

PROTEST EVENT ANALYSIS MEETS AUTOCRACY: COMPARING THE COVERAGE OF CHINESE PROTESTS ON SOCIAL MEDIA, DISSIDENT WEBSITES, AND IN THE NEWS*

Christian Göbel and H. Christoph Steinhardt†

How accurate is media-elicited protest event data from autocracies where the state censors the media? Based on a source-specific model of event selection and a multisource dataset of over 3,100 protests from three Chinese megacities, we demonstrate the substantial advantages of using social media data, capturing 115 times more protests than English-language international news, 74 times more than domestic news, and 10 times more than dissident websites. Social media are most likely to cover small and nonviolent events that other sources often ignore. Aside from antiregime protests, they are less affected by censorship than often assumed. A validity test against public holidays and daily rainfall shows that social media data outperform dissident websites and traditional news. Social media, and to a lesser extent dissident media, are promising new sources for protest event analysis in autocracies. Scholars should treat news media-based event data from heavily censoring regimes with caution.

Driven by the improved access to media-based event data, quantitative research on protests and social movements has increased exponentially,¹ and the focus of analysis has diversified. Previously confined mainly to industrialized democracies in Western Europe and North America, protest event research now covers authoritarian regimes worldwide.² Another noteworthy phenomenon is the recent availability of protest- and conflict-event datasets that stretch across continents and regime types.³ These developments promise unprecedented comparative insights into the dynamics of popular protests. Yet how good are media-based event data from autocracies?

It is well established that news media-generated protest data are affected by selective coverage.⁴ This bias is most severe in the international news media (Mueller 1997; Nam 2006), on which most transnational datasets rely heavily. In democracies, the domestic news media are a much richer alternative source (Nam 2006). In autocracies, the usefulness of the domestic news is probably severely undermined by censorship. Scholars of contentious politics in such contexts have begun extracting event data from dissident websites and social media. To date, however, relatively little is known about the quality of event data of any source in an authoritarian context.

More broadly, existing research has not yet sufficiently addressed three important issues. The first is the quality of new sources. Scholars of protest in democracies and non-democracies increasingly draw on political activist websites.⁵ However, research on the value of protest event data from these sources is in its infancy (Almeida and Lichbach 2003; Zhang and Pan 2019). The second is the comparative performance of different types of media. Examinations of event data tend to be restricted to one type and therefore offer no insights into how patterns of event selection differ by media category.⁶ The effect of censorship on protest event data is a third issue that

* The research for this article was funded by the Hong Kong Research Grants Council, Early Career Scheme (Grant No. 24615215), the European Research Council (Grant No. 678266), and the Chiang Ching-kuo Foundation for International Scholarly Exchange (Grant No. CS002-U-18). The authors would like to thank Anita Gohdes, Swen Hutter, Tianguang Meng, Zachary Steinert-Threlkeld, Yang Su, Nils Weidmann, and five anonymous reviewers. They provided very valuable feedback on previous drafts of this article. The authors dedicate this article to Lu Yuyu and Li Tingyu for their path-breaking contribution to science.

† Christian Göbel (ORCID 0000-0002-6940-6127) is professor and H. Christoph Steinhardt (ORCID 0000-0003-1521-8566) is associate professor in the Department of East Asian Studies, University of Vienna. Please direct all correspondence to: hc.steinhardt@univie.ac.at.

scholars have not yet sufficiently addressed. Most of the existing methodological research has analyzed media coverage of protests in democracies,⁷ so research has not systematically conceptualized and assessed the impact of censorship on protest event data. Likewise, the comparative utility of the domestic and the international news media remains unexamined in an autocratic environment.

Building on the existing insights on the coverage of protests by traditional news media, we seek to drive the debate forward by developing a source-specific model of protest coverage under conditions of information control. Our model divides the event selection process into three stages: information gathering, reporting, and control. We then derive four hypotheses from it and test them on 3,107 hand-coded and nonredundant protests in three Chinese megacities between 2014 and 2016. We collected data from social media, dissident websites, and domestic and international news. We follow previous research by comparing different sources to examine differences in coverage patterns (Almeida and Lichbach 2003; Dowd et al. 2020; Mueller 1997; Strawn 2008). We then compare daily protest- and participant frequencies with two objective external benchmarks: public holidays and rainfall. We choose this approach because the often-used ‘ground truth’ of police permits for demonstrations is neither available nor meaningful in China. Here and in other autocracies where protest is not institutionalized, demonstrations with police permits are rare. Permit data would therefore not be helpful even if the state would publish it.

Our findings demonstrate that social media coverage is several times broader than any other source. Compared to data from the other source types, social media data are less affected by the event intensity (Myers and Caniglia 2004; Snyder and Kelly 1977) and aligns most accurately with objective external benchmarks. We find much less evidence of government censorship in social media than is often assumed. However, we do find strong indications that regime-challenging protests are censored. Although China is a specific case, it arguably represents a challenging test for the quality of social media-elicited event data in autocracies. It is one of the few authoritarian countries consistently blocking Twitter and Facebook.⁸ Only local surrogates are allowed, and these are tightly controlled by the world's most sophisticated online censorship system (Fu, Chan, and Chau 2013; King, Pan, and Roberts 2013; Roberts 2018; Qin, Strömberg, and Wu 2017). As social media emerge as promising sources even under these conditions, there is reason to believe that social media may be even more useful in autocracies where searchable international platforms such as Twitter are not banned.

In the following, we spell out our source-specific model of protest event selection and infer hypotheses from it. Then, we describe our methods and data. The empirical section compares our four types of sources for total coverage, selection on intensity, and selection on political issues. After that, we compare sources against objective external benchmarks. The conclusion summarizes the main findings and their implications for protest event research.

A SOURCE-SPECIFIC MODEL OF PROTEST EVENT SELECTION

Our model of event selection distinguishes between three stages: gathering, reporting, and controlling information. The result of these selection rounds is whether a particular media type covers a protest or not. Our distinction between stages is necessary because we study sources of a very different nature, which are subject to varying conditions at all stages. Selection during information gathering equals the likelihood that the media actors (international and domestic journalists and their editors, editors of dissident websites, and social media users) in a given media type will learn about a protest event. At the information reporting stage, selection equals the likelihood that these actors will report a protest they have learned about. Finally, information control through censorship adds a third stage of selection. We summarize the comparative strength of event selection effects during the three stages and their associated selection mechanisms on the four media types in table 1.

Table 1. Source-Specific Model of Protest Event Selection

<i>Selection Stages</i>	<i>Mechanism</i>	<i>International Media</i>	<i>Dissident Media</i>	<i>Domestic Media</i>	<i>Social Media</i>
Information Gathering	Group size	***	****	**	*
	Scope	****	***	**	*
	Informational distance	****	**	***	*
Information Reporting	Editorial space	****	**	***	*
	Editorial inclination	***	***	****	*
Information Control	Censorship	*	**	****	***

Note: Comparative strength of selection effects from weakest * to most substantial ****.

Information Gathering

The likelihood that media actors gather information about protests is a function of the *size of their group*, the geographical and thematic *scope* of their attention, and their *informational distance* to protests (in particular, to what extent they have access to event participants and eye-witnesses, or have to rely on indirect information from other media).

International media correspondents are a small group that reports on whole countries—China in the present case. The amount of time they can spend on protests is minimal. Much of their work involves filtering information they receive from their contacts, the domestic news, and social media. They learn about protests not primarily because they actively look for them but because a protest event is important enough to be picked out among the deluge of information that competes for a reporter’s attention (Almeida and Lichbach 2003: 254; Interview with China reporter of a major international news organization, December 16, 2018).

The editors of dissident websites are likely a smaller group than international journalists. Although they also have to pay attention to the whole country, they focus on the narrow spectrum of information they consider helpful for their antigovernment cause. Protests belong to this type of information. Moreover, dissidents are also embedded in networks devoted to sharing such information. Hence, dissident websites can be expected to gather more information on protest events than the international media (Almeida and Lichbach 2003: 253–54).

In China, the group of domestic journalists far exceeds that of international journalists and dissident website editors.⁹ While the national press also attends to the whole country, journalists working for local newspapers focus on smaller geographical areas. Locally embedded domestic journalists are likelier to learn about events through participants or bystanders than their international colleagues. Sometimes protesters actively reach out to local media because they know that public attention increases the likelihood of the issue being resolved in their favor (Cai 2010: 28). In sum, domestic reporters can be expected to gather more protest information than either international journalists or dissidents (Almeida and Lichbach 2003; Nam 2006: 283).

Social media users are the largest group of all media actors (note 9). Thus, the scope any one user pays attention to is the smallest among all media types. It is reasonable to assume that for most protest events, there will be at least one social media user among the protesters or spectators. Therefore, social media users’ informational distance to events is smaller than for the actors in any other media type. Hence, we expect that social media gather more protests than any other media type.¹⁰

While the above dynamics affect the share of all protests a media-type will gather, some protest types likely diverge from this broad pattern. We assume that media actors are likelier to learn about events that generate attention in their information channels. The most consistent determinant of the attention-worthiness of a protest is event intensity (e.g. Almeida and Lichbach

2003; Barranco and Wisler 1999; Myers and Caniglia 2004; Snyder and Kelly 1977)— such as a protest's size, the extent of violence or the presence of repression. The larger a media-type's informational distance from protests, the stronger the effect of the event intensity is on gathering event information. Thus, international and domestic media should be most affected by intensity during the gathering, dissident media somewhat less so, and social media the least. Moreover, in the context of an authoritarian regime, political protests attract additional attention among certain groups. Dissident media and, to a lesser extent, international journalists can be expected to be plugged into such channels of information and thus have a higher likelihood of learning about political protests than the other two media types.

Information Reporting

The likelihood that media actors in a given media-type report the protests they have learned about is a function of editorial space constraints, which limit the number of protests that can be published, and the editorial inclination decides how the space is filled.

International media are most constrained by space. Protests in a given country must compete with all global events. Hence, international news can publish the least amount of events and select most strongly on event intensity (Mueller 1997). Moreover, as antiregime protests are more interesting to an international audience than bread and butter protests, these kinds of protests can also be expected to be reported disproportionately often in international news. Consequently, domestic news media outlets can devote more space to protests in their country than international ones (Nam 2006: 282–83). However, in China and other regimes with information control, the domestic media's editorial decision making is subsumed by the logic of information control.

Dissident media are less pressed for space than traditional news media. Websites or blogs do not have limited page numbers, and their focus is narrow (Almeida and Lichbach 2003: 254). Hence, dissident media should publish more events and be less affected by event intensity than the news media. However, because of their political agenda, we expect dissident sources to be much more inclined to report political protests (Lankina 2015; Li 2017; Robertson 2013).

Social media users have no strong editorial inclinations that would have them prefer one type of protest over another. Space on social media is practically unlimited, and the cost of snapping a picture and writing a post about a protest event one witnesses is negligible. Moreover, most social media users will only occasionally witness a protest. Therefore, a post about a protest will elicit interest among a user's followers. All this makes it likely that a very high proportion of events gathered by social media users will also be reported (or posted). Social media in China, however, are also affected by information control.

Information Control

Information control is a complex amalgam of state *censorship* of publications (both before and after reporting), state censorship of sources that media actors rely on, and self-censorship, which is hard to disentangle. International news media and dissident blogs are not subject to state censorship because they are published or hosted outside of the countries they cover (Li 2017: 8). Nonetheless, international journalists, as well as editors and informants for dissident websites, sometimes face state harassment, which may induce self-censorship. In China, state interventions in protest reporting by international news media are rare but happen for highly sensitive events (Interview 2018). Since dissident editors and informants do not have the protection of a foreign passport and the official journalist status, the impact on them is probably stronger. Both media types are affected by the censorship of their domestic sources. Overall, however, we expect the combined censorship effect on these two sources to be much weaker than on the other two.

In democracies, the domestic news media's advantages during information gathering and reporting can make it a very efficient source for event collection (e.g., Barranco and Wisler 1999; Oliver and Meyer 1999)—and a much better one than international news (Nam 2006). In autocracies, however, the decision to report a protest or not is dominated by state censorship and

self-censorship. For example, in China, reports on protests are no longer completely absent from the news, but they remain a sensitive topic that is reported at best sporadically (Huang, Boranbay-Akan, and Huang 2019; Lorentzen 2014; Steinhart 2015). Hence, we expect that censorship erases the vast comparative advantages that domestic media has vis-à-vis international media in uncensored contexts.

The most popular Chinese social media platforms are Sina Weibo and Weixin. They are subject to censorship rules that are both strict and opaque. Social media companies are required to set up censorship departments to implement government demands (Fu, Chan, and Chau 2013; Qin, Strömberg, and Wu 2017: 120–22), but it is doubtful if they can filter out all sensitive content among the hundreds of millions of posts created by users every day. Social media users probably know that protests are considered sensitive by the government, which might induce self-censorship. However, Molly Roberts shows that awareness of censorship does not dissuade but encourage users to post about such topics (Roberts 2018: ch. 4). Scholars have also found no evidence that users open separate accounts for posts with protest-related messages or that accounts are discontinued after posting such content (Qin, Strömberg, and Wu 2017: 127). According to a recent estimate, less than 7% of protest-related posts on Weibo are censored after posting (Zhang and Pan 2019: 38). Research finds that protest-related content is available on a very large scale on Chinese social media (Qin, Strömberg, and Wu 2017; Zhang and Pan 2019). Thus, we expect the overall censorship effect for social media to be less severe than for domestic news.

Research has not yet examined how protest type affects the strength of censorship in domestic news and on social media in China. We expect that domestic news will not report on political protests and that social media posts on such events are particularly likely to be erased. Moreover, in their study of censorship on discussion forums, King, Pan, and Roberts (2013) find that protest-related content is censored when it attracts high user attention. Suppose their observation is also valid for social media posts. In that case, protest intensity should increase the degree of attention an event receives on social media and, in turn, strengthen the effect of censorship. Scholars have also observed that protests and other events that generate attention online often push the censored Chinese domestic media to report them (Hassid and Repnikova 2016, 388–89; Lorentzen 2014, 410; Steinhart 2015: 129–31). Thus, in contrast to social media, we expect that event intensity in the domestic news event diminishes the effect of censorship.

HYPOTHESES

The vast advantages social media have over all other media types during information gathering and reporting, combined with the limited censorship effect we expect, should mean that they cover the largest number of protests among all media types. Conversely, despite comparative advantages during information gathering and reporting, the heavy censorship of the domestic media should result in their covering the fewest events of all. Although dissident websites and international media are not directly censored, we expect that the relative disadvantages during gathering and reporting should lead these sources to cover fewer protests than social media but more than domestic media. Furthermore, although the number of dissident website editors is smaller than international journalists, dissident websites should have an edge over international news media due to their smaller geographical scope and informational distance to protests (gathering stage) and their larger editorial space (reporting stage). Table 2 on the next page summarizes all our hypotheses and their operationalizations, which we discuss below.

H1: Total amount of events covered: social media > dissident media > international media > domestic media.

Because protest intensity is the only known factor that might penetrate the censorship wall for domestic media in China, we expect selection into domestic media to be most affected by intensity. International media should also be heavily inclined to select on intensity. Due to its large informational distance to protests, it relies heavily on indirect information from other media

Table 2. Hypotheses, Methods, Dependent Variables and Key Independent Variables

<i>Hypotheses</i>	<i>Methods, Dependent and Independent Variables</i>
<i>H1:</i> The total amount of events covered: Social media > dissident media > international media > domestic media	<i>Methods:</i> Frequencies <i>DVs:</i> Event selection into media-types <i>IVs:</i> None
<i>H2:</i> Likelihood of intense protest to be selected: Domestic media > international media > dissident media > social media	<i>Methods:</i> Logistic regressions <i>DVs:</i> Event selection into media-types <i>IVs:</i> Large size, protester violence, state coercion
<i>H3:</i> Likelihood of political protests to be selected: Dissident media > international media > social media > domestic media	<i>Methods:</i> Logistic regressions <i>DVs:</i> Event selection into media-types <i>IVs:</i> Discontent with state and regime
<i>H4:</i> Association between event data and public holidays/daily rainfall: Social media > all other sources	<i>Methods:</i> Poisson and negative binomial regressions <i>DVs:</i> Protests per city-day, total and average participants per city-day <i>IVs:</i> Public holidays, daily rainfall

during information gathering. Moreover, international media have the least amount of editorial space for protests during the reporting stage. Both factors should increase the impact of intensity. Compared to international news, dissident media have a shorter informational distance to protests during a gathering and is much less pressed for space during reporting. It should therefore be less likely to select on intensity. Finally, because of users' close proximity to event information during a gathering, almost unlimited space, and few editorial inclinations during reporting, in addition to increased censorship scrutiny for high-intensity events, we assume that the likelihood of intensive protests to be selected into social media is the smallest.

H2: Likelihood of intense protest to be selected: domestic media > international media > dissident media > social media.

As dissident media actors are plugged into dedicated information streams during a gathering, and follow a political editorial agenda during the reporting stage, we expect that the impact of a political issue on event selection will be most pronounced. International journalists, too, are close to information on political protests during a gathering, and their editors have incentives to select these protests for their international audiences during reporting. Therefore, censorship should target political events on social media and make them practically absent from domestic media.

H3: Likelihood of political protests to be selected: dissident media > international media > social media > domestic media.

Our model suggests that social media have enormous advantages over all other media types during information gathering and reporting. Information control will somewhat diminish these advantages, but we expect the overall effect to be limited. Hence, we presume that protest event data gathered from social media reflect real-world protests more closely than data from any other media type.

Without reliable data on the true population of protests, we examine our data against two objective external benchmarks: public holidays and daily rainfall. Educational institutions, most businesses, construction sites, and government offices are closed during public holidays. In addition, many people use the occasion for traveling. Hence, most of the protest targets and many potential participants are unavailable (Steinhardt 2021: 563). Rainfall is increasingly used in

protest event analysis to instrumentalize protest mobilization (Negro and Olzak 2019; Ritter and Conrad 2016). The idea is that rain will impose an additional cost on potential protesters and thus depress turnout. Therefore, protest organizers should be disinclined to mobilize on national public holidays and rainy days, and the number of protests should be reduced. If protests take place on such days, turnout should be depressed.

H4: Association between event data and public holidays/daily rainfall: social media > all other sources.

DATA

We test our hypotheses with a hand-coded dataset of 3,107 protest events in three Chinese megacities (Guangzhou, Shanghai, and Chongqing) between January 2014 and May 2016. The three cities are regional political centers and the leading commercial and cultural hubs of Southern, Eastern, and Western China. They provide some of the best conditions for protests to occur and to be covered media sources. First, the cities are among those with the largest population in the country. Moreover, Chinese protesters target regional political centers (Chen and Cai 2021). Therefore, the number of protests should be higher than in most other locations. Second, media penetration (Internet, mobile phone, and news media bureaus) is most pronounced in large cities that perform important commercial, cultural and political roles, and news media tend to pay more attention to protests in such cities (Myers and Caniglia 2004). These data collection sites thus provide ‘most likely’ conditions for media coverage. If a source performs well here, it does not imply it performs equally well elsewhere in China. However, if a source performs poorly here, it will most likely perform even worse elsewhere.

We follow Tilly’s definition of a protest event as a contentious gathering: ‘an occasion on which a number of people (here, a minimum of ten) outside of the government gathered in a publicly accessible place and made claims on at least one person outside their own number, claims which if realized would affect the interests of their object’ (Tilly 1993: 270). We adhered to the Dynamics of Collective Action project manual to avoid duplication of protest gatherings. Hence, we coded gatherings for which the protest location, group, and issue were the same, and there were not more than twenty-four hours between them as one protest event. When a gap between an event and a preceding event was larger than twenty-four hours but the other conditions applied, the protests were recorded as separate events in a series of protests (McAdam, McCarthy, Olzak, and Soule 2009).

We collected data from four types of sources: (1) international English and Chinese language news media, which include news media outlets from Hong Kong and Taiwan; (2) dissident media; (3) domestic news media; and (4) social media. Apart from twenty-five cases from nondomestic English news media coded by one of the authors, three trained research assistants independently coded the data under close supervision by one of the authors. After establishing sufficient interrater reliability during coder training, we retested reliability on five occasions and 139 cases throughout the coding process to ensure that coding remained consistent. Among these cases and the coded variables used here, Brennan and Prediger’s κ ranged from 0.89 to 0.98 and percent agreement from 95% to 99% (for details, see appendix A at <https://phaidra.univie.ac.at/o:1424766>).

We manually extracted international English-language news media articles through keyword queries from the LexisNexis database, which contains more than 1,000 global news outlets (see appendix B at <https://phaidra.univie.ac.at/o:1424766> for the list of keywords). Using similar procedures, we collected international Chinese and domestic media from the Wisenews database, which contains over 600 Chinese-language newspapers from Mainland China, Hong Kong, Macao, Taiwan, and South East Asia. In addition, we screened three major Chinese language news media platforms (BBC Chinese, Duowei News, and Radio Free Asia Chinese) manually or with search engine keyword queries, depending on the source. Finally, we collected Chinese dissident media sources from six websites run by political dissidents and well known for reporting sensitive political news and protests in Mainland China (Boxun, China

Labour Bulletin, Canyu, Minsheng Guancha, Zhongguo Molihua Geming, and 64tianwang). Again, we screened them manually and through search engine queries. Typically, cross-national protest event databases examine only English language news, and studies using dissident media draw on only one source. Hence, our international and dissident media collection can be considered exceptionally broad.

We gathered events posted on social media data from the Wickedonna blog (for its description, see also Göbel 2021), which consists of more than 200,000 social media posts related to more than 70,000 protests, collected and published online by activists Lu Yuyu and Li Tingyu. The collection begins in July 2013 and ends in June 2016, when Lu and Li were arrested (China Labour Bulletin 2017). The activists manually searched primarily Sina Weibo for material, mostly eyewitness accounts of a particular protest. They inferred protest issues and location from the text of posts and examined each post's veracity by cross-checking with other sources, sometimes by contacting the author. They wanted to collect all protest events independent of issue, locality, intensity, or other factors. After verification, they published one entry per event on their website (Wang 2014). Protests from the three cities and the observation period were extracted from the website through manual inspection. Although a comparison with a protest detection algorithm applied to data collected from Weibo indicates that the two activists caught most of the protests on this platform (Zhang and Pan 2019), they used semi-systematic collection. Thus, our collection from social media is certainly not optimized.

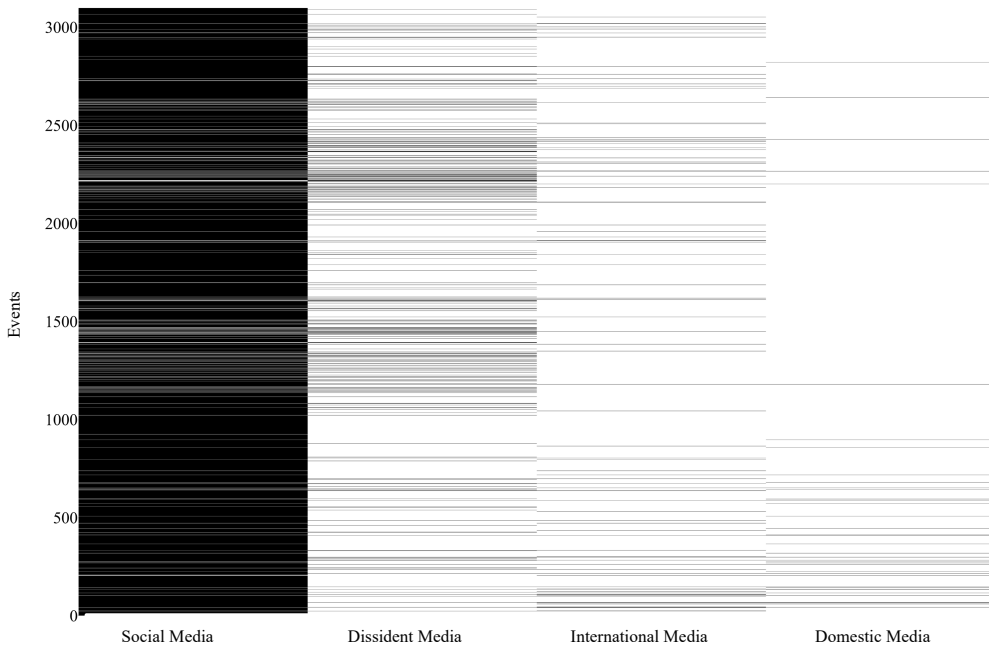
We measure our key variables as follows: our dependent variables to test hypotheses 1 to 3 are whether a protest was selected into one of the four source types, measured as dummy variables as described above. However, we did not include China Labour Bulletin as a dissident website because 96% of their events were collected from the Wickedonna blog. To test hypothesis 4, we use the number of events per city-day and the total and average protest participants per city-day as our dependent variables (see online appendix A). To tap event intensity, we draw on three commonly used indicators—large size (participants ≥ 500), protester violence against persons, and state coercion. Previous event research from China has shown that most protests are very small (Göbel 2019; Göbel and Ong 2012). In our sample, the mean number of participants is 113, and the median is thirty. Just 4% of events have at least 500 participants. To measure political issues, we distinguish between protests driven by discontent with the state (conflicts with administration and law enforcement) and those driven by dissatisfaction with the political regime (protests with explicit claims on political, civil, or human rights or protests in support of dissidents). We coded national public holidays as a dummy variable in line with official announcements by the Chinese central government. Daily rainfall data is not available from Chinese meteorological institutions. Hence, we measure rainfall with the twenty-four-hour precipitation recorded in the US National Oceanic and Atmospheric Administration (NOAA) Daily Summaries. We log transform this measure because we expect rainfall's negative effect on contentious action decrease with each additional unit of rain.¹¹ Summary statistics and further details on variable coding can be found in the online appendix available at <https://phaidra.univie.ac.at/view/o:1424766>

FINDINGS

We test our hypotheses in three steps. First, we examine the total number of events selected by the four media types. Second, we ascertain their “relative bias” (Strawn 2008; see also, Almeida and Lichbach 2003; Mueller 1997) by examining how the intensity and political issues affect the selection of events into these sources. Third, we test how event data generated from the different sources perform against public holidays and daily rainfall.

Events covered

The share of events covered by each source is the first yardstick for judging data quality (figure 1). Social media capture 93.24% (2,897 events) of all protests in our sample, and dom-

Figure 1. Events Covered by Different Sources

Note: Dark coloring indicates that an event was covered by the source listed on the x-axis

estic media accounts for 1.26% (thirty-nine events). International media catch 3.06% (95 events). Chinese international media cover most of these (2.70%, 84 events). English-language international media perform even worse than domestic media—they catch just 0.80% (twenty-five cases). Dissident media fall between these extremes, albeit towards the lower end—8.98% of all events (279 cases) appear in this source. All sources cover a certain number of events exclusively. 88.70% of all events are exclusively covered in social media, 5.12% in dissident media, 0.45% in international media, and 0.58% in domestic media. This implies that, in all likelihood, the actual number of protests is larger than the combined sample.

These findings support our hypothesis 1. The superior performance of social media is rooted in its vast advantages in information gathering and reporting. The only alternative explanation would be that a much lower level of censorship in social media than other sources permits it to capture most events. This is implausible because social media is censored directly, while international and dissident media are not. Conversely, the meager coverage by the domestic news is hard to account for without censorship. Without censorship, domestic media should have manifest advantages during information gathering and reporting. Although it is more difficult to pinpoint the exact reasons for the low coverage by international and dissident media, their relative performance is in line with our expectations. The effect of indirect censorship should be equal for both sources, but dissident sources should have an edge during information gathering and reporting.

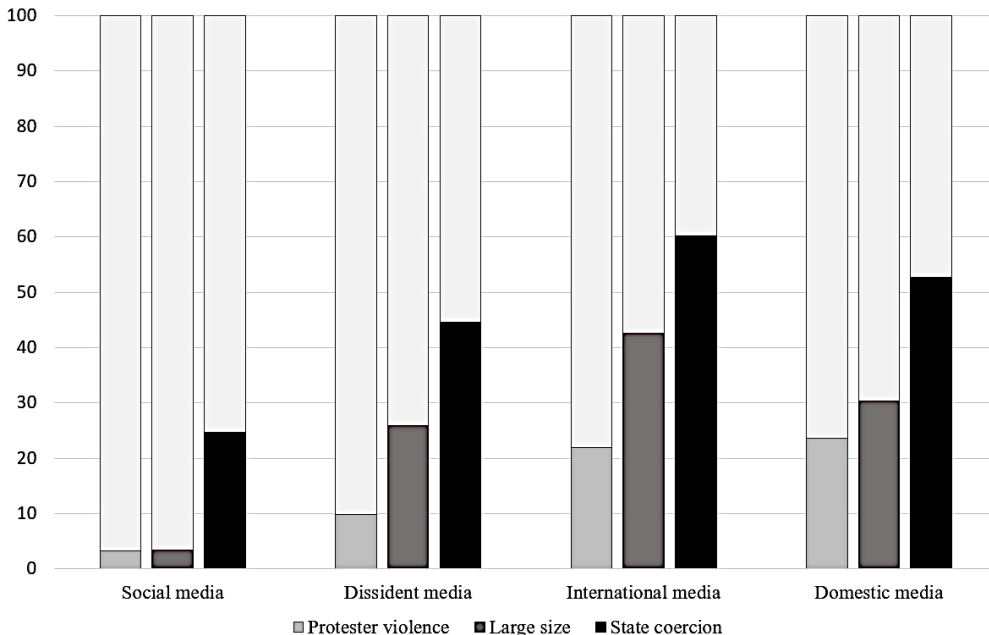
It is worthwhile to put the findings on the domestic media in context. A study of protests from the 1960s to the 1990s in Swiss cities reported a coverage rate of 80% in four domestic newspapers (Barranco and Wisler 1999); another, focusing on protests in a US state capital in the 1990s, found coverage of 46% in two domestic newspapers (Oliver and Meyer 1999). For Minsk in Belarus during the transition from communism, scholars calculated coverage of 38% in domestic news (McCarthy et al. 2008). For protests in Washington DC in 1982 and 1991, a study elicited coverage of 13% and 7%, respectively, in two domestic newspapers and three television channels (McCarthy, McPhail, and Smith 1996).

All these studies benchmark fewer sources than we use against entire populations of protests. We include more sources, but our benchmark consists of less than the complete population of events. All else equal, we should elicit higher coverage rates for domestic news media than these studies. What we find are lower coverage rates. The proportions of protests in Chinese urban centers covered by domestic media are therefore dismal by any standard. The coverage by international news media is similarly scant. Our findings on domestic media also stand in sharp contrast to their superior performance vis-à-vis international sources in a democratic context (Nam 2006).

Event intensity

In a first step, we examine the proportions of events with high intensity covered in the four different source types (figure 2). Our hypothesis 2 proposes that domestic media should be most affected by intensity, international and dissident media less, and social media the least. In line with our expectations, social media select much higher proportions of small and nonviolent events than the other sources. No clear-cut differences between the other three sources are discernable.

Figure 2. Percentage of High-Intensity Events among Protests Covered by Different Sources



We then run logistic regressions with selection/nonselection in the different media sources as dependent variables and high-intensity indicators and controls as the independent variables (table 3). Odds ratios for all three intensity indicators are significant for dissident, international and domestic media, and the explained variances are relatively large. Dissident and international media have the highest odds ratios for large size. International and domestic media have the highest odds ratios for protester violence and state coercion. Hence, contrary to our hypothesis 2, domestic media are not the media type most affected by intensity. Other criteria besides intensity thus seem to be more critical for domestic media when they decide whether to break the usual silence on protests or not. Instead, over all three indicators, international media display the highest susceptibility to intensity. The intensity of protest size also affects dissident media more than we expected. The large effects of event size in international and dissident media may reflect that small protests are not regarded as sufficiently interesting by the editors of these sources.

Table 3. Event Selection and Protest Intensity

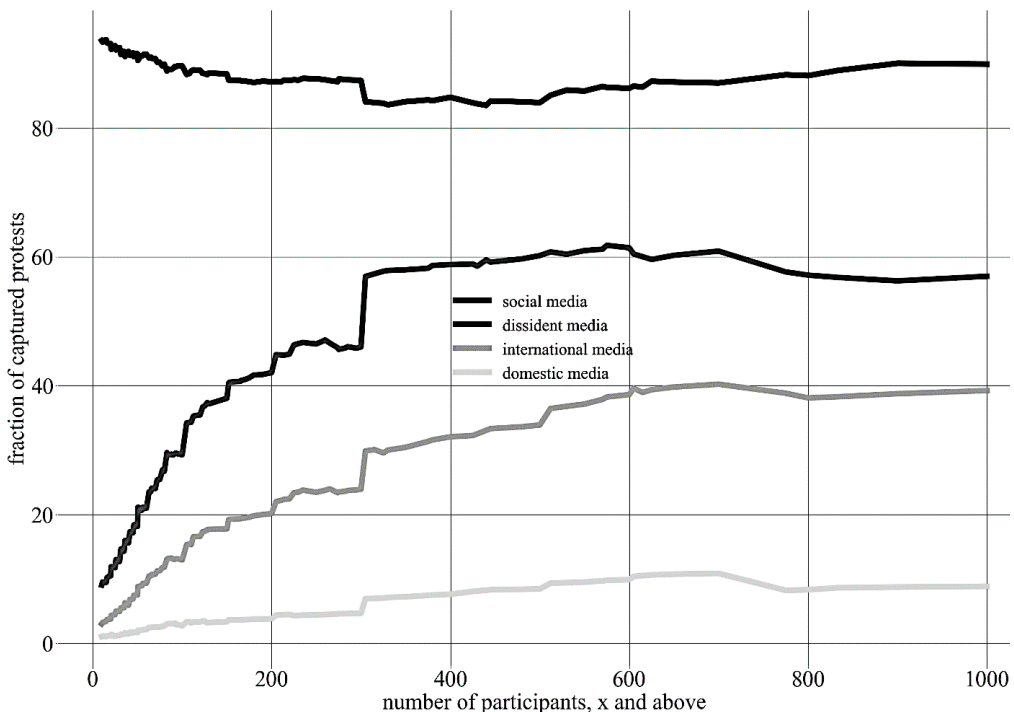
	<i>Social Media</i>	<i>Dissident Media</i>	<i>International Media</i>	<i>Domestic Media</i>
Large protest	0.35 (0.10)**	14.68 (3.39)**	13.80 (3.76)**	4.29 (2.10)*
Protester violence	1.26 (0.53)	1.74 (0.49)*	4.39 (1.52)**	4.86 (2.68)*
State coercion	1.10 (0.20)	1.90 (0.29)**	3.37 (0.88)**	2.55 (1.07)*
Protest series	Y	Y	Y	Y
Days (log)	Y	Y	Y	Y
City	Y	Y	Y	Y
Constant	29.33 (6.06)**	0.03 (0.01)**	0.50 (0.14)*	0.02 (0.01)**
N	2,298	2,298	2,298	2,298
Pseudo R2	0.05	0.17	0.27	0.27

Note: ** p ≤ 0.001, * p ≤ 0.05. Cell entries display logit regression odds ratios and standard errors in brackets.

For social media, protester violence and state coercion have no significant effect on the likelihood of coverage, while large size has a significant negative effect. The explained variance is by far the smallest. This is in line with our expectations, and underlines social media’s vastly lower selective pressure in the information gathering and reporting stages.

Has censorship, which we expected to result in negative selective pressure for high-intensity events in social media, contributed to the patterns we identify? Social media have caught 83.19%, 90.61%, and 92.16% among large, violent, and coercive protests, respectively. The share of large events is notably lower than its overall share of events covered (93.24%). We, therefore, examine the coverage of events at different sizes in greater detail. Figure 3 depicts the proportions of events covered by different sources when the protests become larger. For social media, the share of events covered declines from 92.73% for events ≥ twenty protesters to 83.19% for ≥ 500 protesters. For larger events, the proportion reverses to 88.75% (≥1000 protesters). By contrast, the curves for all other media types point upwards and then flatten as protests become larger than 600 participants.

Figure 3. Percentage of Events Covered by Number of Participants and Source Type



Based on these findings, we cannot exclude that censorship contributes to the slight decline of social media's proportion of events covered in the middle ranges. At the same time, our finding could result from other media types catching more events as event size increases. In any case, we find no compelling evidence that high-intensity protests on social media are subject to extensive censorship.

Political Events

In terms of selective pressure on specific protest issues, we expected to find the most clear-cut differences between sources in terms of their coverage of political protests. To further differentiate political protests in terms of political sensitivity, we distinguish between “discontent with the state” and “discontent with the regime.”

Figure 4 displays the relative shares of these two types of protest issues in the different sources. As we expected in hypothesis 3, domestic media cover no regime-challenging protests and report the smallest share of discontent with the state (10.35%). In comparison, dissident media cover the largest share of political protests (38.02% are protests against the state and 15.97% contain claims against the regime). Social and international media lie in between these poles. Notably, protests featuring discontent with the government make up just 0.60% of events on social media.

Figure 4. Percentage of Political Events Covered by Different Sources

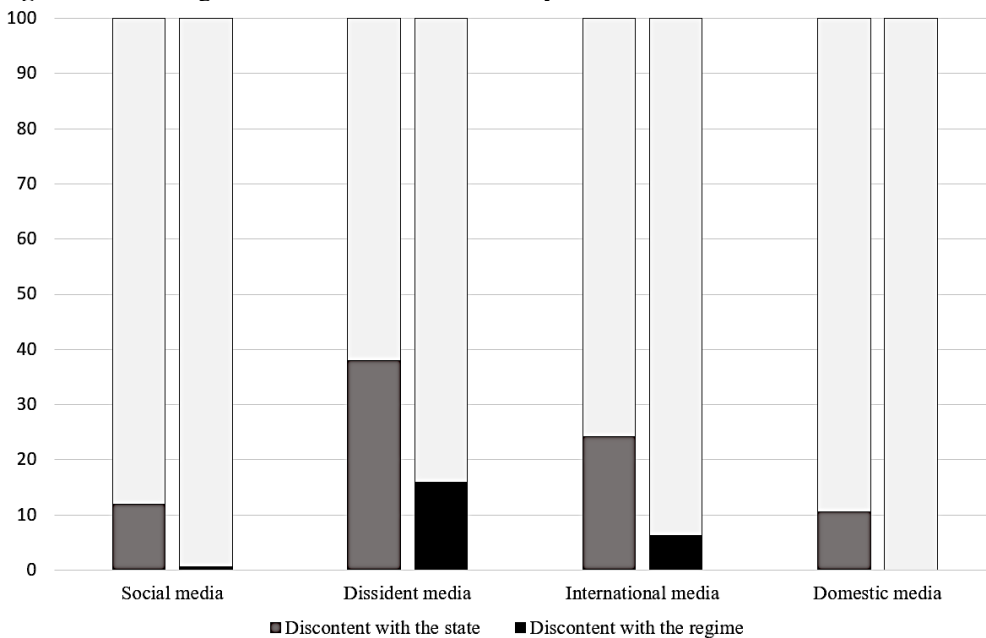


Table 4 displays logistic regressions with selection/nonselection in the different media sources as dependent variables and the political issues plus controls as independent variables. The significant negative odd ratios for both types of political protest underline that these issues are underrepresented in social media. The effect is most pronounced for antiregime events. By contrast, political protests are strongly overrepresented in dissident media—antiregime protests most strongly so. In the other sources, the effects are much smaller and mostly not significant. The coefficient for antiregime protests in the domestic news model is missing because there are no such cases in the sample.

These findings support our hypothesis 3, which expected coverage of political protests to be most likely in dissident media and least likely in the directly censored social and domestic

Table 4. Event Selection and Political Protests

	<i>Social Media</i>	<i>Dissident Media</i>	<i>International Media</i>	<i>Domestic Media</i>
Discontent with the state	0.31 (0.06)**	3.05 (0.55)**	0.91 (0.29)	0.44 (0.27)
Discontent with the regime	0.03 (0.01)**	20.36 (7.04)**	3.81 (2.11)*	-
Large protest	Y	Y	Y	Y
Protester violence	Y	Y	Y	Y
State coercion	Y	Y	Y	Y
Protest series	Y	Y	Y	Y
Days (log)	Y	Y	Y	Y
City	Y	Y	Y	Y
Constant	44.85 (10.50)**	0.02 (0.00)**	0.02 (0.01)**	0.02 (0.01)**
N	2,187	2,187	2,187	2,188
Pseudo R2	0.20	0.27	0.28	0.28

Note: ** p ≤ 0.001, * p ≤ 0.05. Cell entries display logit regression odds ratios and standard errors in brackets. ‘-’ = no variation with the dependent variable.

media. We also find that the differences are much more pronounced for protests with a regime-challenging connotation than for discontent with the state. The substantial coverage of both types in dissident media, and to a lesser extent, international news, can be explained by the media actors in these sources being particularly well plugged into relevant information networks and their editorial inclinations. The complete lack of coverage of protests with regime-challenging claims and the very few events with claims against the state in domestic news can be attributed to censorship. The only explanation would be that domestic journalists gather less information about political protests than their international peers. This is exceedingly unlikely.

To what extent can censorship account for the coverage of political protests in social media? For conflicts with the state, social media cover 319 out of 391 protests (81.59%) of the events. That is less than its average share of protests covered (93.24%). Censorship may be a cause behind this gap. If it is, its effect is somewhat limited. However, among the 59 events featuring regime-challenging claims, social media catches just sixteen (27.12%). This gap is very substantial and almost certainly the result of censorship. Otherwise, it would have to be explained with systematically less information on such events available to social media users compared to other sources or other incentives not to post this information. Both explanations are improbable.

Performance against External Benchmarks

We now seek to assess the performance of data generated from different sources by comparing daily protest and participant frequencies against daily rainfall and public holidays—both of which should suppress the amount of protest. Since case numbers for domestic and international news media are small, we combine all traditional news media (domestic and international). We estimate Poisson regression models on the number of protests per city-day and negative binomial regression models for the total number of estimated protest participants and the average number of protest participants per city-day (both variables are strongly over-dispersed). For both, we drop extreme outliers of above 2,500 participants representing, e.g., fifteen and ten city-days out of all city-days for the total and average participants, respectively, in the social media data. The respective means are 118.17 and 75.93 per city-day (see Appendix A). The models for the daily number of protests control for the number of labor protests. Previous research has shown that labor protests are the most frequent category and display a distinct seasonality (Göbel 2019). All models further control for the city, the weekday, the month, a lagged variable for the dependent variable on the previous day, and a month counter variable to control for trends over time.

Table 5. Daily Protest Frequency Models

	<i>Social Media Protests</i>	<i>Dissident Media Protests</i>	<i>News Media Protests</i>
Public Holidays	-1.05 (0.20)**	-0.78 (0.53)	-1.88 (0.75)*
Rainfall (log)	-0.07 (0.02)*	-0.07 (0.07)	0.06 (0.11)
Number of Workers' Protests	0.21 (0.01)**	2.49 (0.21)**	3.40 (0.32)**
City	Y	Y	Y
Weekday	Y	Y	Y
Month	Y	Y	Y
Month Counter	Y	Y	Y
Lagged Protest	Y	Y	Y
Constant	-0.35* (0.13)	-3.42 (0.49)**	-4.12 (0.81)**
N	1,553	1,553	1,553
Pseudo R2	0.13	0.19	0.24

Note: ** $p < 0.001$, * $p < 0.05$. Cell entries display unstandardized Poisson regression coefficients and standard errors in brackets.

The results displayed in tables 5 and 6 estimate separate models for data recorded in different media-types. Most coefficients for public holidays and rainfall are negative, indicating that all sources pick up the expected depressing effect on contentious action to some extent. However, social media are the only sources where these effects are almost consistently statistically significant and the only source in which rainfall is significant at all. Public holidays become significant for predicting the number of events in the news media sample, but only when workers' protests are controlled for (the model without workers' protest is not displayed).

Our hypothesis 4, which expected that data generated from social media is more closely aligned with public holidays and daily rainfall than data from any other source, is supported. This provides good reasons to assume that due to social media's major advantages during information gathering and reporting and a somewhat limited censorship effect, its event data represents real-world protests more closely than data gathered from any other media type.

Table 6. Models on Total and Average Daily Protest Participants

	<i>Social Media</i>		<i>Dissident Media</i>		<i>News Media</i>	
	Participants	<i>x</i> participants	Participants	<i>x</i> participants	Participants	<i>x</i> participants
Public Holidays	-0.84 (0.39)*	-0.50 (0.37)	-1.18 (1.60)	-2.02 (1.61)	-2.37 (4.00)	-2.28 (3.97)
Rainfall (log)	-0.17 (0.05)*	-0.16 (0.05)*	-0.09 (0.26)	-0.11 (0.27)	0.00 (0.47)	0.00 (0.47)
City (Guangzhou)	Y	Y	Y	Y	Y	Y
Weekday	Y	Y	Y	Y	Y	Y
Month	Y	Y	Y	Y	Y	Y
Month Counter	Y	Y	Y	Y	Y	Y
Lagged Participants	Y	Y	Y	Y	Y	Y
Constant	3.58 (0.32)**	3.18 (0.30)**	1.60 (1.22)	1.37 (1.20)	-0.64 (1.84)	-0.69 (1.84)
N	1,543	1,546	1,546	1,548	1,545	1,545
Pseudo R2	0.00	0.01	0.01	0.01	0.01	0.01

Note: ** $p \leq 0.001$, * $p \leq 0.05$. Cell entries display unstandardized negative binomial regression coefficients and standard errors in brackets. Extreme outlier city-days with ≥ 2500 protest participants have been excluded from the analysis.

CONCLUSION

In this study, we tested a source-specific model of protest event selection in the context of China. Our findings suggest that social media cover considerably more protests than any other source due to distinct advantages in information gathering and reporting. Due to censorship, the coverage of domestic news is dismal. We also find that the overall coverage of international news—particularly English international news—is similarly scant. The coverage of dissident sources is far behind that of social media but better than that of domestic and international news.

We reconfirmed previous research in that selection into the news media is strongly and positively affected by event intensity. New findings emerging from this study are: social media are least affected by intensity, confirming our expectations that users are the least picky about the kinds of events they share. We could not find compelling evidence for an intensity-driven censorship effect on social media. However, we found evidence suggesting that the few protests with regime-challenging claims are strongly affected by censorship in social media. By contrast, discontent with the state is, at most, weakly censored. Social media also performed best among all sources when we evaluated daily events and participant estimates against rainfall and public holidays. In spite of its relative advantages in space, dissident media are strongly affected by intensity. We further confirmed our hypothesis of a strong selection effect for political protests in international and dissident media.

The performance of social media is good news for China scholars and even better news for those who study autocracies where censorship is less comprehensive than in China and international social media are available. Especially Twitter, which is searchable and has recently introduced an API for academic use, is a promising source for protest event data. Although Twitter, Facebook, and Instagram also respond to government censorship requests, they require a certain level of proof that the objected content is contravening local laws and do not enforce every request.¹² However, there are indications that international social media, to an unknown extent, pre-emptively self-censor to avoid conflict with local authorities (Fisher 2018).

Social media data also have three additional limitations. First, it is very event-centered. The more the desired information extends beyond the event itself, the less of its social media data can provide. Second, domestic social media are not the best source for protests featuring regime-challenging content for China. Scholars should consult at least one additional source—dissident media in particular—to make up for this weakness. Third, social media data are difficult to sample. Finding protest-related information in billions of posts is particularly challenging. Scholars have taken the first steps to show how a machine-learning approach may help to overcome classification challenges (Muchlinski et al. 2021; Z. Steinert-Threlkeld, Chan, and Joo Forthcoming; Teo and Fu 2021; Zhang and Pan 2019). If such methods can be further improved, social media may perform even better than the hand-collected social media archive our study relied on.

About the use of traditional media for protest event collection in a censored autocracy, our results are sobering. We show that international and domestic news media catch only a tiny segment of events even though our selection of sources was extensive. The contrast to the good coverage, particularly of domestic news media in democracies, is stark (Nam 2006). News media may perform better in countries with less comprehensive control than China. However, given that we examined large megacities with optimal conditions for news media coverage, it is likely that the coverage of traditional media in authoritarian countries with comparable or slightly weaker information control regimes will be similarly scant. If we take the 2017 Freedom of the Press report as a yardstick (Freedom House 2017), we would expect comparably weak coverage in countries with lower scores than China (implying stricter control), including Cuba, Eritrea, Iran or Uzbekistan, and those in the same decile as China, including Ethiopia, Russia, Saudi Arabia, Sudan or Vietnam. This cautions against including authoritarian regimes in transnational protest event datasets, particularly when the data are based on English media sources only (for a more general critique of cross-national conflict databases, see Herkenrath and Knoll (2011).

Selection bias in protest event data remains a problem that can never be fully resolved. This study will hopefully help highlight specific challenges that scholars in authoritarian contexts have to consider. Most important, perhaps, is that social media data have emerged as a new and promising instrument in the scholarly toolbox of protest event analysis in autocracies and beyond. Future research should explore if and how our findings travel to other nondemocratic regimes.

NOTES

¹ For the year 1990, the Google Scholar search engine lists only 18 publications that match the search terms “protest” and “dataset.” For 2021, the number of publications matching both search criteria is 6,790. Authors’ calculations, January 2022.

² For single-country studies, see e.g., Elfstrom and Kuruvilla (2014), Ferrara (2003), Göbel (2019), Lankina (2015), Li (2017; 2018), McCarthy, Titarenko, McPhail, Rafail, and Augustyn (2008), Robertson (2013) or Tong and Lei (2014). For multi-country and multi-regime research e.g., Barry, Clay, Flynn, and Robinson (2014), Hellmeier, Weidmann and Rød (2018), Johnson and Thyne (2018), Sutton, Butcher and Svensson (2014), Weidmann and Rød (2019) or Woo and Conrad (2019).

³ For regional datasets, see e.g., the Armed Conflict Location & Event Data Project (ACLED) (<https://www.acleddata.com>) or the Social Conflict Analysis Database (SCAD) (<https://www.strausscenter.org/scad.html>). The Mass Mobilization in Autocracies Database (MMAD) (<https://mmadatabase.org>) documents protest events in autocracies. The Integrated Conflict Early Warning System (ICEWS) (<http://dataverse.harvard.edu/dataverse/icews>), the Nonviolent and Violent Campaigns and Outcomes (NVAVC) (https://www.du.edu/korbel/sie/research/chenow_navco_data.html) or the Social, Political and Economic Event Database (SPEED) (<https://clinecenter.illinois.edu/project/human-loop-event-data-projects/SPEED>) collect protest events worldwide.

⁴ See Almeida and Lichbach (2003), Baum and Zhukov (2015), Jenkins and Maher (2016), McCarthy et al. (1996), Myers and Caniglia (2004), Nam (2006), Oliver and Maney (2000), Ortiz et al. (2005), Snyder and Kelly (1977) or Strawn (2008).

⁵ See, e.g., Almeida and Lichbach (2003), Carter and Carter (2020), Elfstrom and Kuruvilla (2014), Kousis, Guigni and Lahusen (2018), Lankina (2015), Robertson (2013) or Steinhardt (2021)), and, more recently, on social media to collect information on protests events (see, e.g. Göbel (2019; 2021), Steinert-Threlkeld (2018), Steinert-Threlkeld, Chan and Joo (Forthcoming), Steinhardt (2021), Teo and Fu (2021), Tsai (2021), Zhang and Pan (2019).

⁶ Almeida and Lichbach’s (2003) and Dowd, Justino, Kishi and Marchais’ (2020) studies are exceptions to this rule.

⁷ Partial exceptions are studies examining the domestic coverage of protest in transitioning Belarus (McCarthy et al. 2008), the domestic coverage of foreign protests in Middle Eastern autocracies (Baum and Zhukov 2015), the coverage of protests in Eastern Germany in non-domestic media (Mueller 1997), and the performance of algorithms to detect protests in social media content from China (Zhang and Pan 2019).

⁸ Out of the sixty-five countries covered in the 2020 Freedom of the Net Report, only two (China and Iran) have blocked Facebook and Twitter consistently. In three more countries (Kazakhstan, Venezuela and Ethiopia) Twitter or Facebook were subject to intermittent blackouts (<https://freedomhouse.org/report/freedom-net/2020/key-internet-controls>, accessed 24 March 2021). Twitter is further blocked in North Korea (<https://www.theguardian.com/world/2016/apr/01/north-korea-announces-blocks-on-facebook-twitter-and-youtube>, accessed 24 March 2021) and Turkmenistan (<https://www.opendemocracy.net/od-russia/naz-nazar/how-turkmenistan-spies-on-its-citizens>, accessed 24 March 2021).

⁹ In 2014, there were 258,000 domestic journalists and 275,000,000 microblog users (referring to Twitter-like applications) in China (All-China Journalists Association 2014; China Internet Network Information Center 2019). There were 636 international journalists (excluding journalists from Hong Kong and Taiwan) registered with the Chinese Ministry of Foreign Affairs in 2015 (2014 data not available) (Ministry of Foreign Affairs 2015). Data about dissident media editors are not available. Given these organizations’ tight resources and the small number of websites, it is safe to assume that they are the smallest group.

¹⁰ Recent studies that use Twitter to identify collective action events seem to confirm this. Steinert-Threlkeld et al. (2022) find many more events for their five cases than any dataset based on local or international news media data, the same applies to cases of electoral violence investigated by Muchlinsky et al. (2021). Note, however, that ACLED, which collects local news sources, outperforms Twitter for protest events in Kenya (Dowd et al. 2020).

¹¹ For rainfall data, see <https://www.nci.noaa.gov/metadata/geoportals/rest/metadata/item/gov.noaa.ncdc:C00861/html>, last accessed February 22, 2021. The data have a larger number of missing values. No data were received on 41% of the days under consideration. To mitigate this problem, we adopted two procedures: first, we checked if the days of missing values are correlated to our daily event and participant estimates. All three dependent variables from all source-samples used below are not significantly correlated to missing days. Second, we imputed missing data with rainfall from the geographically closest available weather station (Hangzhou for Shanghai, Heyuan for Guangzhou and Nanchong for Chongqing). These reduced days with missing data to 30%. We then re-estimated all models with the imputed data. All findings remain stable.

¹² For instance, between mid-2013 and mid-2020, Facebook and Instagram complied with 2872 content restriction requests from Vietnam (<https://transparency.facebook.com/content-restrictions>, accessed 23 March 2021), an unspecified number of which include posts on demonstrations. Vietnam’s party state is in many aspects highly comparable with China’s, except that it does not block these services. In the same period, Twitter reported no such requests from Vietnam (<https://transparency.twitter.com/en/removal-requests.html>, accessed 23 March 2021).

REFERENCES

- All-China Journalists Association. 2014. "Chinese Journalism Development Report (2014)." http://www.chinaja.org.cn/2015-04/07/c_137851976_6.htm. accessed 19 February 2021.
- Almeida, Paul, and Mark Lichbach. 2003. "To the Internet, from the Internet: Comparative Media Coverage of Transnational Protests." *Mobilization: An International Quarterly* 8 (3): 249–72. doi:10.17813/maiq.8.3.90441650652801xl.
- Barranco, José, and Dominique Wisler. 1999. "Validity and Systematicity of Newspaper Data in Event Analysis." *European Sociological Review* 15 (3): 301–22.
- Barry, Colin M, K Chad Clay, Michael E. Flynn, and Gregory Robinson. 2014. "Freedom of Foreign Movement, Economic Opportunities Abroad, and Protest in Non-Democratic Regimes." *Journal of Peace Research* 51(5): 574–88. doi:10.1177/0022343314537860.
- Baum, Matthew A, and Yuri M Zhukov. 2015. "Filtering Revolution: Reporting Bias in International Newspaper Coverage of the Libyan Civil War." *Journal of Peace Research* 52(3): 384–400. doi:10.1177/0022343314554791.
- Cai, Yongshun. 2010. *Collective Resistance in China: Why Popular Protests Succeed or Fail*. Stanford, CA: Stanford University Press.
- Carter, Erin Baggott, and Brett L. Carter. 2020. "Focal Moments and Protests in Autocracies: How Pro-Democracy Anniversaries Shape Dissent in China." *Journal of Conflict Resolution* 64(10): 1796–1827. doi:10.1177/0022002720932995.
- Chen, Chih-Jou Jay, and Yongshun Cai. 2021. "Upward Targeting and Social Protests in China." *Journal of Contemporary China* 30 (130): 511–25 doi:10.1080/10670564.2020.1852735.
- China Internet Network Information Center. 2019. "Statistical Report on Internet Development in China." Beijing. <https://www.cnnic.com.cn/IDR/ReportDownloads/201911/P020191112539794960687.pdf> accessed 19 February 2021.
- China Labour Bulletin. 2017. "Lu Yuyu and Li Tingyu, the Activists Who Put Non News in the News." *China Labour Bulletin* August 18. Accessed 11 February, 2022 (<https://clb.org.hk/content/lu-yuyu-and-li-tingyu-activists-who-put-non-news-news>).
- Dowd, Cairiona, Patricia Justino, Roudabeh Kishi, and Gauthier Marchais. 2020. "Comparing 'New' and 'Old' Media for Violence Monitoring and Crisis Response: Evidence from Kenya." *Research & Politics* 7(3). doi:10.1177/2053168020937592.
- Elfstrom, Manfred, and Sarosh Kuruvilla. 2014. "The Changing Nature of Labor Unrest in China." *Industrial and Labor Relations Review* 67 (2): 453–80.
- Ferrara, Federico. 2003. "Why Regimes Create Disorder: Hobbes's Dilemma During a Rangoon Summer." *Journal of Conflict Resolution* 47(3): 302–25. doi:10.1177/0022002703252366.
- Fisher, Max. 2018. "Inside Facebook's Secret Rulebook for Global Political Speech." *New York Times*, December 27. Accessed February 5, 2022. (<https://www.nytimes.com/2018/12/27/world/facebook-moderators.html>).
- Freedom House. 2017. "Freedom of the Press 2017: Table of Country Scores." Washington, DC: Freedom House. Accessed on March 23, 2021 (https://freedomhouse.org/sites/default/files/FOTP1980-FOTP2017_Public-Data.xlsx).
- Fu, King-wa, Chung-hong Chan, and M. Chau. 2013. "Assessing Censorship on Microblogs in China: Discriminatory Keyword Analysis and the Real-Name Registration Policy." *IEEE Internet Computing* 17(3): 42–50. doi:10.1109/MIC.2013.28.
- Göbel, Christian. 2019. "Social Unrest in China: A Bird's Eye Perspective." Pp 27-45 in *Handbook of Dissent and Protest in China*, edited by Teresa Wright. Cheltenham: Edward Elgar Publishing.
- . 2021. "The Political Logic of Protest Repression in China." *Journal of Contemporary China* 30(128): 169–85. doi:10.1080/10670564.2020.1790897.
- Göbel, Christian, and Lynette H. Ong. 2012. "Social Unrest in China." Long Briefing, Europe China Research and Academic Network (ECRAN). <https://papers.ssrn.com/abstract=2173073>.
- Hassid, Jonathan, and Maria Repnikova. 2016. "Why Chinese Print Journalists Embrace the Internet." *Journalism* 17(7): 882–98. doi:10.1177/1464884915592405.
- Hellmeier, Sebastian, Nils B. Weidmann, and Espen Geelmuyden Rød. 2018. "In the Spotlight: Analyzing Sequential Attention Effects in Protest Reporting." *Political Communication* 35(4): 587–611. doi:10.1080/10584609.2018.1452811

- Herkenrath, Mark, and Alex Knoll. 2011. "Protest Events in International Press Coverage: An Empirical Critique of Cross-National Conflict Databases." *International Journal of Comparative Sociology* 52 (3): 163–80. doi:10.1177/0020715211405417.
- Huang, Haifeng, Serra Boranbay-Akan, and Ling Huang. 2019. "Media, Protest Diffusion, and Authoritarian Resilience." *Political Science Research and Methods* 7(1): 23–42. doi:10.1017/psrm.2016.25.
- Jenkins, J. Craig, and Thomas V. Maher. 2016. "What Should We Do about Source Selection in Event Data? Challenges, Progress, and Possible Solutions." *International Journal of Sociology* 46 (1): 42–57. doi:10.1080/00207659.2016.1130419.
- Johnson, Jaelyn, and Clayton L. Thyne. 2018. "Squeaky Wheels and Troop Loyalty: How Domestic Protests Influence Coups d'État, 1951–2005." *Journal of Conflict Resolution* 62(3): 597–625. doi:10.1177/0022002716654742.
- King, Gary, Jennifer Pan, and Margaret E. Roberts. 2013. "How Censorship in China Allows Government Criticism but Silences Collective Expression." *American Political Science Review* 107(2): 326–43. doi:10.1017/S0003055413000014.
- Kousis, Maria, Marco Giugni, and Christian Lahusen. 2018. "Action Organization Analysis: Extending Protest Event Analysis Using Hubs-Retrieved Websites." *American Behavioral Scientist* 62 (6): 739–57. doi:10.1177/0002764218768846.
- Lankina, Tomila. 2015. "The Dynamics of Regional and National Contentious Politics in Russia: Evidence from a New Dataset." *Problems of Post-Communism* 62(1): 26–44. doi:10.1080/10758216.2015.1002329.
- Li, Yao. 2017. "A Zero-Sum Game? Repression and Protest in China." *Government and Opposition* 54(2): 309–335 1–27. doi:10.1017/gov.2017.24
- . 2018. *Playing by the Informal Rules: Why the Chinese Regime Remains Stable despite Rising Protests*. New York: Cambridge University Press.
- Lorentzen, Peter. 2014. "China's Strategic Censorship." *American Journal of Political Science* 58(2): 402–14. doi:10.1111/ajps.12065.
- McAdam, Doug, John McCarthy, Susan Olzak, and Sarah Soule. 2009. "Documentation. Dynamics of Collective Action." <https://web.stanford.edu/group/collectiveaction/cgi-bin/drupal/>
- McCarthy, John D., Clark McPhail, and Jackie Smith. 1996. "Images of Protest: Dimensions of Selection Bias in Media Coverage of Washington Demonstrations, 1982 and 1991." *American Sociological Review* 61(3): 478–99. doi:10.2307/2096360.
- McCarthy, John D., Larissa Titarenko, Clark McPhail, Patrick Rafail, and Boguslaw Augustyn. 2008. "Assessing Stability in the Patterns of Selection Bias in Newspaper Coverage of Protest During the Transition from Communism in Belarus." *Mobilization: An International Quarterly* 13 (2): 127–46. doi:10.17813/mai.q.13.2.u45461350302663v.
- Ministry of Foreign Affairs. 2015. "Register of Foreign News Organizations in China (Waiguo Zhu Hua Xinwen Jigou Mingce)." Beijing.
- Muchlinski, David, Xiao Yang, Sarah Birch, Craig Macdonald, and Iadh Ounis. 2021. "We Need to Go Deeper: Measuring Electoral Violence Using Convolutional Neural Networks and Social Media." *Political Science Research and Methods* 9(1): 122–39. doi:10.1017/psrm.2020.32.
- Mueller, Carol. 1997. "International Press Coverage of East German Protest Events, 1989." *American Sociological Review* 62(5): 820–32. doi:10.2307/2657362.
- Myers, Daniel J., and Beth Schaefer Caniglia. 2004. "All the Rioting That's Fit to Print: Selection Effects in National Newspaper Coverage of Civil Disorders, 1968–1969." *American Sociological Review* 69 (4): 519–43. doi:10.1177/000312240406900403.
- Nam, Taehyun. 2006. "What You Use Matters: Coding Protest Data." *PS: Political Science & Politics* 39(2): 281–87. doi:10.1017/S104909650606046X.
- Negro, Giacomo, and Susan Olzak. 2019. "Which Side Are You On? The Divergent Effects of Protest Participation on Organizations Affiliated with Identity Groups." *Organization Science* 30(6): 1189–1206. doi:10.1287/orsc.2019.1302.
- Oliver, Pamela E., and Daniel J. Meyer. 1999. "How Events Enter the Public Sphere: Conflict, Location, and Sponsorship in Local Newspaper Coverage of Public Events." *American Journal of Sociology* 105 (1): 38–87. doi:10.1086/210267.
- Oliver, Pamela E., and Gregory M. Maney. 2000. "Political Processes and Local Newspaper Coverage of Protest Events: From Selection Bias to Triadic Interactions." *American Journal of Sociology* 106(2): 463–505. doi:10.1086/ajs.2000.106.issue-2.
- Ortiz, David, Daniel Myers, Eugene Walls, and Maria-Elena Diaz. 2005. "Where Do We Stand with Newspaper Data?" *Mobilization: An International Quarterly* 10 (3): 397–419. doi:10.17813/mai.q.

- 10.3.8360r760k3277t42.
- Qin, Bei, David Strömberg, and Yanhui Wu. 2017. "Why Does China Allow Freer Social Media? Protests versus Surveillance and Propaganda." *Journal of Economic Perspectives* 31(1): 117–40.
- Ritter, Emily Hencken, and Courtenay R. Conrad. 2016. "Preventing and Responding to Dissent: The Observational Challenges of Explaining Strategic Repression." *American Political Science Review* 110 (1): 85–99. doi:10.1017/S0003055415000623.
- Roberts, Margaret E. 2018. *Censored: Distraction and Diversion Inside China's Great Firewall*. Princeton, NJ: Princeton University Press.
- Robertson, Graeme B. 2013. "Protesting Putinism." *Problems of Post-Communism* 60(2): 11–23. doi: 10.2753/PPC1075-8216600202.
- Snyder, David, and William R. Kelly. 1977. "Conflict Intensity, Media Sensitivity and the Validity of Newspaper Data." *American Sociological Review* 42(1): 105–23.
- Steinert-Threlkeld, Zachary C. 2018. *Twitter as Data*. New York: Cambridge University Press.
- Steinert-Threlkeld, Zachary, Alexander Chan, and Jungseock Joo. Forthcoming. "How State and Protester Violence Affect Protest Dynamics." *Journal of Politics*. doi:10.1086/715600.
- Steinhardt, H. Christoph. 2015. "From Blind Spot to Media Spotlight: Propaganda Policy, Media Activism, and the Emergence of Protest Events in the Chinese Public Sphere." *Asian Studies Review* 39 (1): 119–37.
- . 2021. "Defending Stability under Threat: Sensitive Periods and the Repression of Protest in Urban China." *Journal of Contemporary China* 30 (130): 526–49. doi:10.1080/10670564.2020.1852741.
- Strawn, Kelley. 2008. "Validity and Media-Derived Protest Event Data: Examining Relative Coverage Tendencies in Mexican News Media." *Mobilization: An International Quarterly* 13(2): 147–64. doi:10.17813/maiq.13.2.b3j3p1104244u073.
- Sutton, Jonathan, Charles R Butcher, and Isak Svensson. 2014. "Explaining Political Jiu-Jitsu: Institution-Building and the Outcomes of Regime Violence against Unarmed Protests." *Journal of Peace Research* 51(5): 559–73. doi:10.1177/0022343314531004.
- Teo, Elgar, and King-wa Fu. 2021. "A Novel Systematic Approach of Constructing Protests Repertoires from Social Media: Comparing the Roles of Organizational and Non-Organizational Actors in Social Movement." *Journal of Computational Social Science* 4: 787–812. doi:10.1007/s42001-021-00101-3.
- Tilly, Charles. 1993. "Contentious Repertoires in Great Britain, 1758-1834." *Social Science History* 17 (2): 253–80. doi:10.2307/1171282.
- Tong, Yanqi, and Shaohua Lei. 2014. *Social Protest in Contemporary China, 2003-2010: Transitional Pains and Regime Legitimacy*. Abingdon, Oxon: Routledge.
- Tsai, Chia-Yu. 2021. "Protest and Power Structure in China." *Journal of Contemporary China* 30(130): 550–77. doi:10.1080/10670564.2020.1852742.
- Wang, Yaqiu. 2014. "Meet China's Protest Archivist." Foreign Policy. Blog, April 3, 2014. Accessed February 11, 2022. (<https://foreignpolicy.com/2014/04/03/meet-chinas-protest-archivist/>).
- Weidmann, Nils B., and Espen Geelmuyden Rød. 2019. *The Internet and Political Protest in Autocracies*. Oxford: Oxford University Press.
- Woo, Ae sil, and Courtenay R. Conrad. 2019. "The Differential Effects of 'Democratic' Institutions on Dissent in Dictatorships." *Journal of Politics* 81(2): 456-70. doi:10.1086/701496.
- Zhang, Han, and Jennifer Pan. 2019. "CASM: A Deep-Learning Approach for Identifying Collective Action Events with Text and Image Data from Social Media." *Sociological Methodology* 49(1): 1-57. doi:10.1177/0081175019860244.