qua non of all research" (Shadish, Cook, and Campbell 2002, 98), especially since other subtle factors related to the statistical assumptions could mask real effects (i.e., "statistical conclusion validity" in the language of Shadish et al.). It could also be that the constructs are not

[3] The paradox concerns the twin possibilities of pre-treatment effects (e.g., Druckman and Leeper 2012) in which the communication effects are already taking place before the analysis starts or alternative paths to influence that exist outside the mass media, such as when individuals communicate with each other (e.g., McClurg 2006; Ryan 2010). Again, design choices may help contend with these possibilities.

properly operationalized or measured (i.e., "construct validity"). At a very general level, though, we are perhaps most interested in the issue of generalizability, or "external validity" in the parlance of Shadish, Cook, and Campbell 2001; Campbell and Stanley 1963). Often external validity is narrowly interpreted in terms of the units being studied (e.g., descriptive characteristics of survey respondents). However, the search for generalizable effects is broader than that. It goes beyond the units to include types of treatments, contexts in which those treatments were delivered, and outcome measures. As such, we adopt a macro view of media effects.

Data and Methods

We seek to estimate media effects across locations, time, outcomes, and designs. To do so, we face uncommon and formidable data acquisition challenges. First, we need data that span geographic borders. This rules out commonly used and high quality datasets like the American National Election Studies (ANES) or National Annenberg Election Study (NAES). Likewise, we would like to be able to study media events over time. Again, temporally isolated multi-country studies, such as those conducted by Gallup or Pew, are excluded. Finally, we need surveys that are broad with respect to outcomes and media exposure measures; often surveys possess one or the other, but not both.

One of the few data collections meeting all these requirements is the European Social Survey (ESS). This is a cross-national public opinion survey conducted bi-annually since the 2002. In the first five rounds, which we study, an average of 26 countries appeared in each survey round; many of the same countries are surveyed repeatedly (e.g., the United Kingdom, Belgium), but occasionally other countries enter and leave the sample (e.g., the Russian Federation, Lithuania, Norway). Across the first five rounds, there were nearly 2,000 respondents per country and the average response rate was 62 percent across the rounds.[4] Most surveys are in the field for a few weeks, although some fieldwork periods are longer. Importantly for our study, interviewing takes place throughout the year with some temporal overlap.[5]

The ESS surveys are of particular interest because the survey collection teams record events that take place in each of the countries that could conceivably affect response patterns. In the first five rounds of the ESS, researchers affiliated with the data collection efforts in each of the countries documented more than 8,000 events of interest (n=8,142).[6] As an example, in the second wave of the ESS there was an event reported for October 14, 2004 concerning a parliamentary struggle between the Prime Minister of Portugal and the new elected leader of the most important opposition party. The event dataset for the ESS elaborates on the feud and suggests that this event might be expected to produce less satisfaction toward the way the government is acting, which is an outcome variable in the ESS data.[7] While this event was likely relevant for Spanish respondents, not all of the events were as isolated; many concerned developments in other countries or event events in countries outside of the ESS sample (e.g., the U.S. presidential election or developments in China). To focus our attention on the events most likely to produce an effect, we had two coders unfamiliar with the project characterize all of the

---

[4] The average number of respondents was 1,923 with roughly the same number in each round (round 1 average=1,925, round 2=1,887, round 3=1,891, round 4=1,968, and round 5=1,943). Likewise, in most rounds the ESS approached the target response rate of 70% with averages in the low 60s for each round (61, 62, 63, 62, and 60 for each of the rounds respectively).

[5] See http://www.europeansocialsurvey.org/ for more details on the surveys and methodology.

[6] Most of the events are in round 5 (n=2,153) while the least are in round 1 (n=717). Most events are single day events (80%). Not all countries have events recorded, but of those that do, 11% occur within 30 days of the survey start and 77% take place within the interview period.

[7] Specifically, the ESS event data characterizes this event as follows: "Santana Lopes had his first parliamentary debate on October 14, since nominated in July 2004. This was simultaneously the first parliamentary confrontation between the Prime Minister and the newly elected leader of PS, Socialist Party and the most important opposition party. The Prime Minister has avoided the polemic of the Marcelo crisis and the main subjects of debate were the economy, the rents and the SCUT (highway pays). *José Sócrates has doubted the legitimacy of Santana Lopes to the place of Prime Minister and accused him of not winning national elections (he was substituting José Barroso, the previous Prime Minister, which went to European Commission, without any election).* Discussion about the State Budget was nearly absent" (emphasis added).

events in terms of whether they were domestic or international as well as whether they were major or minor.[8] Roughly 20 percent of the events were categorized as major (1,663 of 8,142, or 20.4%) and most were domestic (5,812 or 71.4%). Our inquiry considers nearly 900 events that were both major and domestic (n=880). The events covered many different topics, but the main ones were economic issues, scandals/resignations, crime, disasters, elections, and strikes.[9] Roughly 100 of these occurred in the thirty days prior to the survey events in each country while the remainder took place while the surveys were in the field.[10]

We consider the effects of these ESS events on three variables: trust in politicians, economic satisfaction, and satisfaction with the government.[11] The other key individual-level variables in our inquiry were the media exposure measures. To determine whether a respondent was exposed to the media, we created a trichotomous measure of media exposure made up of an index of television, radio, newspaper usage.[12] As individual-level controls, we employed the standard battery of demographic considerations (e.g., education, income, age, race, and gender).[13]

[8] In a randomly selected sample of fifty events, the two research assistants achieved relatively high intercoder reliability statistics for domestic (Krippendorf alpha=.92) and major vs. minor distinctions (Krippendorf alpha=.60).
[9] We created dummy variables for each of these relative to the omitted baseline of non-economic national events.
[10] The appendix contains details on the countries and events by ESS round as well as other coding decisions.
[11] The trust in politicians question was, "Using this card, please tell me on a score of 0-10 how much you personally trust each of the institutions I read out. 0 means you do not trust an institution at all, and 10 means you have complete trust. Firstly...trust in politicians." The economic satisfaction question was an 11 point scale (from 0=extremely dissatisfied, 1=extremely satisfied) of "On the whole how satisfied are you with the present state of the economy in [country]?" Finally, the government satisfaction item used the same 11 point scale in response to "Now thinking about the [country] government, how satisfied are you with the way it is doing its job? These variables have the ESS neumonics of TRSTPLT, STFECO, and STFGOV.
[12] The media index was an additive scale built from the responses to 8-point measures of "on an average weekday, how much of your time watching television is spent watching news or programmes about politics and current affairs?" for television and similar items for radio, and newspapers. The answer choices were time-based increments ranging from "no time at all" to "more than three hours."
[13] The education item was a seven point measure from less than lower/secondary to higher tertiary education above an MA degree. Race was a binary indicator of whether the respondent belonged to "a minority ethnic group" in the country. Income was a twelve point measure of household net total income from less than €1800 to €120000 or more. All independent and dependent variables were rescaled to the 0 to 1 interval.

Beyond the individual-level, we create measures for media system freedom based upon Freedom House scores (see Schoonvelde 2013 for similar measures).[14] Variables representing a country's political system (1=parliamentary system; 0=otherwise), compulsory voting, and gross national income also were included. The country-specific factors one might include are potential limitless; we limited our attention to factors that have established effects in previous work on this topic. As a precaution, we report analyses with fixed-effects terms for countries in a series of robustness checks.

In the empirical analyses that follow, we estimate as many models as we can for the three dependent variables subject to data constraints dictated by three designs. For the baseline design, we study events occurring 30 days or less from the start of the survey period and we focus on the media exposure coefficient. That estimate is then compared with the two rival designs discussed earlier: (1) a within-survey/within-subjects (WS/WS) design and (2) a difference-in-differences (DID) design. For the WS/WS comparison, the design imposed an extra restriction of having a similar dependent variable which was not influenced by media exposure but one that could plausibly tap baseline levels of trust. For this we employed trust in the UN (i.e., each trust in politicians variable was differenced by levels of trust in the UN at the individual-level). For the DID analysis, we needed observations before and after the key media event. That meant studying

[14] The media freedom measure is a continuous measure that rates countries based on government interference in their media sectors. In its original form, it is scaled from 0 (most free) to 100 (least free) and is constructed from 23 items that are subdivided into three equally weighted subcategories: legal environment, political environment and economic environment. See Schoonvelde (2013) for a detailed description of the subcategories, but broadly they cover laws and the legal regulatory environment (legal), political control over media content (political), and ownership structures (economic). The variable was inverted and rescaled to the 0 to 1 interval so that higher values convey more freedom.

a different set of events than the first two designs (exposure by country and WS/WS within a country). [15]

Unlike other media effects studies, our quantities of interest are the regression output from hundreds of statistical models. Specifically we examine the absolute value of the t-values for models in each of the designs with the goal of uncovering which designs produce the "most significant" results. Of course, another quantity of interest is the subset of cases that exceed the 1.96 significance threshold for $p < .05$ (two-tailed) findings. So, in auxiliary analyses we also consider that specification. However, both of the preceding analyses have to do with *statistical* significance. To examine *substantive* significance, we attempt to look at the size of the coefficients (i.e., effect sizes) in yet another auxiliary analysis. These analyses proceed as two-step multilevel models (e.g., Jusko and Shively 2005), in which the data is the output from hundreds of models estimating media effects. To adjust for the repeated observations by event (i.e., some events are present in all three designs, producing three entries for each model), we cluster the standard errors and apply White's correction to offset any potential heteroskedasticity (Lewis and Linzer 2005).

Empirical Results

Table 1 provides some basic descriptive information on the events we studied. In particular, the table contains all designs aggregated as well as each of the separate designs. We have 839 observations for the trust in politicians outcome, and the mean t-value (in absolute terms) was 1.829 with a standard deviation of 2.120. The range was essentially zero (.006) to more than 24 (24.247), which is an extremely large t-value. The other outcome variables have

[15] We make use of 741 unique events for which models could be estimated due to data requirements (i.e., occurring at right moment relative to the survey interview period). Some of these events are repeated in the dataset when analyzed by different designs.

slightly higher means (1.990 for economic satisfaction and 2.262 for government satisfaction), both of which are—on average—significant findings in the sense that they would be above the 1.96 threshold for findings at the 95% confidence level using both tails of the distribution. The average number of observations was around 10,000 for all three outcomes with a range of fewer than twenty to more than 20,000.

The aggregate patterns for all of the designs together mask a considerable amount of variance. Specifically, for the nearly 100 events we studied using the media exposure design (n=98), the average t-values were much smaller for all three outcomes (i.e., never larger than 1.133 on average and never more than a value of three). The average number of observations was also more modest at 568 with a range of 17 to 1,889. The WS/WS design had the same sample size average and range for the one outcome we could study (due to the lack of a counterfactual outcome on the satisfaction measures). Likewise, the average t-value was under a value of 1 and never rose beyond 2.3. The last part of Table 1 foreshadows patterns that will be seen in the regression analyses discussed next. For the 643 events we could study using the DID approach, the average t-value was comfortably above $p < .05$ levels since they were above two for all three outcomes and the sample sizes were near 12,000 on average. Thus, the descriptive statistics tell an important story about variation across the designs with a decisive edge going to the DID design.

INSERT TABLE 1 HERE

Perhaps it is not surprising, then, when we examine the regression output in Table 2 that the coefficients for the designs are statistically significant and signed in directions that mirror what we saw in the descriptives. For the first dependent variable of trust in politicians, the entries in the first column show that the WS/WS design has smaller absolute t-values than the media exposure design (omitted category) baseline by roughly a quarter point (-.245 with a standard error of .074, $p < .01$ two-tailed). In contrast, the t-values in the DID design were two and a half points larger than the media exposure designs net of the other factors we considered (coeff.=2.529, $p < .01$). This is a pattern that was accentuated for the other two outcome variables. T-values in the DID design were bigger, by 4.218 for economic satisfaction and 5.439 for government satisfaction ($p < .01$ for both). Thus, the DID design is much more likely to detect significant media effects than the typical media exposure design.

INSERT TABLE 2 HERE

While it is the case that the DID designs offered more observations per event studied (and presumably more statistical power), the next methodological factor we considered shows that having numerous cases does not necessarily mean more significant results. In particular, the log of the number of observations available is significantly ($p < .01$) associated with smaller t-values for all three outcome variables.[16] This means that the DID design has an advantage that is not simply due to the edge in statistical power; if anything, having more cases tended to produce fewer statistically significant results with this design. This finding is counterintuitive and at odds with the conventional wisdom concerning the need for statistical power in media effects studies (e.g., Zaller 2002).[17]

The last methodological factor we consider relates to the temporal dimension of our study. In particular, the t-values were smaller in more recent ESS rounds for two of the three

[16] We used the log of the number of observations instead of the count to produce more meaningful results, but we obtain the same finding with the unlogged counts for all three outcome variables; the coefficients are negatively signed and significant at $p < .01$, two-tailed.

[17] Interactions with the design dummy variables and the number of observations (logged) reveal negative and significant coefficients for the DID design interacted with the number of observations ($p < .01$ for trust in politicians and $p < .10$ for the satisfaction outcomes). For trust in politicians model where we are able to contrast the WS/WS technique, that interaction term between the WS/WS design and the number of observations is also negative and significant ($p <.05$); the term is positive and significant ($p < .05$), suggesting that additional observations in the omitted exposure design baseline are associated with higher t-values.

outcome variables (trust in politicians and economic satisfaction), but the significance levels were weaker ($p < .10$). In other analyses (not shown), we employed dummy variables for each ESS round instead of the additive term that is shown in Table 2. In those regressions, the latest rounds are much less likely to produce large t-values as compared with the initial ESS surveys in the early-2000s for the satisfaction outcomes.[18] We hesitate to speculate on the cause of this null effect, but it is a potentially unsettling development for media effects researchers and one that provides suggestive evidence—though far from conclusive—in support for arguments concerning a new era of minimal media effects.

Our next class of variables shown in Table 2 relates to country-level factors. In particular, we studied institutional variables as well as relative wealth. The only factor that seems to matter consistently is media freedom. For all three outcome measures, the coefficient on media freedom is negative and statistically significant ($p < .01$). In contrast, the other country-level factors are almost never statistically significant; only the parliamentary dummy in the trust model approaches conventional significance levels ($p < .10$). From this it seems that informal institutions, such as the level of freedom in a country's media system structure, tend to make it harder to find statistically significant effects for all three outcomes we studied.[19] As for why media freedom tends to undercut the statistical significance of media effects, we suspect it may be related to complimentary trends that tend to co-occur in open societies such as free information exchange beyond the mass media. In other words, if information flows freely,

[18] For economic satisfaction, a round 5 ESS dummy variable has a coefficient of -.976 with a standard error of .47, $p < .05$ (the baseline is round 1). For government satisfaction, the coefficient is = -1.105 with a standard error of .658, $p < .10$. The dummies for rounds 2-4 are also negatively signed, but most are $p > .05$.

[19] Once again, there is some evidence that the effect is specific to the DID design based upon interactions with the design and media freedom (all three interaction term coefficients are negative, but the *p*-values range from .08 to .16.

country-wide events may influence everyone more readily, irrespective of whether they report high media exposure or not.

While the media freedom finding is provocative, we hesitate to read too much into these preliminary analyses without additional research. In particular, our models may have omitted other important country-level factors. One way of diagnosing potential omitted variable problems is to include fixed effect dummy variables for each country that can, in essence, stand in for country level factors that have been omitted from the models. When we do this, the coefficients remain negatively signed, but the standard errors rise considerably to the point where the media freedom findings become statistically insignificant for all three outcomes. In contrast, the findings concerning methodological factors (i.e., design dummies and observation counts) remain even when we include the country fixed effects. All of this suggests that researchers studying variations in media systems might want to be even more cautious when conducting cross-national comparisons. Attempts to find countries that are otherwise similar may be worthwhile, and researchers have been exploring ways to identify states or regions that can serve as counterfactuals based upon matching (e.g., Abadie, Diamond, and Hainmueller 2010).

The final set of factors we consider in Table 2 are related to issues. Of the dummy variables that capture differences in the substantive content of the events, two prove to be statistically significant in many of the models. Events related to crime tend to have smaller t-values (although less so on government satisfaction) and the same holds for stories about disasters (mainly for political trust and government satisfaction, both $p < .01$). Here the omitted baseline comparison group is non-economic national events and stories. Our interpretation of the issue findings relies on the same logic we introduced earlier. For sensational crime/scandal stories or those relating to national disasters, information is disseminated widely. Searching for an exposed subset of the population, when most people are exposed, becomes harder.

The findings thus far make use of the absolute t-values. To guard against any biases that may be related to conceptualizing the dependent variable this way, we also estimated our models with two other versions of the outcome measures. The first alternative dependent variable dichotomizes the t-values measure so that values of 1.96 or greater are scored as 1 and all others take zero. This type of measure considers when we have "significant" media effects using the $p < .05$ threshold for significance at the 95% level for two-tails. Roughly one-third of our models turned up significant effects for each of the three dependent variables.[20] These models (see the appendix for tables of output) largely confirm that patterns reported earlier. Design choices, the number of observations, media freedom, and issues all matter in the same ways when it comes to discovering significant effects or not. For example, the WS/WS design makes it 14 percent less likely to observe significant findings for trust in politicians than the media exposure baseline (marginal effect=.137, se=.049, $p < .05$) while the DID design elevates the likelihood by 28 percent (marginal effect=..279, $p < .01$). The DID design detects significant effects even more powerfully for the other outcome variables (35 and 41 percent improvements, both $p < .01$, for economic and government satisfaction respectively).

But statistical significance (i.e., t-value) or finding statistically significant results (dichotomizing $t > 1.96$) does not necessarily mean the results would be *substantively* significant. To consider the relative magnitude of the effects, we changed the dependent variable to the media effects coefficient. In those models (reported in the appendix), the DID design generates

[20] For trust in politicians, 33.7% of the models produced media effects of 1.96 or greater (mean=.337, sd=.473). For economic satisfaction, the mean was similar (mean=.331, sd=.471) and for government satisfaction, there were a few more significant effects (mean=.364, sd=.481).

larger coefficients holding other factors constant. Other factors, like the log of the number of observations and media system factors are negatively related to the (absolute) size of the coefficient (always $p < .01$ for all three outcome measures).[21] The issue related factors are intermittently significant, with disasters producing slightly smaller coefficients on average. Thus, considering statistical as well as substantive significance, the same patterns appear. Design choices powerfully shape both the statistical and substantive effects across hundreds of events and dozens of countries in surveys spanning a decade.

Discussion

Our analyses were unconventional. Instead of focusing on isolated events, issues, or methodologies, we cast a wide net. The consistency of findings—across outcome measures and measurement choices—was reassuring. But consistency does not necessarily mean consistently significant. In fact, nearly two-thirds of the models we estimated produced statistically *in*significant coefficients. In the course of research, analysts sometimes cycle through many different specifications in a search for publishable findings. These specification searches (Leamer 1978) appear to be related to professional pressures related to editorial standards, leading to publication bias in favor of significant results (see Gerber, Green, and Nickerson 2000 or Gerber and Malhotra 2008 on the "file-drawer problem" in political science). This study avoids that problem by reporting everything; we reveal the contents of the entire filing cabinet, or at least several drawers of it when it comes to media effects research.

While the breadth of our study may have virtues, it comes with downsides as well. One drawback relates to notion that some of our findings, even if statistically significant, may be due

[21] In addition to the same set of variables we considered earlier, we include the standard error of the coefficient as a precaution on the idea that a big coefficient might not be meaningful except in relation to the size of the standard error.

to chance variation. That is, even at the *p* <.05 significance level, we would expect one in twenty coefficients to be significant. Given that we estimated nearly 2,300 coefficients for all three dependent variables for just the original dependent variable specifications (N's of 839, 741, and 741 for each model in Table 2, which sum to 2,321), this means that there are probably more than 100 spurious findings in the set of significant findings (Type I errors). Of course, there are probably an offsetting number of insignificant findings which are truly significant (i.e., Type II errors). One way to correct for this problem, Bonferroni-type adjustments, could lead to more conservative conclusions.

A second weakness of the present study pertains to factors that could not be included due to data availability. For example, it is natural to wonder what the results would look like with alternative measure of media exposure; self-reported exposure measures were used in some of the designs and they have been critiqued by some scholars (e.g., Prior 2009). Another line of inquiry might include a measure representing the proximity of the event to the survey. On average, our events were roughly eleven days prior to the start of the survey (average = 10.7). This information is available for two of our designs (media exposure and WS/WS), but the way we calculated the DIDs meant that there was always a thirty day window before and after the event so the timing could not be considered explicitly. Yet another limitation pertains to variation in the expected relationship between the event and the dependent variable (e.g., some events might be expected to produce a negative finding while others are expected to be positive related to the outcome measure). All three versions of the models reported earlier ignore the direction of the finding, but one could imagine altering our variables so as to capture this information.[22]

Finally, there have been efforts to expand the ESS, and the data source itself was designated as an exceptionally valuable European research asset in 2013.[23] There is little doubt about the utility of the surveys for a great number of outcomes, but a different set of priorities might emerge from the perspective of someone searching for media effects based upon the events data. In particular, there was a great deal of heterogeneity in how the studies are conducted in each country. Earlier we reported the high degree of variation in the survey response rates across the countries, but there was also variation in when the surveys took place, how they were conducted, and the question that were asked, among other things. More research is needed on the ESS events file itself to make those data more useful (also see the appendix). We only used a faction of the entire events file (i.e., ten percent of the events that were both major and domestic), and there might be other subtle patterns in terms of the country-based reporting which could alter the effects. For instance, some countries contributed greatly to the events report—the top countries with more than five percent of the events were Spain, Portugal, Belgium, Germany, Hungary, and Israel. Other countries registered far fewer events. Bulgaria, Italy, Iceland, Cyprus, Luxembourg, Russia, Ireland, and Slovakia contributed fewer than two percent of the cases. Thus, while geographically broad, there might be patterns in terms of the

[22] We were able to include a term on the right-hand side which captured whether the coefficient was negative or positive. Those “negative coefficient” terms are themselves negative and significant ($p < .05$), and their inclusion did not change the patterns reported earlier.

[23] In November of 2013, the ESS was awarded ERIC (European Research Infrastructure Consortium) status. According to the news release (http://www.europeansocialsurvey.org/about/news.html), “ERICs are facilities for the scientific community, allowing researchers access to archives and tools to conduct top-level research. Member States, Associated and Third Countries and intergovernmental organisations may become members of an ERIC.”

distribution and quality of the events file and survey data which could be influencing our results.[24]

Conclusion

Our macroscopic study of media effects suggests that design choices weigh heavily on the findings. Against the backdrop of the traditional media exposure model, some research designs accentuate (DID) or diminish effects (WS/WS) across a range of outcomes and settings. A subtle factor related to design choices—the statistical power of the model—seems to have counterintuitive effects. While the number of cases is important in traditional media exposure model design as Zaller (2002) showed and our results confirm, different designs that elevate the importance of counterfactuals demonstrate that the number of cases is less important and may even result in fewer statistically significant findings. Thus, as Shadish, Cook, and Campbell remind us, research design choices often trump statistical considerations (2002, 105).

Another finding which cuts against the conventional wisdom concerns the role of institutional factors. Formal institutions were almost never significant (i.e., parliamentary system or compulsory voting) whereas informal institutions related to media freedom did matter—but the direction of the influence was negative. That is, significant media effects were less likely to be observed in countries with “freer” media systems. We urge readers to view this result with caution since it was not robust to alternative specifications; there appear to be other country-level factors that make the negative media freedom effect diminish. However, even showing no effects for media institutions should be of interest given the state of the literature (Fraile 2013; Hallin

[24] Other questions concerning the events arise too, such as the relationship of events to actual coverage. Others who study events (e.g., Smetko et al. 2003; Ladd and Lenz 2009; Stevens et al. 2011) show that they do generate coverage.

and Macini 2004; Iyengar et al. 2010; Schoonvelde 2013). The isolated instances of system significance may be outliers in the larger population of possible media effect studies.

Nevertheless, media effects can be identified on a large-scale across many different outcomes and methodological choices. Whether or not our own findings on the generalizability of media effects are generalizable is unknown. But, with more studies of studies, we will be able to make assertions that span designs, time, space, outcomes, and contexts. Seeing the entire forest rather than individual trees reveals quite a bit even if some details are lost in the process.

Appendix

This appendix provides a description of the events data as well as details on data processing and coding that was necessary to undertake prior to our analysis. Replication data and code will be available on the authors' website after publication.

Description of ESS Events Data

The European Social Survey (ESS) is a cross-national study that has been conducted every two years since 2001 in various countries across Europe. In conjunction with the individual-level data sets for each round, the ESS team has also released data designed to capture the political context within the participating countries. The political structure of Europe is such that there are likely to be shared environmental factors affecting sets of countries, as well as domestic factors specific to individual nations. The ESS event file offers an expansive, publicly-available data source for researchers looking to integrate these factors into their analyses.

Each event report typically provides several pieces of information, including a substantive description (e.g. “UK house prices have fallen for an 11th consecutive month”) and categorization ("[e]vents concerning the national economy, labour market”) of the event, start and end dates, and potentially connected items from the survey instrument. Responsibility for collecting these data appears to be decentralized, falling to separate research teams in each country involved in the broader study. Each group follows a set of common instructions on how to collect and record media-reported events. This delegation of collection duty to the numerous local teams has advantages with respect to accurately capturing events occurring in many locales at once. On the other hand, one drawback likely attributable to this arrangement is heterogeneity in what gets reported by each team. For instance, some events are sourced, while others are not. There are also practical differences in formatting and structure between subsets of the countries.

For those who may wish to construct new variables or employ the events data in a modeling capacity, standardization is an obvious imperative. We transformed the data set in several ways to improve its usefulness in our analyses. Many of the issues we outline below are likely to be encountered by all users upon first opening the unprocessed events file. Our corrections are often generally applicable. The corrected events data set and the underlying code are available in the replication materials for this paper.

Appendix Table 1 summarizes the cumulative events data file for all countries participating in any of the first five ESS studies. This table shows which survey rounds each country participated in, as well as counts of events in the data set. We first show the total number of events reported by a country, and then subdivide this number into events reported in the thirty days prior to the start of one of a given country's survey rounds, and number of events reported during one of a given country's survey rounds. The table also displays separate counts for one subcategory of events we deemed particularly useful ("Domestic/Major"; we discuss this distinction later). Ignoring for a moment these final three columns, several features of the data are worth noting. First, the pattern of inclusion in the five rounds varies considerably across the set of countries. Less than half of the participating nations were present for all rounds (i.e.

Denmark, UK). Others are included for only a single year (Austria), while the rest participate in some continuous (Ukraine) or non-continuous (Netherlands) subset of rounds.

INSERT APPENDIX TABLE 1 HERE

Similarly, there is a large degree of variance in the overall number of stories reported in each country. This may be due in some part to substantive differences in political context between the participant nations, but there are also systematic differences in reporting frequencies that seem difficult to justify on substantive grounds. For example, Spain and the United Kingdom are both large countries that participated in all five waves. However, the former reported nearly three times as many events (1,441) as the latter (484). Such extreme discrepancy likely reflects differences between the reporting patterns of the separate ESS teams rather than real variance in the political environment within the associated countries. Caution is advised in using these data for any application that might require comparable between-country counts of events.

Figures A1a and A1b graphically illustrate a few of the ways in which event reporting differed between countries, again using Spain and the United Kingdom as examples. The two countries first vary in terms of the time frame and length of survey interview periods, as depicted by the horizontal lines within the chart space. Likewise, there are also substantial differences in the timing of event reports, denoted by the rug plot (i.e., the black vertical lines) positioned above the X-axis. Spain reported more events than the United Kingdom overall (see Appendix Table 1), and reporting closely coincides with the timing of the five ESS rounds. On the other hand, the United Kingdom team reported many events in the intervening period between rounds.

The data set includes media-reported events occurring both internationally and domestically. An election in the United States, for instance, might be reported if it receives significant coverage. While international events could be utilized in other settings, the most useful reports for our analyses were those reflecting unique qualities of the political environment within a single country. To identify this category of events, coders read through every entry in the cumulative file, and judged whether each operated at the international (i.e. an election in the USA reported by the UK team) or domestic level (an election in the UK reported by the UK team).

Finally, events within the file vary considerably in terms of their magnitude of importance. Perceptions of importance are, of course, subjective to a degree, but some events clearly stood out to us as more likely to have perceptible effects on ESS survey responses than others. Our coders also made entries denoting which events appeared to be “major” compared to the others reported. To illustrate, we judged an attempted car bombing at Glasgow airport to be major, while a report about an isolated factory closing was judged to be minor. The cross-section of events that were both domestic and major was of greatest interest. As shown in the rightmost columns of Appendix Table 1, these events comprise a small subset of the overall reporting.

In addition to our new coding, we also corrected numerous existing issues within the data:

Creating Consistent Date Formats

Maintaining a uniform date format for each record is necessary to effectively use the events file with statistical software. Unfortunately, the date entries in the unprocessed file fluctuate between four different primary formats: mm/dd/yyyy for single dates, and either dd-dd/mm/yyyy, dd/mm/yyyy-dd/mm/yyyy, or dd/mm-dd/mm/yyyy in cases where an event spanned multiple days. There are also dozens of entries with idiosyncratic formatting that does not match any of these patterns. Dates with non-standard formatting were coerced to one of the four main styles. In order to form a uniform date indicator, we first ran a program to automatically identify the format of each date entry. This information is used to parse information on the day, month, and year of each event. We use this information to construct separate variables for the start and end date of each event. For single-day events, both of these variables take the same value. We store these dates in a single, common format (mm/dd/yyyy) easily read by modern software.

Identification of the Survey Field Period

Another problem related to dates involved the published beginning and end of the ESS survey periods for each country. Documentation on the ESS website provides a set of “fieldwork period” dates corresponding to each country for each round. However, these dates often fail to match the earliest and/or latest interview dates recorded in the individual-level survey data. Having an accurate sense of when events occurred relative to the beginning of each survey was important for many of our analyses. Thus, we constructed our own survey start and end variables from the dates of the actual interviews in the survey data.

Removal of Duplicate Events

We deleted a total of 207 duplicate entries in the events file. Many of these entries were a result of multiple reporting of a single, ongoing event. This type of duplicate appeared in an inconsistent pattern throughout the data, so we settled on the convention of keeping only a single event report in these cases. Some others duplicates had no immediately obvious reason for being repeated, and were also removed.

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. “Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California’s Tobacco Control Program.” *Journal of the American Statistical Association* 105 (June): 493-505.

Barabas, Jason, and Jennifer Jerit. 2009. “Estimating the Causal Effects of Media Coverage on Policy-Specific Knowledge.” *American Journal of Political Science* 53 (Jan.): 73-89.

Barabas, Jason, and Jennifer Jerit. 2010. “Are Survey Experiments Externally Valid?” *American Political Science Review* 104 (May): 226-42.

Barker, David C. and Adam B. Lawrence. 2006. "Media Favoritism and Presidential Nominations: Reviving the Direct Effects Model." *Political Communication* 23(1):41-60.

Bartels, Larry M. 1993. “Messages Received: The Political Impact of Media Exposure.” *American Political Science Review* 87 (June): 267-85.

Bennett, W. Lance, and Shanto Iyengar. 2008. “A New Era of Minimal Effects? The Changing Foundations of Political Communication.” *Journal of Communication* 58: 707-31.

Berinsky, Adam J., and Donald R. Kinder. 2006. “Making Sense of Issues Through Media Frames: Understanding the Kosovo Crisis.” *Journal of Politics* 68 (August): 640-56.

Campbell, Donald T. and Julian C. Stanley. 1963. Experimental and Quasi-Experimental Designs for Research. Boston: Houghton-Mifflin.

Curran, James, Shanto Iyengar, Anker Brink Lund, and Inka Salovaara-Moring. 2009. “Media System, Public Knowledge, and Democracy: A Comparative Study.” *European Journal of Communication* 24: 5-26.

Dalton, Russell J., Paul A. Beck, and Robert Huckfeldt. 1998. “Partisan Cues and the Media: Information Flows in the 1992 Presidential Election.” *American Political Science Review* 3(1): 111-26.

Dilliplane, Susanna, Seth K. Goldman, and Diana C. Mutz. 2013. “Televised Exposure to Politics: New Measures for a Fragmented Media Environment.” *American Journal of Political Science* 57 (Jan.): 236-48.

Dunaway, Johanna. 2008. “Markets, Ownership, and the Quality of Campaign News Coverage.” *Journal of Politics 70*: 1193–1202.

Druckman, James N. 2005. “Media Matter: How Newspapers and Television News Cover Campaigns and Influence Voters." *Political Communication* 22(4): 463-81.

Druckman, James N., and Thomas J. Leeper. 2012. “Learning More from Political Communication Experiments: Pretreatment and Its Effects.” *American Journal of Political Science* 56 (Oct.): 875-96.

Druckman, James N., and Michael Parkin. 2005. "The Impact of Media Bias: How Editorial Slant Affects Voters." *Journal of Politics* 67 (Nov.): 1030-49.

Fair, Christine, Neil Malhotra, and Jacob N. Shapiro. 2012. "Faith or Doctrine? Religion and Support for Political Violence in Pakistan." *Public Opinion Quarterly* 76 (Winter): 688-720.

Finseraas, Henning, Niklas Jakobsson, and Andreas Kotsadam. 2011. "Did the Murder of Theo van Gogh Change Europeans' Immigration Preferences?" *Kyklos* 64 (3): 396-409.

Finseraas, Henning, and Ola Listhaug. 2013. "It Can Happen Here: The Impact of The Mumbai Terror Attacks on Public Opinion in Western Europe." *Public Choice* 156: 213-228.

Fraile, Marta. 2013. "Do Information-Rich Contexts Reduce Knowledge Inequalities? The Contextual Determinants of Political Knowledge in Europe." *Acta Politica*. 48: 119-43.

Gerber, Alan, and Neil Malhotra. 2008. "Do Statistical Reporting Standards Affect What Is Published? Publication Bias in Two Leading Political Science Journals." *Quarterly Journal of Political Science* 3: 313-26.

Gerber, Alan S., Donald P. Green, and David Nickerson. 2000. ''Testing for Publication Bias in Political Science.'' *Political Analysis* 9:385-92.

Gomez, Brad T., and J. Matthew Wilson. 2008. "Political Sophistication and Attributions of Blame in the Wake of Hurricane Katrina." *Publius* 38 (Fall): 633-650.

Hallin, Daniel C., and Paolo Mancini. 2004. Comparing Media Systems: Three Models of Media and Politics. New York: Cambridge University Press.

Hetherington, Marc. 1996. "The Media's Effect on Voters' National Retrospective Economic Evaluations in 1992." *American Journal of Political Science* 40: 372-95.

Hutchings, Vincent L. 2001. "Political Context, Issue Salience, and Selective Attentiveness: Constituent Knowledge of the Clarence Thomas Confirmation Vote." *Journal of Politics* 63 (Aug.): 846-68.

Iyengar, Shanto, Mark D. Peters and Donald R. Kinder. 1982 "Experimental Demonstrations of the 'Not-So-Minimal' Consequences of Television News Programs." *American Political Science Review* 76: 848-58.

Iyengar, Shanto, Kyu S. Hahn, Heinz Bonfadelli, and Mirko Marr. 2009. "'Dark Areas of Ignorance' Revisited: Comparing International Affairs Knowledge in Switzerland and the United States." *Communication Research* 36: 341-58.

Iyengar, Shanto, James Curran, Anker Brink Lund, Inka Salovaara-Moring, Kyu S. Hahn, Sharon Coen. 2010. 2010. "Cross-National versus Individual-Level Differences in Political Information: A Media Systems Perspective." *Journal of Elections, Public Opinion & Parties* 20 (3): 291-309.

Jerit, Jennifer, Jason Barabas, and Toby Bolsen. 2006. “Citizens, Knowledge, and the Information Environment.” *American Journal of Political Science* 50 (April): 266-82.

Jusko, Karen Long, and W. Phillips Shively. “Applying a Two-Step Strategy to the Analysis of Cross-National Public Opinion Data.” *Political Analysis* 13 (Autumn): 327-44.

Kahn, Kim Fridkin, and Patrick J. Kenney. 2002. “The Slant of the News: How Editorial Endorsements Influence Campaign Coverage and Citizens' Views of Candidates.” *American Political Science Review* 96 (2): 381-94.

Keele, Luke, Neil Malhotra, and Colin H. McCubbins. 2013. “Do Term Limits Restrain State Fiscal Policy? Approaches for Causal Influence in Assessing the Effects of Legislative Institutions.” *Legislative Studies Quarterly* 38 (Aug.): 291-326.

Kellstedt, Paul M. 2000. “Media Framing and the Dynamics of Racial Policy Preferences.” *American Journal of Political Science* 44(2):245-260.

Keeter, Scott, Courtney Kennedy, April Clark, Trevor Tompson, and Mike Mokrzycki. 2007. What's Missing from National Landline RDD Surveys?: The Impact of the Growing Cell-Only Population." *Public Opinion Quarterly* 71: 772-92.

Keeter, Scott, Courtney Kennedy, Michael Dimock, Jonathan Best, and Peyton Craighill. 2006. "Gauging the Impact of Growing Nonresponse on Estimates from a National RDD Telephone Survey." *Public Opinion Quarterly* 70: 759-79.

Klapper, Joseph T. 1960. *The Effects of Mass Media*. New York: Free Press.

Krosnick, Jon A. 1990. “Government Policy and Citizen Passion: A Study of Issue Publics in Contemporary America.” Political Behavior 12: 59-92.

Ladd, Jonathan McDonald, and Gabriel S. Lenz. “Exploiting a Rare Communication Shift to Document the Persuasive Power of the News Media.” *American Journal of Political Science* 53 (April): 394-410.

Lassen, David Dreyer. 2005. “The Effect of Information on Voter Turnout: Evidence from a Natural Experiment.” *American Journal of Political Science* 49 (Jan.): 103-18.

Lazarsfeld, Paul and Marjorie Fiske. 1938. “The ‘Panel’ as a New Tool for Measuring Opinion.” *Public Opinion Quarterly* 2 (4): 596–612.

Lazarsfeld, Paul F., Bernard Berelson and Hazel Gaudet. 1944. *The People’s Choice: How the Voter Makes Up His Mind in a Presidential Campaign*. 2nd ed. New York: Duell Sloan and Pearce.

Leamer, Edward E. 1978. *Specification Searches: Ad Hoc Inference with Nonexperimental Data.* New York: Wiley.

Lewis, Jeffrey B., and Drew A. Linzer. 2005. "Estimating Models in Which the Dependent Variable is Based on Estimates." *Political Analysis* 13: 345-64.

Maestas, Cherie, Lonna Atkeson, Thomas Croom and Lisa Bryant. 2008 "Shifting the Blame: Federalism, Media and Public Assignment of Blame Following Hurricane Katrina." *Publius: The Journal of Federalism*. 38(4): 609-632.

McClurg, Scott D. 2006. "The Electoral Relevance of Political Talk: Examining the Effect of Disagreement and Expertise in Social Networks on Political Participation." *American Journal of Political Science* 50 (3): 737-54.

McGuire, William J. 1986. "The Myth of Massive Media Impact: Savagings and Salvaging." In G. Comstock (Ed.), *Public Communication and Behavior*. Orlando: Academic Press, pp. 173-257.

Miller, Beth. 2010. "The Effects of Scandalous Information on Recall of Policy-Related Information." *Political Psychology* 31: 887.

Morgan, Stephen L., and Christopher Winship. 2007. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. New York: Cambridge University Press.

National Research Council. 2013. *Nonresponse in Social Science Surveys: A Research Agenda.* Panel on a Research Agenda for the Future of Social Science Data Collection Committee on National Statistics. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Neuman, W. Russell, Marion R. Just, Ann N. Crigler. 1992. *Common Knowledge: News and the Construction of Political Meaning*. Chicago: University of Chicago Press.

Patterson, Thomas, and Robert D. McClure. 1976. *The Unseeing Eye: The Myth of Television Power in National Elections.* New York: Putnam.

Prior, Markus. 2007. *Post-Broadcast Democracy: How Media Choice Increases Inequality in Political Involvement and Polarizes Elections.* Cambridge University Press.

Prior, Markus. 2009. "The Immensely Inflated News Audience: Assessing Bias in Self-Reported News Exposure." *Public Opinion Quarterly* 73 (Spring): 130-43.

Rosen, Jay. 2001. *What Are Journalists For?* New Haven: Yale University Press.

Ryan, John Barry. 2010. "The Effects of Network Expertise and Biases on Vote Choice." *Political Communication* 27(1): 44-58.

Schoonvelde, Martijn. 2013. "Media Freedom and the Institutional Underpinnings of Political Knowledge." *Political Science Research and Methods* 1 (October): 1-16.

Semetko, Holli A., Wouter Van Der Brug, and Patti M. Valkenburg. 2003. “The Influence of Political Events on Attitudes toward the European Union.” *British Journal of Political Science* 33 (Oct.): 621-34.

Shadish, William R, Thomas D. Cook, and Donald T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inferenc*e. Boston: Houghton Mifflin.

Soroka, Stuart. 2006. “Good News and Bad News: Asymmetric Responses to Economic Information.” *Journal of Politics* 68: 372-85.

Soroka, Stuart, Blake Andrew, Toril Aalberg, Shanto Iyengar, James Curran, Sharon Coen, Kaori Hayashi, Paul Jones, Gianpetro Mazzoleni, June Woong Rhee, David Rowe., and Rod Tiffen. 2013. “Auntie Knows Best? Public Broadcasters and Current Affairs Knowledge.” *British Journal of Political Science* 43: 719-39.

Statham, Paul, and Howard Tumber. 2013. “Relating News Analysis and Public Opinion: Applying a Communications Method as a ‘Tool’ to Aid Interpretation of Survey Results.” *Journalism* 14 (6): 737-53.

Stevens, Daniel. 2008. “Measuring Exposure to Political Advertising in Surveys.” *Political Behavior* 30 (1): 47-72.

Stevens, Daniel, and Susan Banducci. 2013. “One Voter and Two Choices: The Impact of Electoral Context on the 2011 UK Referendum.” *Electoral Studies* 32: 274-84.

Stevens, Daniel, Susan Banducci, Jeffrey Karp, and Jack Vowles. 2011. “Prime Time for Blair? Media Priming, Iraq, and Leadership Evaluations in Britain.” *Electoral Studies* 30 (April): 546-60.

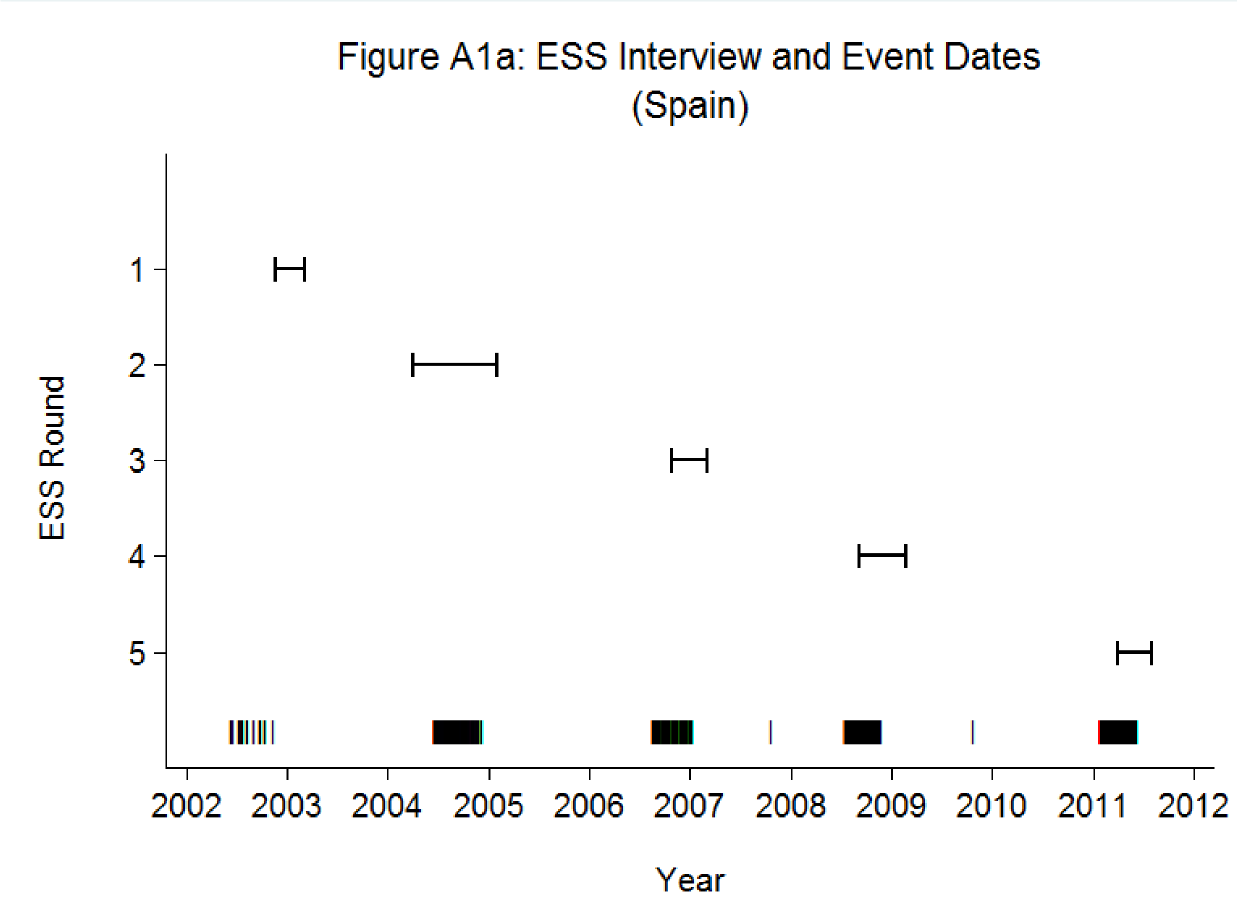
Stevens, Daniel, and Jeffrey A. Karp. 2012. “Leadership Traits and Media Influence.” *Political Studies* 60: 787-808.

Wooldridge, Jeffrey. 2013. *Introductory Econometrics: A Modern Approach,* 5th ed. Mason, OH: South-Western Cengage Learning.

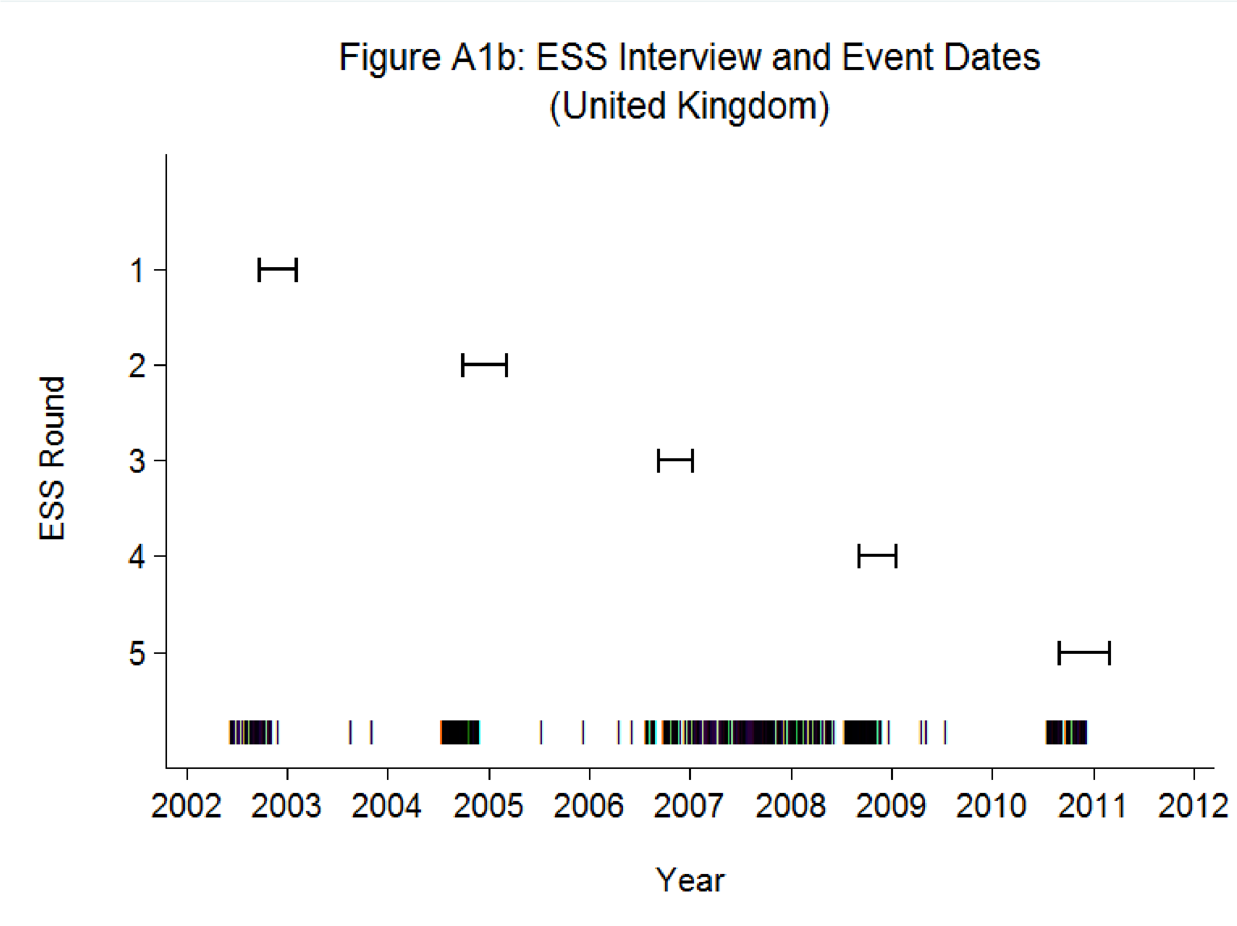
Zaller, John R. 2002. “The Statistical Power of Election Studies to Detect Media Exposure Effects in Political Campaigns.” *Electoral Studies* 21: 297-329.

Zaller, John R. 1996. "The Myth of Massive Media Impact Revived: New Support for a Discredited Idea.” 1996. In Diana Mutz, Richard Brody, and Paul Sniderman (eds.), *Political Persuasion and Attitude Change*. Ann Arbor: University of Michigan Press, pp. 17-79.

Zaller, John R. and Mark Hunt. 1995. “The Rise and Fall of Candidate Perot: The Outsider Versus the Political System - Part II.” *Political Communication* 12: 97-123.



Note: Horizontal bars denote the survey period of each ESS round. Vertical marks along the X-axis denote reported events.



Note: Horizontal bars denote the survey period of each ESS round. Vertical marks along the X-axis denote reported events.

Table 1. Descriptive Information

	Number of Obs.	Mean	Standard Dev.	Minimum	Maximum
Overall (All Designs)				--	--
Trust Politicians |t-value|	839	1.829	2.120	.006	24.287
Econ. Satisifaction |t-value|	741	1.990	3.089	.005	38.943
Gov't Satisfaction |t-value|	741	2.262	3.394	.006	40.455
Cases for Trust Politicians	839	9278	6681	17	20739
Cases for Econ. Satisfaction	741	10399	6223	17	20662
Cases for Gov't Satisfaction	741	10191	6103	17	20230
Media Exposure Design					
Trust Politicians |t-value|	98	1.133	0.737	.046	2.821
Econ. Satisifaction |t-value|	98	0.830	0.654	.065	2.494
Gov't Satisfaction |t-value|	98	0.827	0.649	.018	2.618
Cases for Trust Politicians	98	568	453	17	1889
Cases for Econ. Satisfaction	98	568	453	17	1889
Cases for Gov't Satisfaction	98	568	453	17	1889
Within-Survey/Within-Subjects Design					
Trust Politicians |t-value|	98	0.888	0.638	.006	2.296
Econ. Satisifaction |t-value|	N/A	N/A	N/A	N/A	N/A
Gov't Satisfaction |t-value|	N/A	N/A	N/A	N/A	N/A
Cases for Trust Politicians	98	568	453	17	1889
Cases for Econ. Satisfaction	N/A	N/A	N/A	N/A	N/A
Cases for Gov't Satisfaction	N/A	N/A	N/A	N/A	N/A
Difference-in-Differences Design					
Trust Politicians |t-value|	643	2.078	2.335	.007	24.287
Econ. Satisifaction |t-value|	643	2.167	3.270	.005	38.943
Gov't Satisfaction |t-value|	643	2.481	3.585	.006	40.455
Cases for Trust Politicians	643	11932	5289	235	20739
Cases for Econ. Satisfaction	643	11897	5254	232	20662
Cases for Gov't Satisfaction	643	11657	5159	229	20230

Table 2. Generalized Media Effects: Predicting Model |T-Values|

	DV: Trust Pol.	DV: Econ. Satis.	DV: Gov't Satis.
Methdological Factors [a]			
Design: Within-Survey/Subjects (WS/WS)	-.245 *** (.074)	--	--
Design: Difference-in-Differences (DID)	2.529 *** (.522)	4.218 *** (1.124)	5.439 *** (1.329)
Number of Observations (Logged)	-.470 *** (.141)	-.860 *** (.313)	-1.158 *** (.380)
ESS Survey Rounds 1-5	-.137 * (.073)	-.184 * (.088)	-.186 (.139)
Country Factors			
Media System Freedom	-2.234 *** (.766)	-2.095 *** (.703)	-2.218 *** (.848)
Parliamentary System	-.339 * (.184)	-.021 (.279)	-.117 (.314)
Compulsory Voting	-.010 (.232)	-.148 (.342)	-.256 (.451)
Gross National Income per Capita/1000	.010 (.010)	.009 (.009)	.028 (.017)
Issue Factors			
Economic	-.440 * (.228)	.177 (.292)	-.410 (.331)
Scandal	-.273 (.254)	-.224 (.254)	-.246 (.374)
Crime	-.660 * (.393)	-.802 *** (.339)	-.460 (.555)
Disaster	-1.109 *** (.310)	-.360 (.523)	-1.429 *** (.456)
Election	-.254 (.304)	.202 (.427)	-.063 (.455)
Strike	-.102 (.446)	.905 (.644)	.114 (.653)
Constant	6.552 *** (1.340)	8.100 *** (2.214)	9.827 *** (2.730)
R-squared	.10	.07	.09
F-test	8.20***	5.79***	5.95***
Number of cases (i.e., models estimated)	839	741	741
Number of countries	28	28	28

Note: Coefficients are ordinary least squares estimates with dependent variables of Trust in Politicians (Trust Pol.), economic satisfaction (Econ. Satis.), and government statisfaction (Gov.'t Sat). Robust standard errors, clustered by the event (in cases of repeated events), are in the parentheses.
*** $p < .01$, ** $p < .05$, * $p < .10$ (two-tailed)
[a] Effects vs. omitted baseline design of within a single country, exposed compared to unexposed.

Appendix Table 1. Summary of European Social Survey (ESS) Events

		Full Data Set			Domestic/Major		
Country	Rounds	All	Prior	During	All	Prior	During
Austria	1	18	0	0	4	0	0
Belgium	1-5	1031	93	834	103	5	89
Bulgaria	3-4	20	2	9	5	0	2
Cyprus	3-4	91	1	86	8	0	8
Czech Republic	1-2, 4-5	150	60	67	33	17	13
Denmark	1-5	210	13	196	24	2	22
Estonia	2-5	83	17	45	15	3	7
Finland	1-4	113	19	83	12	2	9
France	3-4	91	3	86	7	0	7
Germany	1-5	468	25	360	56	0	43
Greece	1-2, 4	86	6	47	19	0	13
Hungary	1-5	274	72	181	41	12	26
Iceland	2	14	6	8	2	0	2
Ireland	1-5	247	0	226	25	0	18
Israel	1, 4-5	236	8	212	42	2	38
Italy	1	16	4	0	3	1	0
Luxembourg	2	260	38	209	10	0	10
Netherlands	1, 3-4	241	15	176	39	4	26
Norway	1-5	117	28	84	13	2	11
Poland	1-5	220	24	145	34	4	18
Portugal	1-5	1089	240	742	97	18	72
Romania	4	27	0	0	10	0	0
Russia	3-4	96	23	71	10	2	8
Slovakia	2-5	138	4	133	14	0	14
Slovenia	1-4	119	18	84	25	5	17
Spain	1-5	1441	110	1294	126	11	112
Sweden	1-4	76	17	55	11	6	5
Switzerland	1-5	498	22	417	36	2	31
Turkey	2	120	0	120	8	0	8
Ukraine	2-4	68	30	37	9	4	5
United Kingdom	1-5	484	19	296	39	0	28

Note: The overall count of events (*All)* is disaggregated into events occurring within a 30-day window prior to the start of a country's ESS survey period (*Prior*) and events occurring during the survey period (*During)*. "Domestic/Major" is a subset of events scored as both domestic and major by our coders (see text for details).
Round 1 = 2002, 2 = 2004, 3 = 2006, 4 = 2008, 5 = 2010.

Appendix Table 2. Generalized Media Effects: Predicting Significant Effects (t>|1.96|)

	DV: Trust Pol.	DV: Econ. Satis.	DV: Gov't Satis.
Methdological Factors [a]			
Design: Within-Survey/Subjects (WS/WS)	-.422 ***	--	--
	(.170)		
Design: Difference-in-Differences (DID)	.915 ***	1.495 ***	1.761 ***
	(.288)	(.324)	(.361)
Number of Observations (Logged)	-.088	-.148 *	-.125 *
	(.069)	(.073)	(.072)
ESS Survey Rounds 1-5	-.105 ***	-.156 ***	-.065
	(.040)	(.045)	(.043)
Country Factors			
Media System Freedom	-1.365 ***	-1.935 ***	-.065
	(.482)	(.493)	(.450)
Parliamentary System	-.262 ***	.046	.027
	(.110)	(.117)	(.113)
Compulsory Voting	.124	.164	-.094
	(.164)	(.167)	(.164)
Gross National Income per Capita/1000	.003	-.001	-.006
	(.007)	(.007)	(.007)
Issue Factors			
Economic	-.077	.086	.011
	(.151)	(.155)	(.155)
Scandal	-.272 *	.022	-.022
	(.150)	(.162)	(.160)
Crime	-.399	-.728 ***	.130
	(.279)	(.299)	(.277)
Disaster	-.370	-.169	-.210
	(.280)	(.282)	(.306)
Election	-.313 *	-.053	.129
	(.182)	(.189)	(.185)
Strike	.005	.562	-.084
	(.324)	(.362)	(.386)
Constant	1.418 *	1.697 ***	-.396
	(.663)	(.689)	(.633)
Pseudo R-squared	.07	.08	.07
Wald	57.21***	57.24***	48.32***
Number of cases	839	741	741
Number of countries	28	28	28

Note: Coefficients are probit estimates with dependent variables of Trust in Politicians (Trust Pol.), economic satisfaction (Econ. Satis.), and government statisfaction (Gov.'t Sat). Robust standard errors, clustered by the event (in cases of repeated events), are in the parentheses.

*** $p < .01$, ** $p < .05$, * $p < .10$ (two-tailed)

[a] Effects vs. omitted baseline design of within a single country, exposed compared to unexposed.

Appendix Table 3. Generalized Media Effects: Predicting (Absolute Value) Coefficient Size

	Trust Politicians	Economic Satisfaction	Government Satisfaction
Methdological Factors			
Design: Within-Survey/Within-Subjects (WS/WS)[a]	.001	--	--
	(.002)		
Design: Difference-in-Differences (DID)[a]	.051 ***	.103 ***	.094 ***
	(.010)	(.016)	(.016)
Number of Observations (Logged)	-.009 ***	-.019 ***	-.019 ***
	(.003)	(.005)	(.004)
ESS Survey Rounds 1-5	-.001	-.003	-.001
	(.001)	(.002)	(.002)
Country Factors			
Media System Freedom	-.033 ***	-.040 ***	-.043 ***
	(.010)	(.013)	(.016)
Parliamentary System	-.002	.010 *	.010 *
	(.003)	(.005)	(.004)
Compulsory Voting	-.001	-.008	-.009 *
	(.004)	(.005)	(.005)
Gross National Income per Capita/1000	.000	.000	.000
	(.000)	(.000)	(.000)
Issue Factors			
Economic	-.004	.003	-.006
	(.004)	(.007)	(.006)
Scandal	-.006	-.004	-.011 *
	(.005)	(.007)	(.006)
Crime	-.003	.005	.012
	(.013)	(.022)	(.024)
Disaster	-.020 ***	-.003	-.019 *
	(.006)	(.011)	(.010)
Election	-.006	-.002	-.005
	(.005)	(.007)	(.007)
Strike	-.003	.023 *	.001
	(.008)	(.013)	(.011)
SE of Coefficient	.567 ***	.342	.573 *
	(.132)	(.222)	(.251)
Constant	.100 ***	.160 ***	.160 ***
	(.026)	(.038)	(.037)
R-squared	.16	.14	.16
F-test	12.27***	13.82***	9.96***
Number of cases	839	741	741
Number of countries	28	28	28

Note: Coefficients are ordinary least squares estimates. Robust standard errors, clustered by the event (in cases of repeated events), are in the parentheses. *** $p < .01$, ** $p < .05$, * $p < .10$ (two-tailed)

[a] Effects vs. omitted baseline design of within a single country, exposed compared to unexposed.