Supporting Information for: Unexpected Event during Survey Design: Promise and Pitfalls for Causal Inference

Jordi Muñoz
University of Barcelona

Albert Falcó-Gimeno
University of Barcelona

Enrique Hernández
Universitat Autònoma de Barcelona

Published at Political Analysis (2019)
DOI: 10.0.3.249/pan.2019.27

This research has received financial support from the Spanish Ministry of Science, Innovation, and Universities through the research grant CSO2017-89085-P. We are grateful to Lucas Leeman, Silja Häusermann, Macarena Ares, Guillem Rico, Salomo Hirvonen, the editor and anonymous reviewers of *Political Analysis*, and participants at the University of Zurich IPZ publication seminar, the Autonomous University of Barcelona DEC Seminar, and the 2018 EPSA conference for helpful comments and suggestions. We are especially grateful to Erik Gahner Larsen for his detailed and insightful comments, and for generously sharing with us his collection of UESD references. Enrique Hernández also thanks the University of Zurich Political Science Department for hospitality during the spring of 2018. Replication materials are available at the *Political Analysis Dataverse*: doi.org/10.7910/DVN/RDIIVL (Muñoz et al., 2019).

Contents

A Reachability

A.1 Analysis of reachability in 21 European countries: ESS

To assess the potential problems related to the differential reachability of respondents that might affect studies exploiting the UESD we first assess how the observable characteristics of respondents are related to the time during the fieldwork when they were interviewed in the 21 European countries included in ESS-7.

Table A1: Correlates of fieldwork day when survey interview was completed: 2016 ESS

	(1)	(2)
Age (years)	-0.10	-0.07
	(0.02)	(0.02)
Education (years)	0.31	0.23
	(0.07)	(0.07)
Citizenship	-0.20	-0.19
	(1.13)	(1.10)
Male	0.67	0.42
	(0.50)	(0.49)
In a paid job	0.62	-0.61
	(0.82)	(0.80)
Unemployed	-4.30	-3.70
	(1.22)	(1.19)
Student	-0.73	-1.31
	(1.08)	(1.05)
Retired	-2.86	-2.20
	(1.04)	(1.01)
Housework	-1.94	-1.97
	(0.75)	(0.73)
Attempts to survey		3.81
		(0.10)
Refusals to participate		16.99
		(0.92)
Constant	128.43	121.70
	(2.12)	(2.07)
Country fixed-effects	Yes	Yes
Observations	39697	39697
Countries	21	21
R^2	0.347	0.383

Standard errors in parentheses

Model 1 in Table A1 summarizes the results of an OLS regression model fitted to analyze

the relationship between the sociodemographic characteristics of respondents and the timing of their survey interview. The results indicate that those who are older, less educated, and out of the labor market (unemployed, retired or doing housework) are, on average, more likely to be interviewed earlier during the fieldwork period. These sociodemographic groups should be easier to reach for an interview because presumably they should spend longer periods of time at home and might have more time available to answer surveys. Therefore, they are more likely to be interviewed earlier during the fieldwork.

The second model incorporates variables that are directly related to the differential reachability of respondents: the number of times respondents refused to participate in the survey, and the number of attempts to survey before an interview could be completed. The results indicate that refusing to participate in the survey and a higher number of attempts to survey increase the likelihood of being interviewed later during the fieldwork. Moreover, even when accounting for these variables related to the reachability of respondents the sociodemographic characteristics of those interviewed later during the fieldwork are still significantly different.

A.2 Reachability in the rolling cross-section design

In this section we replicate the analysis of the covariates of interview day using two well-known surveys that explicitly take on the rolling cross-section design: the 2015 Canadian Election Study (CES) and the National Annenberg Election (NAES) Survey 2008 Phone Edition. The rolling cross-section design is explicitly aimed at making the interview date of each respondent random, and it is typically used in the context of election campaigns with the intention of studying campaign effects. The sample is randomly split into groups that are released to the fieldwork operators on a daily basis.

This strategy should prevent imbalances in both observables and unobservables from occurring. However, in practice we might still observe some imbalances: within each block of respondents, those that are more likely to refuse and harder to reach will be concentrated towards the end of the block since they have a few days to be contacted. This means that for any arbitrary partition of the sample on might still find differences. In table A2 we summarize the results of the same analysis conducted with the ESS for the two rolling cross sections. The CES offers information on attempts to survey and refusals, while the NAES does not. In both cases we see some minor imbalances in observable characteristics of the respondents, but it

is especially relevant to notice that in the CES there is a strong correlation between date of interview and refusals and attempts to survey. Therefore, we conclude that in any case it is advisable to check and, eventually, address potential problems of ignorability of the treatment assignment in the UESD context, even in the a priori most favorable sampling conditions.

Table A2: Correlates of interview day. 2015 CES & 2008 NAES

	CES		NAES
	(1)	(2)	(3)
Male	0.13	0.07	-2.39
	(0.37)	(0.36)	(0.89)
Age	0.00	0.00	-0.01
	(0.02)	(0.01)	(0.00)
Education	0.03	0.07	1.07
	(0.09)	(0.09)	(0.19)
Born in country	0.40	0.65	1.48
	(0.53)	(0.52)	(1.61)
Paid work	0.95	0.89	0.44
	(0.58)	(0.57)	(1.41)
Retired	0.42	0.30	3.30
	(0.72)	(0.70)	(1.54)
Unemployed	-1.04	-1.31	11.91
	(1.10)	(1.09)	(3.03)
Student	4.59	4.59	-3.97
	(1.08)	(1.06)	(4.20)
Housework	0.58	0.51	-1.30
	(1.27)	(1.25)	(2.41)
Disabled/Ill	-1.24	-1.28	-3.80
	(1.49)	(1.47)	(2.63)
# Refusals to participate		6.81	
		(0.60)	
# Contacts to participate		1.47	
		(0.47)	
Constant	20.52	17.98	152.27
	(1.17)	(1.25)	(2.32)
R-squared	0.007	0.041	0.001
N	4011	4011	56181

Standard errors in parentheses

B Assessment of the empirical strategies of UESD studies

Table B1 summarizes a detailed assessment of the empirical strategies and robustness tests performed in each of the articles published using the UESD (see Table 2 for the aggregated figures). The criteria used to classify each study for each of the tests/checks are outlined in Table B2. To classify each of the studies we take into consideration both the main text and the appendixes or supplemental information of each study (when available).

Table B1 catalogs the published social science studies that rely on the UESD of which we are aware. These studies analyze the impact of different type of events on multiple outcomes, and they all do so applying the UESD as defined in Section 2. This classification is based on a systematic search of bibliographic databases (Web of Science and Google Scholar), as well as consultation with researchers who have used the UESD.

Table B1, therefore, provides a comprehensive overview of the empirical strategy and robustness checks conducted in UESD studies. The UESD definition provided in Section 2 excludes some published studies that examine the impact of events through surveys, though. This is the case, for example, of studies that rely on a survey fielded some time before the event and another survey fielded some time after the event (or multiple surveys stacked around the event), but without an overlap between the fieldwork and the event (see e.g. Bisgaard and Slothuus, 2018; Brouard et al., 2018). Similarly, we exclude studies that track the association between media coverage of multiple –and sometimes unrelated– events during a prolonged period of time, and time-trends in public opinion (see e.g. Zaller and Hunt, 1995).[1] Studies in which a survey is purposefully fielded to capture the effects of an expected event by including questions directly related to the event are also excluded (see e.g. Motta, 2018). While these studies analyze the impact of events on public opinion through surveys, their design has certain particularities and methodological challenges (e.g. time-series dynamics) not directly addressed in our paper, which focuses on the increasingly common practice of estimating the individual-level effects of unexpected events that occur during the fieldwork of a single survey.

[1] While these studies analyze the impact of campaign events, they do not usually focus on isolating the effect of a single event. Instead they include multiple (expected and unexpected) events and track public opinion reactions to those. Therefore, many of the issues discussed in our paper do no apply to these studies

Table B1: Empirical strategy and robustness checks of published studies drawing on the UESD

	In-depth description of the event	In-depth description of the survey	Balance tests	Multiple bandwidths	Covariate adjustment	Analysis of nonresponse	Placebos in the control group	Inspection of pretrends	Falsification through other surveys/units	Falsification through other outcomes	Pseudo-manipulation checks	Restricted interpretation of treatment effect
Ares and Hernández (2017)	✓		✓	✓	✓				✓	✓		
Balcells and Torrats-Espinosa (2018)	✓	✓	✓	✓	✓	✓						
Berger (2010)				✓	✓			✓			✓	
Bonilla and Grimmer (2013)	✓		✓		✓						✓	
Boomgaarden and de Vreese (2007)	✓		✓								✓	
Boydstun et al. (2018)			✓		✓							
Bozzoli and Müller (2011)	✓				✓							
Branton et al. (2015)			✓		✓							
Burlacu et al. (2018)	✓				✓							
Carey Jr et al. (2014)					✓							
Coupe (2017)					✓						✓	
Das et al. (2009)						✓						
De Vries (2018)			✓		✓							
Dinas et al. (2016)		✓	✓	✓	✓	✓		✓	✓	✓		
Dinesen and Jæger (2013)			✓	✓	✓							
Finseraas and Listhaug (2013)		✓	✓	✓	✓		✓		✓	✓		
Finseraas et al. (2011)	✓	✓	✓		✓			✓		✓		
Flores (2018)	✓	✓	✓	✓	✓					✓		
Geys and Qari (2017)			✓	✓	✓		✓			✓	✓	
Hariri et al. (2015)	✓	✓	✓	✓	✓				✓			
Hofstetter (1969)	✓		✓									
Holbrook et al. (2012)	✓				✓							
Jakobsson and Blom (2014)	✓		✓	✓	✓				✓			
Jensen and Naumann (2016)			✓	✓	✓		✓	✓		✓		
Kim and Kim (2019)	✓	✓	✓		✓	✓		✓	✓	✓		✓
Krosnick and Kinder (1990)	✓		✓		✓							
Larsen (2018)	✓		✓	✓	✓		✓	✓		✓		
Legewie (2013)	✓		✓		✓	✓						
Lever (1969)	✓				✓			✓				
Metcalfe et al. (2011)		✓	✓		✓	✓		✓	✓			
Minkus et al. (2018)	✓		✓		✓		✓	✓	✓			
Muñoz and Anduiza (2019)	✓	✓	✓	✓	✓	✓	✓	✓		✓		
Nagoshi et al. (2007)			✓									
Ojeda (2016)					✓							
Perrin and Smolek (2009)			✓		✓							
Pierce et al. (2016)			✓	✓	✓							
Reeves and de Vries (2016)	✓	✓	✓	✓	✓	✓			✓	✓		
Schaffner and Roche (2017)	✓	✓	✓		✓					✓		
Schulz et al. (2008)	✓		✓		✓							
Silber Mohamed (2013)	✓	✓	✓		✓							
Silva (2018)			✓	✓	✓						✓	
Slothuus (2016)	✓				✓			✓				
Solaz et al. (2018)	✓		✓	✓	✓					✓		
Zepeda-Millán and Wallace (2013)					✓							

Table B2: Criteria for the classification of UESD studies

Test/Check	Criteria
In-depth description of the event	Discusses when did the event happen, the circumstances leading to the event, the circumstances surrounding it, the postevent context, and provides some direct or indirect evidence of the salience or news coverage of the event.
In-depth description of the survey	Provides details about: the survey sampling procedure, the duration and development of the fieldwork, the fieldwork protocols, and the strategies to deal with unit nonresponse (e.g. timing of refusal conversion strategies).
Balance tests	Analyzes balance on pretreatment covariates between the control and treatment group(s).
Multiple bandwidths	Discusses results based on multiple bandwidths around the time of the event.
Covariate adjustment	In order to estimate the effect of the event uses some form of covariate adjustment such as controlling through regression or any other method to improve covariate balance between the treatment and control groups.
Analysis of nonresponse	Analyzes item and/or unit nonresponse before and after the event.
Placebo treatments at different dates of the control group	Conducts placebo tests at arbitrary points at the left of the cutoff point ($t_p < t_e$)
Inspection of preexisting time trends	Examines the relationship between the timing of interviews and the outcome of interest during the preevent period in order to rule out the presence of preexisting time trends.
Falsification tests through other surveys/units	Conducts a falsification test of a simulated event that occurs on the same date of the event (or at a similar time point during the fieldwork) using another edition/round of the survey used to analyze the event of interest.
Falsification tests through other outcome variables	Conducts falsification test by testing for the effects of t on an alternative outcome Y' that is theoretically unrelated to T
Pseudo-manipulation checks	Analyzes if there is a perceptible increase in the salience respondents attribute to problems related to the event of interest (through a "most important problem" type of question). We consider that a study performs this pseudo-manipulation check even if the analysis of the "most important problem question" (or a similar question) is not explicitly conducted as a robustness check to assess noncompliance.
Restricted interpretation of the treatment effect	Interprets effect of the event as an *Intent-To-Treat* (ITT)

C Charlie Hebdo attacks: Event and survey description

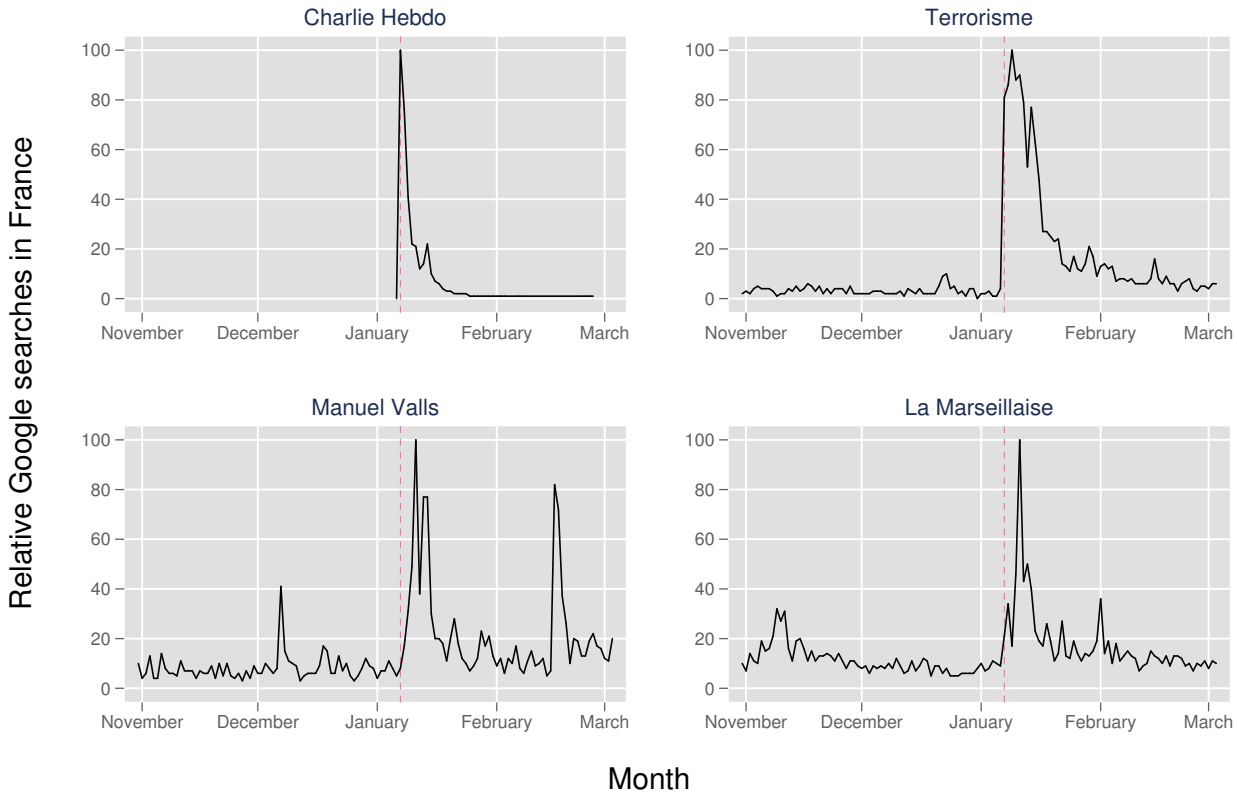
C.1 Characterization of the Charlie Hebdo attacks

The Charlie Hebdo terrorist attacks occurred in Paris on January 7th, 2015. At approximately 11:30 of January 7th, 2015, the brothers Chérif and Saïd Kouachi, identifying as members of Al-Qaeda, burst into the offices of the satirical magazine Charlie Hebdo and killed 12 people and injured 11 others. Immediately after the attack, a manhunt began. The manhunt lasted until January 9th when the two terrorists were shot and killed by the police. In parallel to the manhunt, an acquaintance of the Kouachi brothers (Amedi Coulibaly) shot a policeman on January 8th, and the day after he killed 4 hostages he had taken at a Kosher supermarket in Paris while he was in contact with the Kouachi brothers. Coulibaly was also shot and killed by the police on January 9th. At the time these were the most serious terrorist attacks in France since 1995, which also had an important symbolic impact since, until then, French citizens did not think of themselves as a target of lethal terrorism (Brouard et al., 2018; Moran, 2017).

The attacks sparked multiple reactions. On January 11th a demonstration against the attacks gathered more than 2 million people in Paris, including many presidents and prime ministers of allied countries who stood next to French leaders such as President Hollande, Prime Minister Valls, or former President Nicolas Sarkozy. In this context of intense display of national unity, political leaders such as President Hollande or Prime Minister Valls framed the shootings as an attack on core French values (Moran, 2017). In fact, the national government was "keen to rally public opinion under the 'Je suis Charlie' banner" (Moran, 2017, p. 318). These displays of national unity were accompanied by the deployment of military troops in public spaces and the announcement of numerous anti-terror measures by Prime Minister Valls (Willsher, 2015). This climate of national unity was exemplified on January 13th when, after singing the national anthem, all members of parliament voted in favor of continuing air strikes against Daesh in Iraq.

The Charlie Hebdo attacks were a highly salient political event. In the weeks that followed, the attacks and the reactions they generated were front page news in most French newspapers. In fact, in the ESS media claims dataset, which provides information about the news that are most salient during the fieldwork period as published in the newspapers *Le Figaro* and

Figure C1: Google trends for key search terms during the survey fieldwork



Note: The dashed red line denotes January 7th. All figures in the main text and supporting information were produced using the *plottig* Stata schemes (Bischof, 2017).

Libération, almost all entries (claims) are related to the Charlie Hebdo attacks until January 20th (Fernee et al., 2014). During those days, the actor with the highest number of entries in the claims dataset is the head of the government, PM Manuel Valls. The Internet search trends summarized in Figure C1 also reflect that the terms 'Charlie Hebdo' and 'Terrorisme' rapidly rose in salience in the aftermath of the attacks. Google trends also reveal a clear spike in the number of searches about 'Manuel Valls' and the French national anthem 'La Marseillaise' immediately after the shootings.

The Charlie Hebdo attacks were preceded by a context of unpopularity of the French government, which some attributed to the implementation of austerity policies (see Moran, 2017). Moreover, before the January 7th shootings, three other serious terrorist incidents occurred in December. While these events led to heightened security and the Prime Minister claimed that 'we never faced a greater risk in matter of terrorism' already in December 22nd, we would argue that the Charlie Hebdo attack was still an unexpected event, the specific timing of which was nowhere near foreseeable. The ESS media claims dataset reveals, that the terrorist attacks and

the reactions it generated dominated the news in the weeks following January 7th. It was not until around January 25th that other major political topics, like the 'Law for growth, activity and equal economic opportunity (the 'Macron Law')' gained prominence again.

C.2 Background information on ESS-7 fieldwork in France

The fieldwork of 7th round of the ESS was conducted between October 31st, 2014 and March 3rd, 2015 through face-to-face computer-assisted personal interviews (CAPI). The ESS sampling guidelines establish that samples must be representative of all persons aged 15 or older, that sampling frames of individuals, households and addresses can be used, that individuals are selected by random probability methods at each stage, and that quota sampling or substitution of nonresponding sampled units is not permitted at any point.[2]

In the case of the 7th round in France, the ESS was not granted access to the census sampling frame and the sample was selected by drawing of addresses within each Primary Sampling Unit (municipality or group of municipalities) and then randomly selecting dwellings and individuals within dwellings. Sampled units were contacted in person by interviewers, and non-contact and refusal conversion strategies began immediately after the first failed visit. The absence of sampling quotas, the use of random probability methods at each stage, and the immediate start of refusal conversion strategies are likely to minimize some of the potential biases related to the fieldwork of the survey. However, the long duration of the fieldwork, the use of a 4-stage sampling design, and the late occurrence of the event generate some challenges that must be taken into account when estimating and interpreting the effects of the Charlie Hebdo attacks.

With regard to the development of the fieldwork, it is relevant to note that the fieldwork was not interrupted by the Charlie Hebdo terrorist attacks. Interviews were conducted on January 7th after 11:30, as well as in January 8th and the following days. Even in the region most directly affected by the attacks (Ile de France) interviews were conducted during all those days.

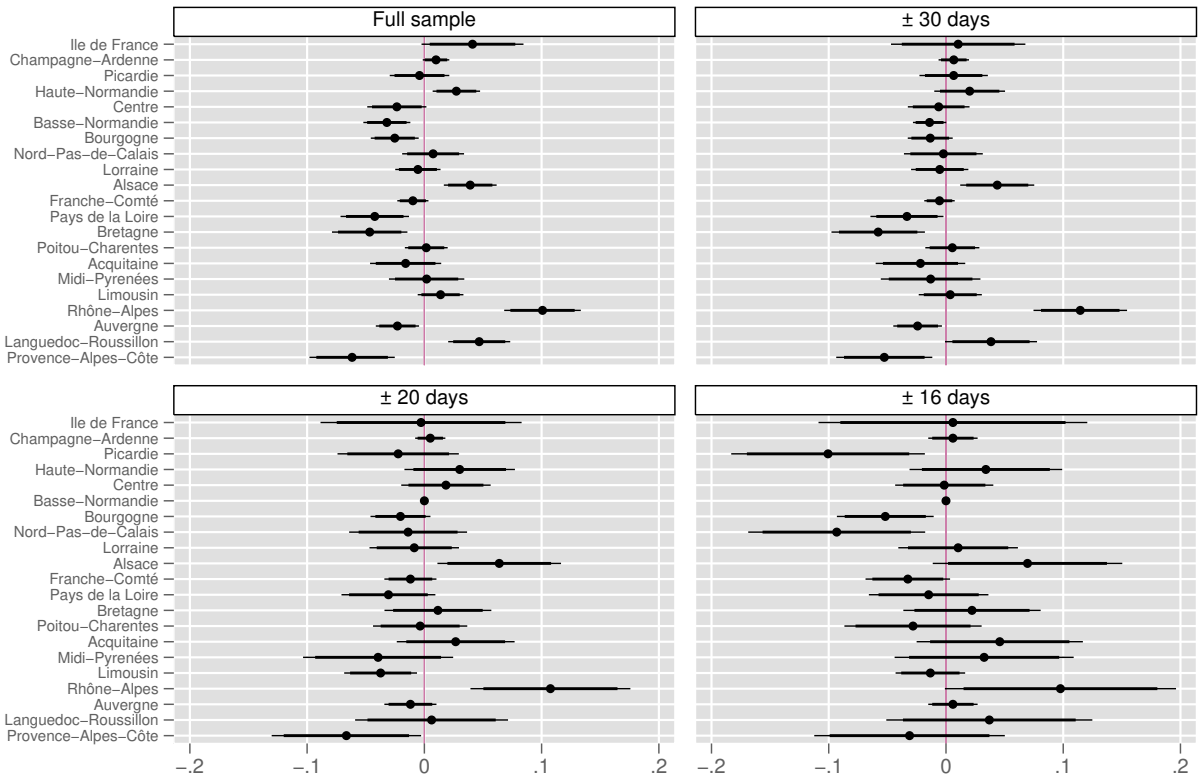
[2]See ESS sampling guidelines.

D Additional analyses of Charlie Hebdo attacks

D.1 Regional imbalances

Figure D1 summarizes the regional imbalances between the treatment and the control groups. The balance tests indicate that even when relying on a ±20 days bandwidth there are still some statistically significant imbalances in the regions of Lorrraine, Limousin, Rhne-Alpes and Provece-Alpes-Cte. Relying on an even narrower bandwidth of ±16 days would not eliminate these imbalances. In fact, for the ±16 days bandwidth some of these imbalances are even larger than for the ±20 bandwidth.

Figure D1: Balance tests: Covariate differences between the treatment and control groups.

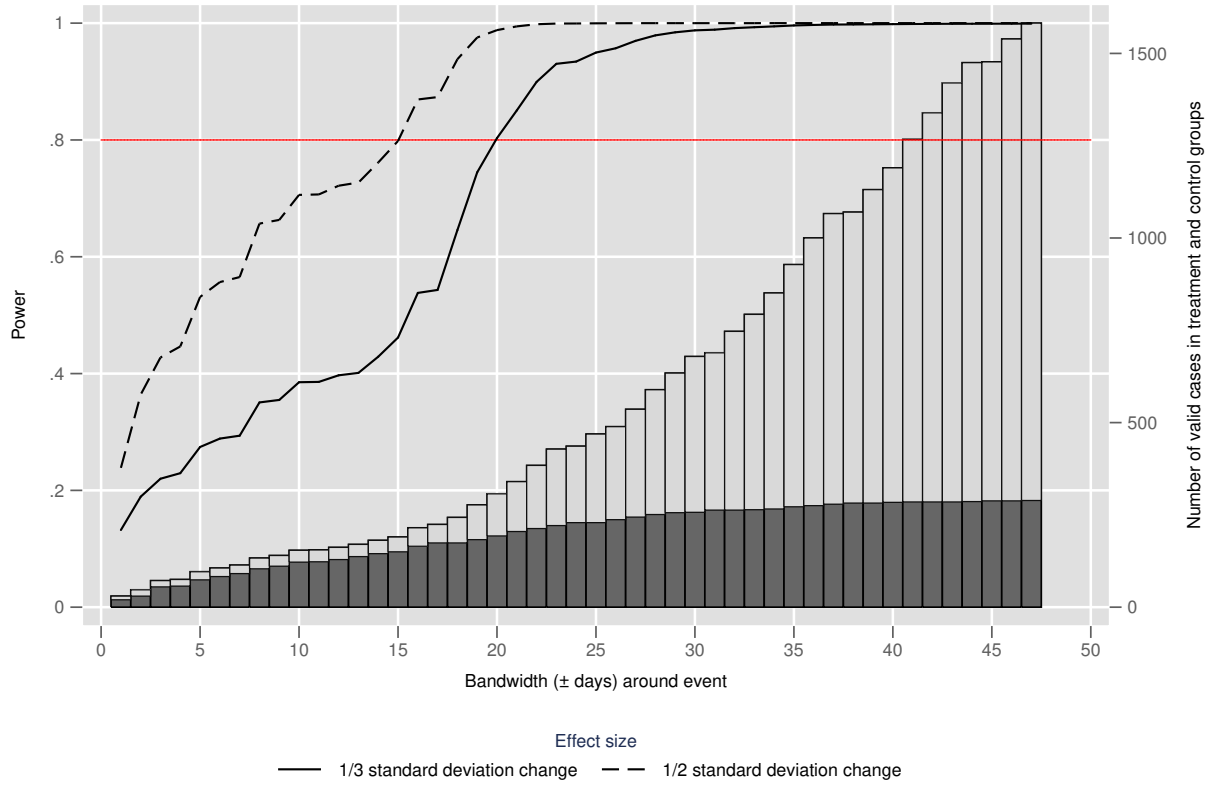


Note: Entries report the difference in the mean of the covariates between the treatment and control groups. Thick and thin lines are 90% and 95% confidence intervals, respectively

D.2 Bandwidths and statistical power

Figure D2 summarizes the power calculations for two different effect sizes. To determine the effect sizes for the power calculation we take as a reference the distribution and the standard deviation of the "Satisfaction with government" variable among those interviewed before the event (which equals 2.0 in this case). As a lower bound for a meaningful effect we consider an increase in the mean level of satisfaction with government equivalent to one third of a standard deviation (+0.67). The ±20 bandwidth is the first bandwidth with a power higher than 0.8 for this effect size. We also estimate the power for a larger effect size: an increase in the mean level of satisfaction with government equivalent to one half of a standard deviation (+1.0). The ±16 bandwidth is the first bandwidth with a power higher than 0.8 for this effect size. In addition to the power calculations, the stacked histogram of Figure D2 summarizes the number of cases included in the treatment and control groups for each bandwidth.

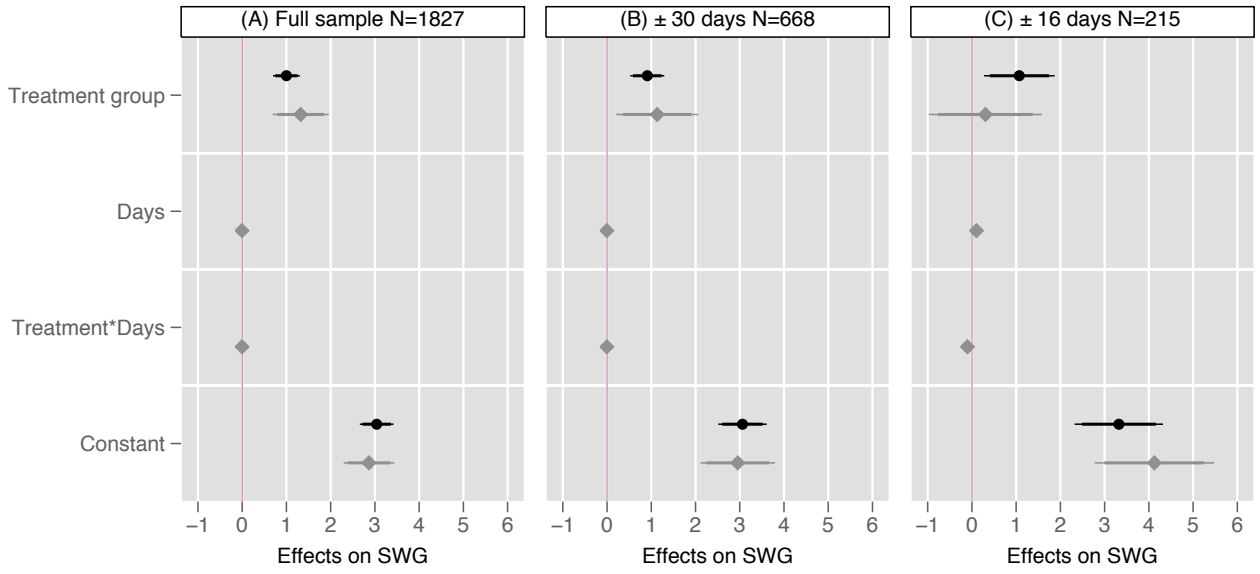
Figure D2: Statistical power for various effect sizes and bandwidths



Note: Power calculations for a 0.05 significance level and based on the standard deviation of the control group. In the histogram the total height of the stacked bars refers to the total number of valid cases included with each bandwidth, and the black and grey bars refer to the number of cases in the treatment and control groups, respectively. Valid cases are those respondents who provide a valid answer for all the covariates used in the estimation of the effects of the Charlie Hebdo attacks

Figure D3 summarizes the effect of the Charlie Hebdo attacks through the bandwidths considered in the balance tests summarized in Figure 3. The table replicates the main models used to estimate the effects of the terrorist attacks summarized in Panel C of Figure 2 (in the main text), which include region-fixed effects and entropy balancing weights. Before estimating these models we have re-estimated the entropy balancing weights for each bandwidth. To achieve convergence with all bandwidths reported in this appendix, the entropy balancing weights used in Figure D3 and Figure D5 only adjust for the first and second moment conditions of the covariates (mean and variance).

Figure D3: Effects of Charlie Hebdo attacks on satisfaction with government (SWG).

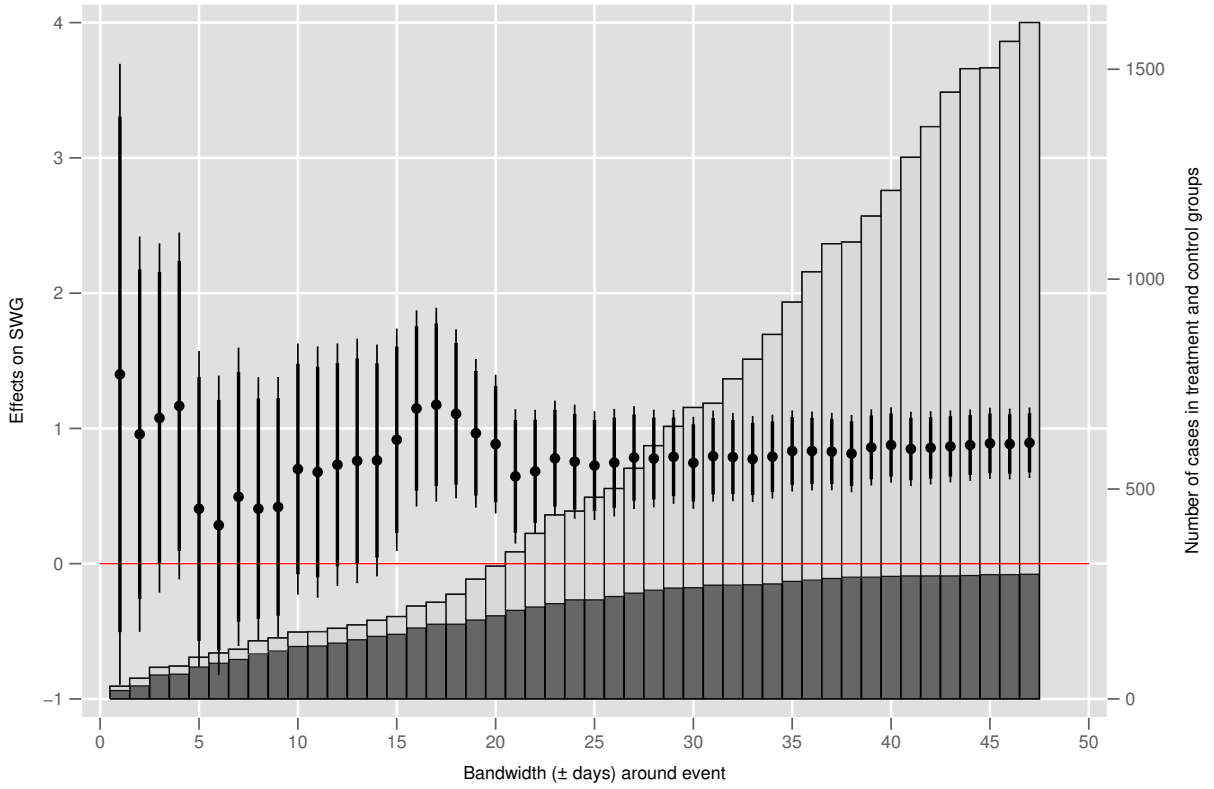


Note: Black spikes (circle markers) summarize results from OLS models that only include the treatment indicator. Grey spikes (diamond markers) summarize the results of models that include an interaction between the treatment indicator and the running variable measuring fieldwork days. Thick and thin lines are 90% and 95% confidence intervals, respectively. All models are estimated with entropy balancing weights and region fixed-effects.

Next, we summarize the results of the Charlie Hebdo attacks increasing the bandwidth around the event day by ±1 days until reaching the last day of the fieldwork. However, as the histograms in the figures show, in the case of narrow bandwidths there is a very limited number of cases, especially in the control group. For this reason, Figure D4 summarizes these results without any type of covariate adjustment. In this case, we observe that the estimated effect of the attacks is consistently positive, although for bandwidths smaller than ±15 the results are not statistically significant. Figure D5 summarizes the results for those bandwidths for which we can balance the control and treatment groups through entropy balancing (with a

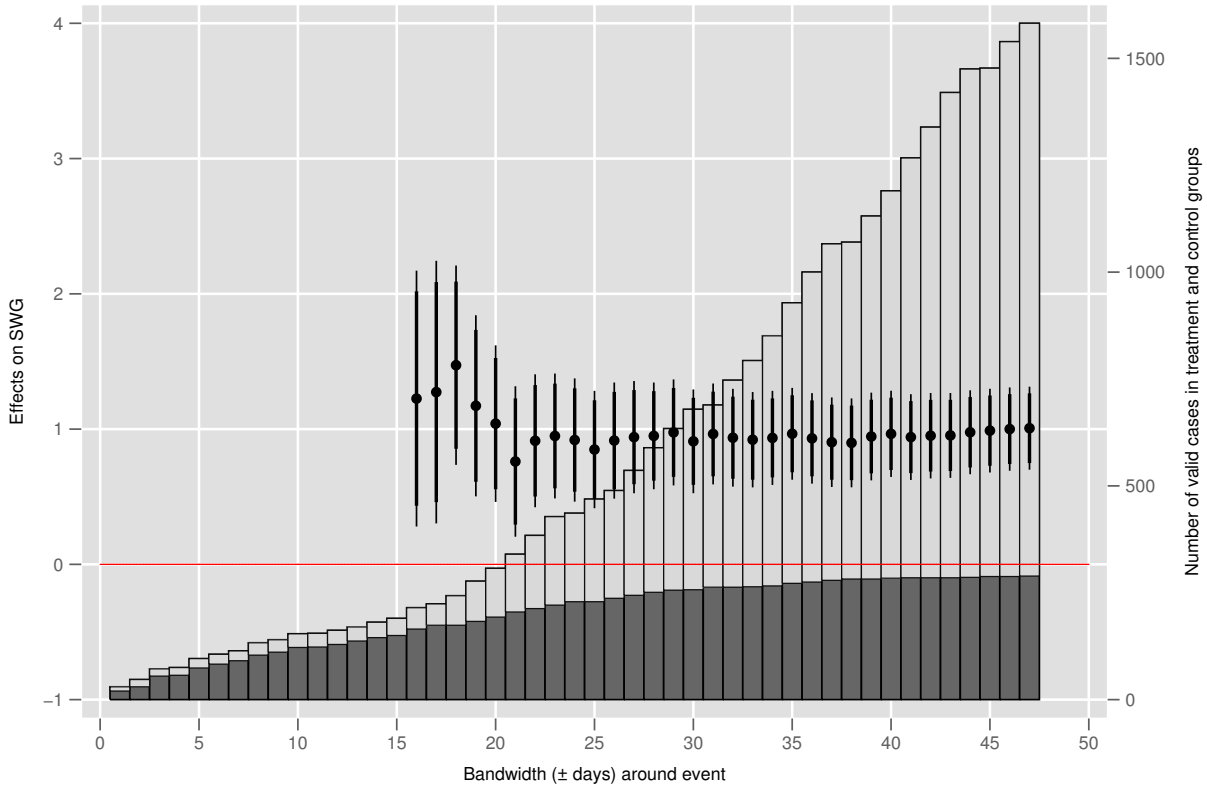
0.015 tolerance) at least when it comes to the mean and variance of the covariates discussed throughout the paper. This limits the analyses to the bandwidts equal or larger than ±16. In this case we observe that the estimated effects of the Charlie Hebdo attacks are consistently positive and statistically significant.

Figure D4: Effects of the Charlie Hebdo attacks for multiple bandwidths: no covariate adjustment



Note: Results from OLS models that only include the treatment indicator and without any type of covariate adjustment. Thick and thin lines are 90% and 95% confidence intervals, respectively. In the histogram the total height of the stacked bars refers to the total number of cases included with each bandwidth, and the black and grey bars refer to the number of cases in the treatment and control groups, respectively.

Figure D5: Effects of the Charlie Hebdo attacks for multiple bandwidths: entropy balancing and region fixed effects



Note: Results from OLS models estimated with entropy balancing weights and region fixed-effects. Entropy balancing weights adjust for two moment conditions (mean and variance) but not for skewness. Thick and thin lines are 90% and 95% confidence intervals, respectively. In the histogram the total height of the stacked bars refers to the total number of valid cases included with each bandwidth, and the black and grey bars refer to the number of cases in the treatment and control groups, respectively. Valid cases are those respondents who provide a valid answer for all the covariates used in the estimation of the effects of the Charlie Hebdo attacks

D.3 Entropy balancing details

Table D1 summarizes the distribution of the observable covariates without the entropy balancing weighting, and Table D2 summarizes this distribution in the reweighted control group (i.e. after adjusting for the first, second and third moment of all the covariates with a tolerance level of 0.015).

Table D1: Covariate distribution (no weighting)

	Treatment group			Control group		
	Mean	Variance	Skewness	Mean	Variance	Skewness
Age (years)	48.052	367.133	0.270	48.921	315.418	0.089
Education (years)	13.332	18.983	0.124	13.570	18.814	-0.067
Citizenship	0.902	0.089	-2.696	0.956	0.042	-4.455
Male	0.420	0.245	0.325	0.544	0.250	-0.176
In a paid job	0.539	0.250	-0.156	0.553	0.249	-0.212
Unemployed	0.078	0.072	3.155	0.053	0.050	4.007
Student	0.114	0.102	2.429	0.096	0.088	2.733
Retired	0.259	0.193	1.100	0.246	0.187	1.182
Housework	0.036	0.035	4.961	0.044	0.042	4.455
Voted last election	0.554	0.248	-0.219	0.649	0.230	-0.625
Voted PS last election	0.135	0.117	2.140	0.219	0.173	1.357
Refusals to participate	0.192	0.208	2.352	0.114	0.120	3.052

Initially, one would also want to balance the treatment and control groups with regards to the number of attempts to survey before the interview was completed. This would allow for a better adjustment of the potential differences in reachability between the two groups. However, like any other method of covariate adjustment, entropy balancing is not an infallible method for achieving covariate balance (Hainmueller and Xu, 2013). If the two groups are too different with regard to a certain characteristic (e.g. the number of attempts to survey) there will not be enough information to identify the counterfactual, and the control group weights generated thorough the entropy balancing might be unrealistic (Hainmueller and Xu, 2013). This seems to be the case for the variable identifying the numbers of attempts to survey each individual. Figure D6 reveals that in the treatment group we can find a substantial number of

respondents who participated in the survey after 8 or more attempts to survey them. However, in the control group only two respondents were surveyed after 8 attempts, and no respondent was surveyed after more than 8 attempts. As a consequence, if one reweights the control group with entropy balancing weights that include the variables measuring the number of attempts to survey the weights of some of the subjects in the control group are unrealistic and are likely to distort the results. Figure D7 summarizes the number of respondents for each value of the 'attempts to survey' variable and the average balance weights for each of these values when considering this variable for the entropy balancing. For example, while the average weight of the 18 respondents who were surveyed after 1, 2 or 3 attempts equals 0.07, the average weight for the two respondents that were surveyed after 8 times equals 32.42. Therefore, when using these weights, the results of the control group could be driven by the two observations that were surveyed after 8 attempts. Since this seems to be a case of what Hainmueller and Xu (2013) characterize as unrealistic weights we exclude the variable measuring the number of attempts to survey from the entropy balancing. Researchers relying on the UESD and adjusting for covariate imbalances should always assess these potential threats and decide which covariates can be adjusted.

Table D2: Covariate distribution (reweighted control group)

	Treatment group			Control group		
	Mean	Variance	Skewness	Mean	Variance	Skewness
Age (years)	48.052	367.133	0.270	48.074	367.291	0.267
Education (years)	13.332	18.983	0.124	13.338	18.993	0.119
Citizenship	0.902	0.089	-2.696	0.902	0.089	-2.704
Male	0.420	0.245	0.325	0.419	0.246	0.328
In a paid job	0.539	0.250	-0.156	0.539	0.251	-0.157
Unemployed	0.078	0.072	3.155	0.078	0.072	3.159
Student	0.114	0.102	2.429	0.113	0.101	2.438
Retired	0.259	0.193	1.100	0.258	0.193	1.106
Housework	0.036	0.035	4.961	0.036	0.035	4.967
Voted last election	0.554	0.248	-0.219	0.555	0.249	-0.221
Voted PS last election	0.135	0.117	2.140	0.134	0.117	2.147
Refusals to participate	0.192	0.208	2.352	0.192	0.208	2.348

Figure D6: Distribution of attempts to survey in the treatment and control groups

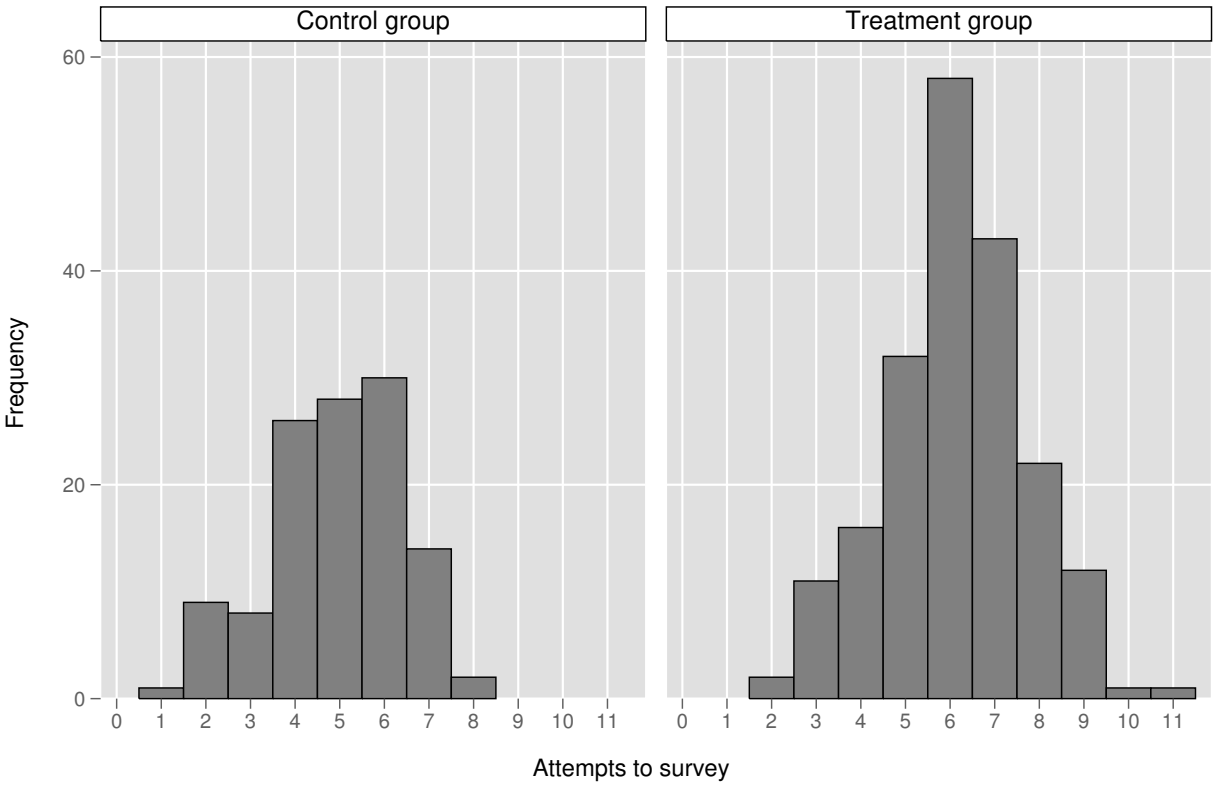
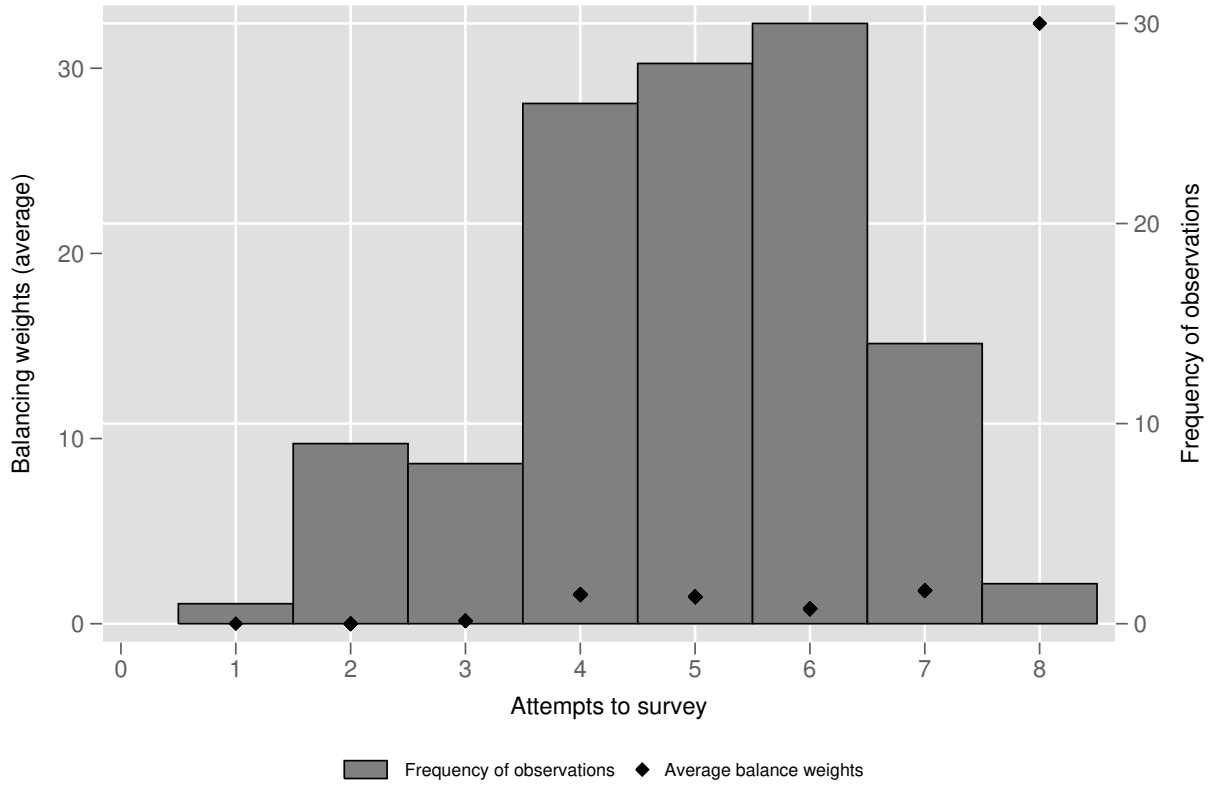


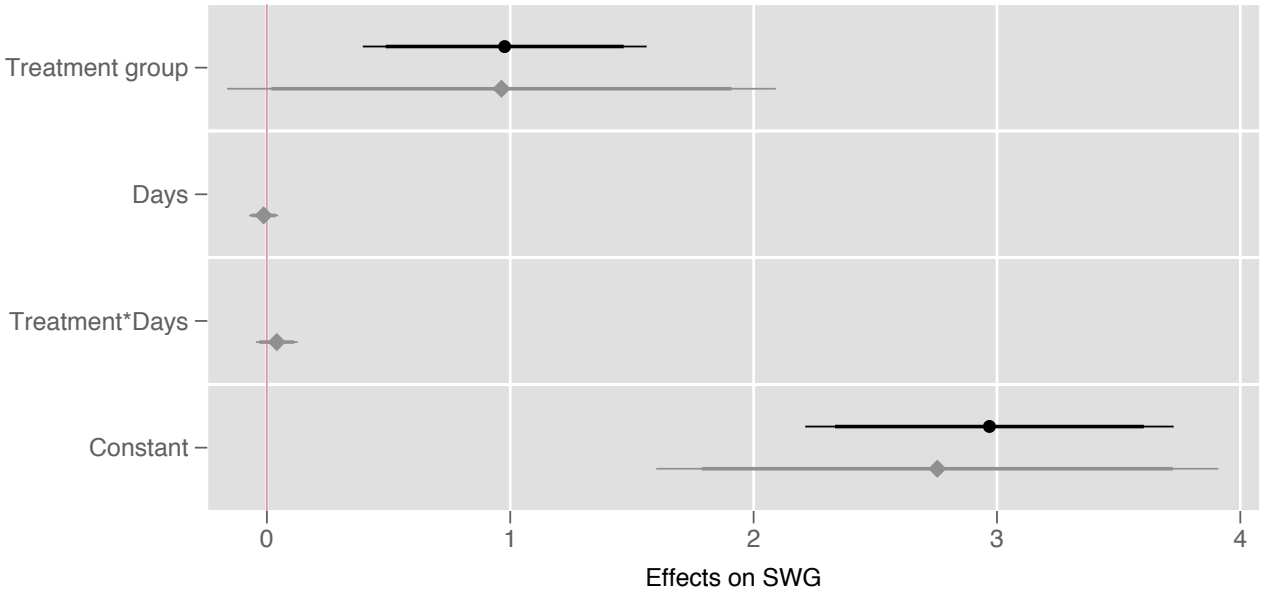
Figure D7: Average entropy balancing weights for each level of attempts to survey in the control group



D.4 Entropy balancing excluding vote recall

Whether respondents voted or not in the previous election, and the party they voted for is reported by the respondents themselves. Despite the fact that vote recall questions refer to an election held prior to the event, vote memories are often known to be biased by current preferences and social desirability (Himmelweit et al., 1978; Van Elsas et al., 2013). Since we expect the Charlie Hebdo attacks to affect evaluations of government, respondents might, therefore, modify their answers to these questions as a consequence of the terrorist attack. Respondents might be less willing to report that they voted for the party in government (or that they turned out to vote) after the attacks. We believe that this is a remote possibility since the survey questions ask about past behaviors and not about a future behavior (e.g. intention to turn out to vote) or an attitudinal orientation towards the party in government (e.g. party identification or closeness to the PS). In any case, to assess the potential posttreatment bias that this variable could induce we re-estimate the main models used to assess the impact of the terrorist attacks without balancing the control and treatment groups with respect to the vote recall variables. The results of Figure D8 are very similar to those of the models summarized in Panel C of Figure 2.

Figure D8: Effects of Charlie Hebdo attacks on satisfaction with government (SWG).



Note: Black spikes (circle markers) summarize results from OLS models that only include the treatment indicator. Grey spikes (diamond markers) summarize the results of models that include an interaction between the treatment indicator and the running variable measuring fieldwork days. Thick and thin lines are 90% and 95% confidence intervals, respectively. The models are estimated with entropy balancing weights and region fixed-effects.

D.5 preexisting time trends

To assess the plausibility of the temporal stability assumption we estimate the effect of a placebo treatment at the left of the cutoff point ($t_p < t_e$) using the empirical median of the control group as the placebo treatment point, which corresponds to December 20. Therefore, the "placebo treatment" variable takes the value 0 for those interviewed between December 17-20 and the value 1 for those interviewed between December 22 and January 6.[3] The results summarized in Table D3 reveal that the change in satisfaction with government is small and not statistically significant, both focusing on the average levels at the two sides of the placebo cutoff (Model 1) and on the jump at the cutoff (Model 2).

Table D3: Placebo treatments. Dependent variable: satisfaction with government. OLS models

	Control group		preevent period	
	(1)	(2)	(3)	(4)
Placebo treatment	-0.21	-0.15		
	(0.50)	(1.31)		
Days		-0.28	0.00	0.03
		(0.37)	(0.00)	(0.02)
Placebo treatment * Days		0.37		
		(0.38)		
Days^2				0.00
				(0.00)
Constant	3.38	2.61	3.14	3.58
	(0.52)	(1.21)	(0.22)	(0.38)
Region fixed-effects	Yes	Yes	Yes	Yes
Observations	115	115	1574	1574
R^2	0.172	0.197	0.018	0.019

Standard errors in parentheses.

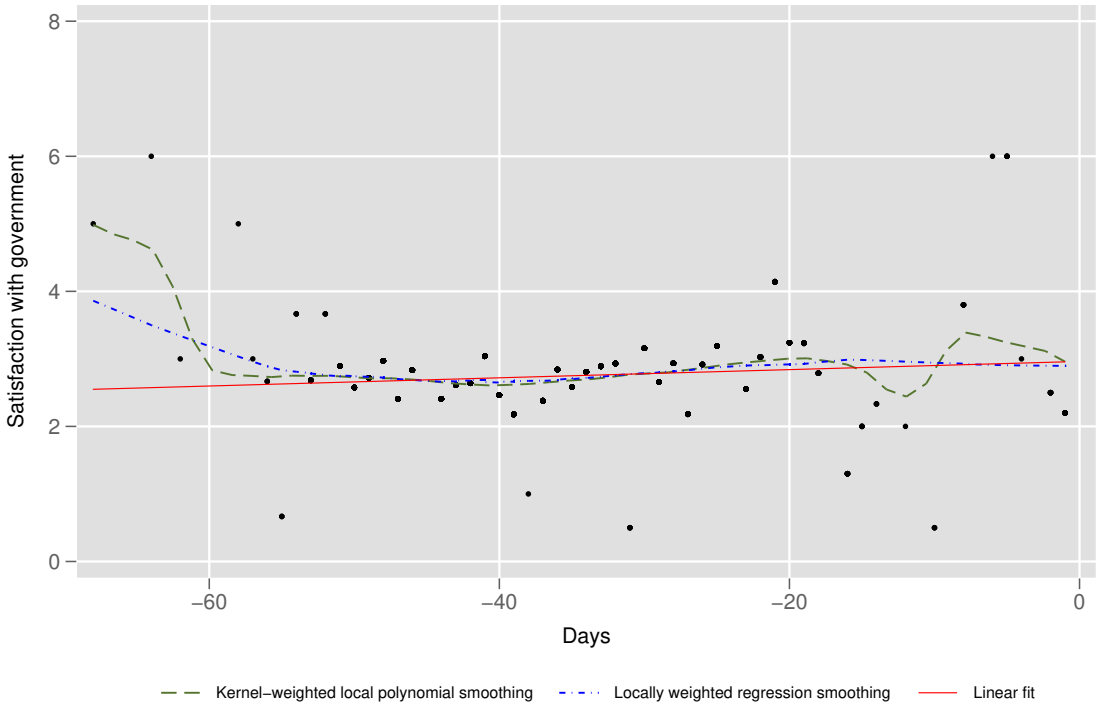
Note: In model 2 the variable days takes the value 0 on December 22.

We also analyze the relationship between the timing of interviews and the outcome variable during the whole survey fieldwork period before the event. Figure D9 illustrates the relationship between satisfaction with government and the timing of interviews in the preevent period, and

[3]There were no interviews on December 21.

Models 3 and 4 of Table D3 test for the existence of a linear and quadratic relationship between these two variables. While these variables are not systematically related, the daily average of satisfaction with government reveals some substantial jumps at points other than the event date. Therefore, in this case, it would not be advisable to draw conclusions about the effects of the event exclusively on the basis of analyses conducted at a specific point in time (e.g. the day immediately after the event).

Figure D9: Average satisfaction with government by day in preevent period (0 = January 8th).



D.6 Attrition and nonresponse

The rate of nonresponse for the "satisfaction with government" question is not affected by the Charlie Hebdo attacks. The proportion of item nonresponse (including refusals and *don't know* answers) in the control period is of 0.025 and of 0.015 in the treatment period. A two-sample t-test (not shown) indicates that this difference is not statistically significant.

Table D4 summarizes the number of sampled units who refused to answer the survey or were not surveyed for any other reason during the control and treatment periods. The proportion of refusals over the total number of attempted interviews is very similar in both periods. When analyzing all forms of unit nonresponse the results also indicate that the proportion of sampled units that do not respond is lower after the event.

Table D4: Survey nonresponse during the control and treatment periods

	Refusals	Unit nonresponse	Total attempted interviews
Control period	159 (11%)	930 (67%)	1396
Treatment period	223 (11%)	1,156 (55%)	2090

Note: The category "Unit nonresponse" also includes those who refused to participate in the survey. Percent over total attempted interviews in each period in parentheses

In order to judge if attrition might bias the findings, though, it is more appropriate to analyze the characteristics of nonrespondents before and after the event. To analyze if attrition and nonresponse might be affected by the event one would ideally rely on information originating from the sample frame. The sample frame generally provides information such as the age and gender of the sampled units, even of those that did not finally participate in the survey. However, in the case of the 7th round of the ESS in France, the survey organization was not granted access to the sampling frame. In the absence of this information we rely on the interviewers' judgment of the age, gender and contextual conditions of the house of the sampled units that were not interviewed during each period (we use the latter as a proxy for the socioeconomic status of nonrespondents). Relying on the interviewers' judgment of the characteristics of respondents imposes an additional limitation on our analysis of nonresponse, since interviewers can only judge the age and gender of those sampled units who refuse to participate in the survey. It is

not possible for interviewers to judge the age and gender of the sampled units that can not be contacted (unit nonresponse).

These analyses, summarized in Table D5, reveal no major differences in the characteristics of those who refused or could not be interviewed during the control and treatment periods. This suggests that our estimations are unlikely to be biased by attrition related to the attacks.[4]

Table D5: Characteristics of sampled units who refused or did not respond during the control and treatment periods

	Refusals			Unit nonresponse		
	Control	Treatment	t-stat.	Control	Treatment	t-stat.
Age (4 categories)	3.09	3.07	0.197			
Gender	0.41	0.45	-0.680			
House physical condition	3.08	3.03	0.513	2.99	2.93	1.156
Amount litter and rubbish	3.80	3.80	0.135	3.78	3.77	0.420
Amount vandalism and graffiti	3.83	3.79	0.673	3.85	3.82	0.991
Observations	127	191		467	693	

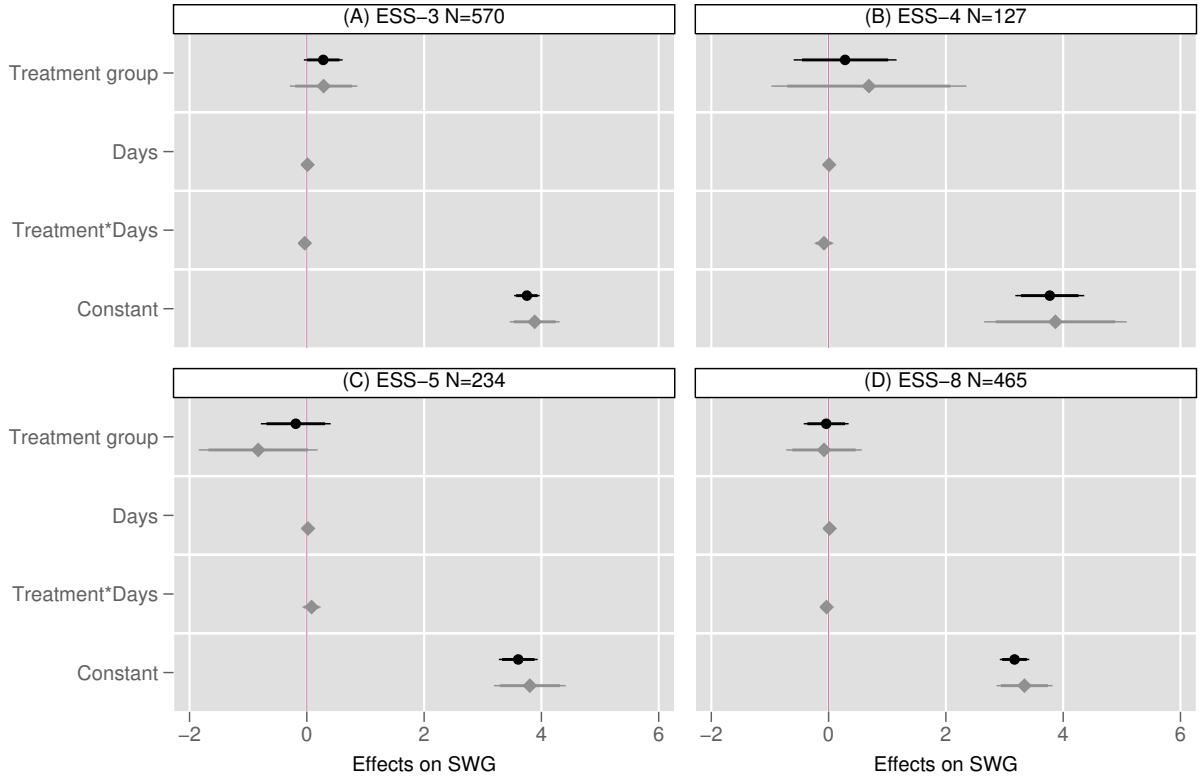
Note: Sampled units that refused or did not respond both during the treatment and control periods are excluded from this table. One observation per sampled unit. All variables are estimated by the interviewer.

[4]The difference in the number of observations between Table D5 and Table D4 is explained by the fact that in Table D5 each sampled unit is included just once, while in Table D4 each sampled unit might be included more than once.

D.7 Falsification tests based on other surveys

To test for the possibility that the estimated effects of the Charlie Hebdo attacks could be driven by aspects related to the timing of the survey administration or by any other cyclical trends related to people's satisfaction with government, we use data from four rounds of the French ESS in which the fieldwork period included January 7th.[5] For each round, we generate a placebo independent variable that mirrors our original treatment indicator, which compares those interviewed before and after January 7th with a bandwidth of ±20 days. The results of these tests summarized in Figure D10 reveal that in none of these rounds the placebo treatment has a statistically significant effect on satisfaction with government that is comparable in size to the effect that we identify for the Charlie Hebdo attacks.

Figure D10: Placebo tests with rounds 3, 4, 5, and 8 of the European Social Survey.



Note: Black spikes (circle markers) summarize results from OLS models that only include the treatment indicator. Grey spikes (diamond markers) summarize the results of models that include an interaction between the treatment indicator and the running variable measuring fieldwork days. Thick and thin lines are 90% and 95% confidence intervals, respectively.

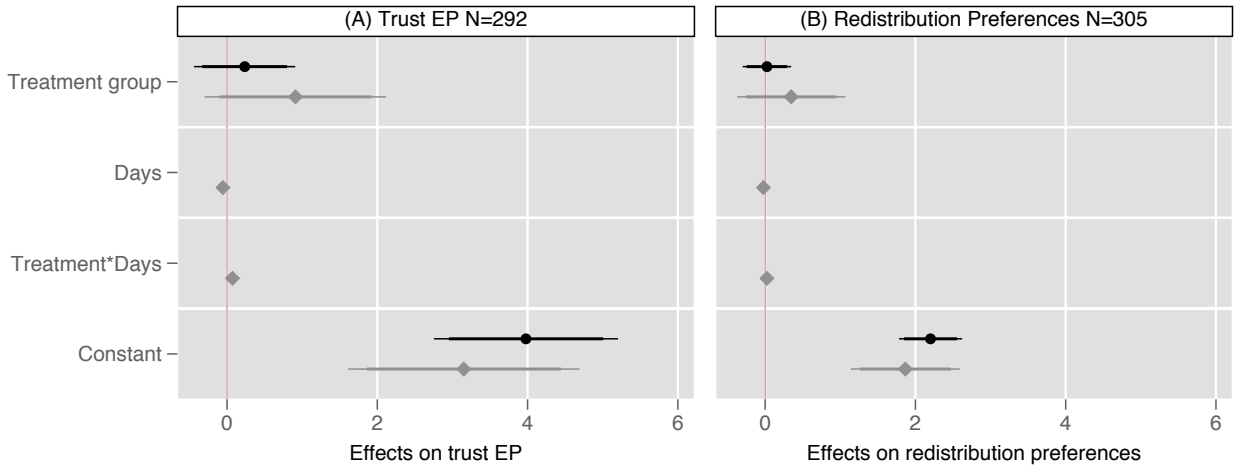
[5]The second round of the ESS in France also includes January 7th. However, we do not consider this round because the date of the interview was recorded differently than in the rounds that followed.

D.8 Falsification tests based on other outcome variables

We also test if the attacks had any impact on a series of variables that are close to our outcome of interest but should not be directly affected by the attacks. These variables measure respondents' trust in the European Parliament (EP) and their redistribution preferences. The first of these variables is selected because like our key variable of interest (satisfaction with government) it measures political support, but for an international institution that should not be directly affected by a "rally round the flag" effect. The second variable is selected in order to capture potential changes in preferences related to the support of a left-wing government.

The results of these tests are summarized in Figure D11. The effects of the treatment indicator are of reduced magnitude and not statistically significant when using any of these alternative outcomes.

Figure D11: Falsification tests: Effect of the Charlie Hebdo attacks on theoretically unrelated dependent variables.



Note: Black spikes (circle markers) summarize results from OLS models that only include the treatment indicator. Grey spikes (diamond markers) summarize the results of models that include an interaction between the treatment indicator and the running variable measuring fieldwork days. Thick and thin lines are 90% and 95% confidence intervals, respectively. The models are estimated with entropy balancing weights and region fixed-effects.

References in Supporting Information

Ares, Macarena and Enrique Hernández. 2017. "The corrosive effect of corruption on trust in politicians: Evidence from a natural experiment." *Research & Politics* 4(2):1–8.

Balcells, Laia and Gerard Torrats-Espinosa. 2018. "Using a natural experiment to estimate the electoral consequences of terrorist attacks." *Proceedings of the National Academy of Sciences* 115(42):10624–10629.

Berger, Eva M. 2010. "The Chernobyl disaster, concern about the environment, and life satisfaction." *Kyklos* 63(1):1–8.

Bischof, Daniel. 2017. "New graphic schemes for Stata: plotplain and plottig." *The Stata Journal* 17(3):748–759.

Bisgaard, Martin and Rune Slothuus. 2018. "Partisan Elites as Culprits? How Party Cues Shape Partisan Perceptual Gaps." *American Journal of Political Science* 62(2):456–469.

Bonilla, Tabitha and Justin Grimmer. 2013. "Elevated threat levels and decreased expectations: How democracy handles terrorist threats." *Poetics* 41(6):650–669.

Boomgaarden, Hajo G. and Claes H. de Vreese. 2007. "Dramatic Real-world Events and Public Opinion Dynamics: Media Coverage and its Impact on Public Reactions to an Assassination." *International Journal of Public Opinion Research* 19(3):354–366.

Boydstun, Amber E., Jessica T. Feezell and Rebecca A. Glazier. 2018. "In the wake of a terrorist attack, do Americans' attitudes toward Muslims decline?" *Research & Politics* 5(4):1–7.

Bozzoli, Carlos and Cathérine Müller. 2011. "Perceptions and attitudes following a terrorist shock: Evidence from the UK." *European Journal of Political Economy* 27(S1):S89–S106.

Branton, Regina, Valerie Martinez-Ebers, Tony E. Carey and Tetsuya Matsubayashi. 2015. "Social Protest and Policy Attitudes: The Case of the 2006 Immigrant Rallies." *American Journal of Political Science* 59(2):390–402.

Brouard, Sylvain, Pavlos Vasilopoulos and Martial Foucault. 2018. "How terrorism affects political attitudes: France in the aftermath of the 2015–2016 attacks." *West European Politics* 41(5):1073–1099.

Burlacu, Diana, Ellen M. Immergut, Maria Oskarson and Björn Rönnerstrand. 2018. "The politics of credit claiming: Rights and recognition in health policy feedback." *Social Policy & Administration* 52(4):880–894.

Carey Jr, Tony E., Regina P. Branton and Valerie Martinez-Ebers. 2014. "The influence of social protests on issue salience among latinos." *Political Research Quarterly* 67(3):615–627.

Coupe, Tom. 2017. "The impact of terrorism on expectations, trust and happiness – the case of the November 13 attacks in Paris, France." *Applied Economics Letters* 24(15):1084–1087.

Das, Enny, Brad J. Bushman, Marieke D. Bezemer, Peter Kerkhof and Ivar E. Vermeulen. 2009. "How terrorism news reports increase prejudice against outgroups: A terror management account." *Journal of Experimental Social Psychology* 45(3):453–459.

De Vries, Catherine E. 2018. *Euroscepticism and the future of European integration*. New York: Oxford University Press.

Dinas, Elias, Erin Hartman and Joost van Spanje. 2016. "Dead man walking: The affective roots of issue proximity between voters and parties." *Political Behavior* 38(3):659–687.

Dinesen, Peter Thisted and Mads Meier Jæger. 2013. "The effect of terror on institutional trust: New evidence from the 3/11 Madrid terrorist attack." *Political Psychology* 34(6):917–926.

Fernee, Henk, Ineke Stoop and Eric Harrison. 2014. "Coding media claims in the European Social Survey, Round 7." Manuscript, `https://www.europeansocialsurvey.org/docs/round7/methods/ESS7_media_claims_guidelines_ed_01_0.pdf`.

Finseraas, Henning, Niklas Jakobsson and Andreas Kotsadam. 2011. "Did the murder of Theo van Gogh change Europeans' immigration policy 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(1-2):213–228.

Flores, René D. 2018. "Can Elites Shape Public Attitudes Toward Immigrants?: Evidence from the 2016 US Presidential Election." *Social Forces* 96(4):1649–1690.

Geys, Benny and Salmai Qari. 2017. "Will you still trust me tomorrow? The causal effect of terrorism on social trust." *Public Choice* 96(3-4):289–305.

Hainmueller, Jens and Yiqing Xu. 2013. "Ebalance: A Stata package for entropy balancing." *Journal of Statistical Software* 54(7):1–18.

Hariri, Jacob Gerner, Christian Bjørnskov and Mogens K. Justesen. 2015. "Economic shocks and subjective well-being: Evidence from a quasi-experiment." *The World Bank Economic Review* 30(1):55–77.

Himmelweit, Hilde T, Marianne Jaeger Biberian and Janet Stockdale. 1978. "Memory for past vote: implications of a study of bias in recall." *British Journal of Political Science* 8(3):365–375.

Hofstetter, C. Richard. 1969. "Political Disengagement and the Death of Martin Luther King." *The Public Opinion Quarterly* 33(2):174–179.

Holbrook, Thomas M., Clayton Clouse and Aaron C. Weinschenk. 2012. "Bringing the president back in: The collapse of Lehman Brothers and the evolution of retrospective voting in the 2008 presidential election." *Political Research Quarterly* 65(2):263–274.

Jakobsson, Niklas and Svein Blom. 2014. "Did the 2011 Terror Attacks in Norway Change Citizens' Attitudes Toward Immigrants?" *International Journal of Public Opinion Research* 26(4):475–486.

Jensen, Carsten and Elias Naumann. 2016. "Increasing pressures and support for public healthcare in Europe." *Health Policy* 120(6):698–705.

Kim, Jin Woo and Eunji Kim. 2019. "Identifying the effect of political rumor diffusion using variations in survey timing." *Quarterly Journal of Political Science* forthcoming.

Krosnick, Jon A and Donald R Kinder. 1990. "Altering the foundations of support for the president through priming." *American Political Science Review* 84(2):497–512.

Larsen, Erik Gahner. 2018. "Welfare Retrenchments and Government Support: Evidence from a Natural Experiment." *European Sociological Review* 34(1):40–51.

Legewie, Joscha. 2013. "Terrorist Events and Attitudes toward Immigrants: A Natural Experiment." *American Journal of Sociology* 118(5):1199–1245.

Lever, H. 1969. "The Johannesburg Station Explosion and Ethnic Attitudes." *The Public Opinion Quarterly* 33(2):180–189.

Metcalfe, Robert, Nattavudh Powdthavee and Paul Dolan. 2011. "Destruction and distress: Using a quasi-experiment to show the effects of the September 11 attacks on mental well-being in the United Kingdom." *The Economic Journal* 121(550):F81–F103.

Minkus, Lara, Emanuel Deutschmann and Jan Delhey. 2018. "A Trump effect on the EU's popularity? The US presidential election as a natural experiment." *Perspectives on Politics* forthcoming:1–18.

Moran, Matthew. 2017. "Terrorism and the *banlieues* : the *Charlie Hebdo* attacks in context." *Modern & Contemporary France* 25(3):315–332.

Motta, Matthew. 2018. "The Polarizing Effect of the March for Science on Attitudes toward Scientists." *PS: Political Science amp; Politics* 51(4):782–788.

Muñoz, Jordi, Albert Falcó-Gimeno and Enrique Hernández. 2019. "Replication Data for: Unexpected Event during Surveys Design: Promise and Pitfalls for Causal Inference.". doi.org/10.7910/DVN/RDIIVL.

Muñoz, Jordi and Eva Anduiza. 2019. "'If a fight starts, watch the crowd' The effect of violence on popular support for social movements." *Journal of Peace Research* forthcoming.

Nagoshi, Julie L., Heather K. Terrell and Craig T. Nagoshi. 2007. "Changes in authoritarianism and coping in college students immediately after the terrorist attacks of September 11, 2001." *Personality and Individual Differences* 43(7):1722–1732.

Ojeda, Christopher. 2016. "The effect of 9/11 on the heritability of political trust." *Political psychology* 37(1):73–88.

Perrin, Andrew J. and Sondra J. Smolek. 2009. "Who trusts? Race, gender, and the September 11 rally effect among young adults." *Social Science Research* 38(1):134–145.

Pierce, Lamar, Todd Rogers and Jason A. Snyder. 2016. "Losing Hurts: The Happiness Impact of Partisan Electoral Loss." *Journal of Experimental Political Science* 3(1):44–59.

Reeves, Aaron and Robert de Vries. 2016. "Does media coverage influence public attitudes towards welfare recipients? The impact of the 2011 English riots." *The British Journal of Sociology* 67(2):281–306.

Schaffner, Brian F. and Cameron Roche. 2017. "Misinformation and motivated reasoning: Responses to economic news in a politicized environment." *Public Opinion Quarterly* 81(1):86–110.

Schulz, Peter J., Kent Nakamoto, Uwe Hartung and Carmen Faustinelli. 2008. "The death of Rosmarie Voser: The not-so-harmful consequences of a fatal medical error." *International Journal of Public Opinion Research* 20(3):386–397.

Silber Mohamed, Heather. 2013. "Can protests make Latinos "American"? Identity, immigration politics, and the 2006 marches." *American Politics Research* 41(2):298–327.

Silva, Bruno Castanho. 2018. "The Non-Impact of the 2015 Paris Terrorist Attacks on Political Attitudes." *Personality and Social Psychology Bulletin* 44(6):838–850.

Slothuus, Rune. 2016. "Assessing the Influence of Political Parties on Public Opinion: The Challenge from Pretreatment Effects." *Political Communication* 33(2):302–327.

Solaz, Hector, Catherine E. De Vries and Roosmarijn A. de Geus. 2018. "In-Group Loyalty and the Punishment of Corruption." *Comparative Political Studies* OnlineFirst:1–31.

Van Elsas, Erika J, Rozemarijn Lubbe, Tom WG van der Meer and Wouter van der Brug. 2013. "Vote recall: A panel study on the mechanisms that explain vote recall inconsistency." *International Journal of Public Opinion Research* 26(1):18–40.

Willsher, Kim. 2015. "France boosts anti-terror measures in wake of Paris attacks.". The Guardian `http://www.theguardian.com/world/2015/jan/21/france-anti-terror-measures-paris-attacks-manuel-valls`.

Zaller, John and Mark Hunt. 1995. "The rise and fall of candidate Perot: The outsider versus the political system—part ii." *Political Communication* 12(1):97–123.

Zepeda-Millán, Chris and Sophia J Wallace. 2013. "Racialization in times of contention: How social movements influence Latino racial identity." *Politics, Groups, and Identities* 1(4):510–527.