American Political Science Review (2020) 114, 4, 1343–1351
doi:10.1017/S0003055420000295

Letter
News and Geolocated Social Media Accurately Measure Protest Size Variation

ANTON SOBOLEV *Yale University*
M. KEITH CHEN *University of California, Los Angeles*
JUNGSEOCK JOO *University of California, Los Angeles*
ZACHARY C. STEINERT-THRELKELD *University of California, Los Angeles*

Larger protests are more likely to lead to policy changes than small ones are, but whether or not attendance estimates provided in news or generated from social media are biased is an open question. This letter closes the question: news and geolocated social media data generate accurate estimates of protest size variation. This claim is substantiated using cellphone location data from more than 10 million individuals during the 2017 United States Women's March protests. These cellphone estimates correlate strongly with those provided in news media as well as three size estimates generated using geolocated tweets, one text-based and two based on images. Inferences about protest attendance from these estimates match others' findings about the Women's March.

INTRODUCTION

News and social media data accurately measure protest size variation, according to estimates of protest size obtained via cellphone location data from 10 million individuals.

Protests are a key tool in the repertoire of social movements (Tilly and Tarrow 2015), and observational (Fassiotto and Soule 2017; Uba 2005; Walgrave and Vliegenthart 2012) as well as experimental (Wouters and Walgrave 2017) studies find that those with more participants are more likely to generate policy change. Methods to directly measure size are too costly to be employed across large numbers of protests, limiting their use in academic studies (McPhail and McCarthy 2004; Schweingruber and McPhail 1999; Yip et al. 2010). Instead, event datasets that record size (usually of protests but also of riots or massacres) rely on numbers provided in media reports, usually newspapers (Earl et al. 2004). This reliance raises concerns about the accuracy of the tallies because researchers do not create the measures themselves; those who do—usually organizers or authority figures—have incentives to misrepresent protest size,[1] and estimates are often provided as vague phrases ("hundreds" or "dozens") whose translation into numbers is open to researcher interpretation (Raleigh et al. 2010). Newspapers can also choose between multiple estimates to find the one that suits their editorial beliefs (Barranco and Wisler 1999; Mann 1974). Concern about measurement validity means many event datasets record size as an ordinal variable (Salehyan et al. 2012), average across reports (Weidmann and Rod 2018), or focus on measuring fatalities (Raleigh et al. 2010).

Using individuals' locations recorded via their cellphones ("cellphone data"), we show that scholars can rely on size estimates reported in newspapers or obtained via geolocated tweets. Since people almost always carry their cellphone, those devices record individuals' locations passively, and other researchers have verified that the locations are not biased politically (Chen and Rohla 2018), we consider the cellphone location to be the most accurate estimate currently available.[2] Cellphone data are especially important because, since 1995, the federal government no longer records protest size in Washington DC (McCarthy et al. 1999), rigorously measuring with enumerators is too

Anton Sobolev, Postdoctoral Associate, Leitner Program on Political Economy, Yale University, anton.sobolev@yale.edu.
M. Keith Chen, Professor, Anderson School of Management, University of California, Los Angeles, keith.chen@anderson.ucla.edu.
Jungseock Joo, Assistant Professor, Department of Communication, University of California, Los Angeles, jjoo@comm.ucla.edu.
Zachary C. Steinert-Threlkeld, Assistant Professor, Luskin School of Public Affairs, University of California, Los Angeles, zst@luskin.ucla.edu.

We thank Pablo Barberá, Brett Carter, Benjamin A.T. Graham, Will Hobbs, Nazita Lajevardi, Natalia Lamberova, John McCarthy, Clark McPhail, Jeremy Pressman, Michael Shalev, and Thomas Zeitzoff for their valuable feedback. The editor and anonymous reviewers also provided constructive criticism. Ryan Fox Squire and Aaron Hoffman assisted with the SafeGraph data. Alexander Chan and Christina Indudhara assisted with detecting duplicate faces. This work is supported by NSF RIDIR #1831848, the UCLA Big Data Transdisciplinary Seed Grant, and a UCLA California Center for Population Research Seed Grant. All remaining errors remain inadvertently. Replication files are available at the American Political Science Review Dataverse: https://doi.org/10.7910/DVN/TRLSJA.

Received: June 11, 2019; revised: January 6, 2020; accepted: April 22, 2020.

[1] Wouters and Camp (2017) finds that protest organizers and police agree on size estimates 21.7% of the time.
[2] We are aware of no work challenging the accuracy of cellphone data.

costly when protests occur in multiple locations (Schweingruber and McPhail 1999; Yip et al. 2010), and, even if possible across locations, is likely to feature response bias (Walgrave and Verhulst 2011; Walgrave, Wouters, and Ketelaars 2016). Therefore, the cellphone data forms the baseline against which we compare estimates from news (print and television) and social media (text and images). Since the cellphone data are a sample, protests' true size is still unknown; the analysis presented below therefore compares protest size variation recorded in different datasets, and does not definitely establish protest size.

Estimates using the January 21, 2017, United States Women's March substantiate this argument. One of the largest mobilizations in American history, that day featured 942 protests comprised of approximately 4.77 million individuals (Beyerlein, Ryan, et al. 2018; Chenoweth and Pressman 2017). The cellphone data shows that variations in protest size reported in newspapers as well as estimated via geolocated tweets are accurate. Of the four sets of estimates we compare, those reported in newspapers or calculated via geolocated tweets containing protest key words are the most accurate, followed by counting the number of faces in protest photos shared on Twitter. Counting the number of accounts that share protest photos is the least accurate. The supplementary materials show that the first principal component of the four sets of estimates recovers the most accurate protest size variation, though this fifth estimate is the most difficult to scale and so not one we recommend for future use.

THE CHALLENGE

A long line of research finds that large protests (those with many participants) are more likely than small ones to elicit policy responses across issues and regimes. During the Vietnam War in the United States, Congresspeople were more likely to engage in roll-call votes in response to protests with more than 10,000 participants (McAdam and Su 2002). This responsiveness exists across branches of the government (Walgrave and Vliegenthart 2012) and in non-Western contexts (Lohmann 1994; Uba 2005). Large protests are especially powerful in authoritarian settings because they decrease the legitimacy of the rulers while increasing the cost of repression (Stephan and Chenoweth 2008). Experimental work that manipulates the reported size of a protest also finds that protest size affects policy decisions (Wouters and Walgrave 2017). Amenta et al. (2010) provide a more detailed overview of how social movements affect policy outcomes.

Previous estimates of protest size rely on teams of well-trained observers, surveys of participants, or size estimates reported in newspapers. Using observers who understand the square footage of a protest site, the density of the crowd, and the percentage of the site the crowd occupies is considered the most accurate estimation method; on average, a standing person requires about five square feet of personal space (McPhail and McCarthy 2004). New technology, such as unmanned aerial vehicles (Choi-Fitzpatrick, Juskauskas, and Sabur 2018) and high-resolution satellite imagery, may allow this methodology to scale across simultaneous events. Trained observers can also disperse within a protest and count the size of clusters of participants (Schweingruber and McPhail 1999; Yip et al. 2010). Surveying protesters, during or after an event, can also generate size estimates (Beyerlein, Barwis, et al. 2018; Opp and Gern 1993; Walgrave and Wouters 2014; Walgrave, Wouters, and Ketelaars 2016).

Scholars most often rely on estimates of protest size in newspapers (Earl et al. 2004; Woolley 2000). Though newspapers have known biases in their coverage (Baum and Zhukov 2015; Gerner and Schrodt 1998; Myers and Caniglia 2004), the concern about measurement validity for protest size does not stem from their preference towards violent events in urbanized areas. The concerns are two. First, newspaper estimates of protest size are secondary, usually coming directly from protest organizers or state authorities (Wouters and Camp 2017). The estimates could be too high or too low, but the bias would not be correctable since no independent size estimate exists. Second, newspapers prefer to report on large protests, generating inaccurate records of protest events (McCarthy, McPhail, and Smith 1996; McCarthy et al. 1999). See Section S1 of the Online Appendix for how other datasets report protest size.

DATA

That news and geolocated social media can accurately measure protest size variation is demonstrated using the 2017 United States' Women's March. Occurring on January 21, 2017, the simultaneous protests across the country constitute the largest single-day mobilization in the United States.[3]

This letter compares seven estimates of protest size from three sources: cellphone location data (Chen and Rohla 2018), the Crowd Counting Consortium (Chenoweth and Pressman 2017), and Twitter. The cellphone data are from a data broker and shared with an author of this paper. The Crowd Counting Consortium (CCC) data are publicly available, and the tweets have been collected by another author of this paper (Steinert-Threlkeld 2018).[4] Table 1 provides summary statistics of the three sources, and they are available at Sobolev et al. (2020).[5]

The cellphone estimates use data provided from SafeGraph. Cellphones record owners' location and let downloaded applications use that information to provide, ostensibly, services. The applications and telecommunications providers then sell the location data to brokers or directly to companies (Cox 2019; Valentino-DeVries et al. 2018). The dataset contains location

[3] See Andrews, Caren, and Browne (2018), Fisher, Jasny, and Dow (2018), and Tarrow and Meyer (2018) for more detail on the event.

[4] See Davidson and Berezin (2018) for a case-study combining multiple types of event data.

[5] All tables were made using Hlavac (2018).

TABLE 1. Summary Statistics

Measure	*N*	Mean	Standard Deviation	Min	Max
Cellphone Data	399	238.479	596.154	2	3,082
CCC: News	326	10,526.630	52,362.250	3	725,000
CCC, All Estimates (Best Guess)	614	6,727.480	42,490.840	1	725,000
CCC, All Estimates (Low)	614	5,804.056	34,443.560	1	550,000
Twitter: Text Accounts	1,045	10.793	70.643	2	1,522
Twitter: Images Faces	244	2,601.361	10,568.640	0	132,708
Twitter: Images Accounts	269	926.301	2,209.269	3	20,521

Note: *CCC* refers to the Crowd Counting Consortium data and uses the "Best Guess" variable unless otherwise indicated. *Twitter: Text Accounts* uses keywords to identify protesters. *Twitter: Images Faces* sums the number of faces in protest images. *Twitter: Images Accounts* counts the number of unique accounts that share protest photos.

information for just more than 10 million individuals. We have no identifying information on these individuals (de Montjoye et al. 2018). Section S2 uses actual and imputed vote share by county in Texas to validate the political representativeness of the cellphone data.

Identifying protesters from the SafeGraph data requires several steps. First, we only search in cities identified in the CCC data. Second, in those cities, we only consider pedestrians by excluding pings that were registered while the owners of mobile phones used private or public transportation; mode of transit is identifiable by the frequency of location pings.[6] Third, we deduplicate individuals by averaging their location within a 30-minute window, assigning that location as the individual's location, and keeping only their first appearance.[7] On average, our data has approximately three pings from every cellphone every thirty minutes. Fourth, we take the size distribution of seven-digit geohashes on January 21, 2017, the day of the Women's March, and keep the geohashes above the 99th percentile per city.[8,9] Figure A2 illustrates how geohashes identify locations, and Figures A11–A19 compare these geohashes with actual protest geohashes to demonstrate that this approach recovers locations where protests occur. Fifth, any individual in those geohashes is counted as a protester. The size of this group represents a conservative estimate of the total size of the protest in each city.[10]

To summarize, the number of people in a city's seven-digit geohashes whose density is in the 99th percentile or greater is the number of protesters to which we compare the other estimates.[11]

The CCC is an open-source, hand-coded event dataset established to record reports of protest size. The CCC dataset is agnostic regarding source type. Most are newspapers or local television broadcasts, but a substantial minority (26.84%) of the estimates derive from tweets or Facebook statuses. It also reports high and low estimates, based on these sources. Unless otherwise indicated, we use the Best Guess variable, usually an average of the adjusted low and high estimates in an article, as the protest size estimate. The term *CCC: News* refers to events for which the source is a newspaper or local television broadcast. The term *CCC, All Estimates* includes events for which only social media provides estimates,[12] and *CCC, All Estimates (Low)* uses the Low estimate, not Best Guess, and ignores source type.[13]

[6] One important question is whether protests that bring large numbers of people together at the same time degrade location data accuracy or compromise counts made from it. We have conducted several diagnostic checks related to this concern—for example, looking for evidence that ping frequencies per active user are less stable during large gatherings—and find no such irregularities. We expect this fidelity exists because location data is logged through a cellphone's data connection (as opposed to call services), which is more robust to congestion because these data are asynchronous, events are time-stamped locally (on a user's smartphone), and ping quality is not degraded by data buffering.

[7] We still assume that all counted individuals are protesters, not bystanders. Since some are likely to be vendors, police, or the curious, the size estimates are likely slightly inflated.

[8] A geohash is an alphanumeric code that corresponds to a grid cell. The seven-digit code corresponds to cells with sides of 152 meters. Because geohashes are alphanumeric strings and not points, location matching is much quicker than using spatial coordinates.

[9] We choose the 99th percentile as a threshold for two reasons. First, the density of these geohashes is noticeably distinct from the rest of their city's geohashes; see Figure A10. Second, geohashes with density above the 99th percentile are very close to the protest locations declared by the organizers; see Section S9.2 and Figures A11–A19. Table A21 in Section S10 shows that using a threshold of 95% does not change results.

[10] All the results in the paper are generated without deduplicating across 30-minute geohashes. Deduplicating users across the 30-minute geohashes generates the same results because there is almost no overlap in the protest geohashes with respect to individuals.

[11] We explored an additional step to identifying protesters, based on the path of their walking. In response to concerns from the data provider, we dropped this methodology. Initial results suggested that this extra step degraded results.

[12] We include television broadcasts as part of news because all estimates come from the source type's website, and there is no facile method to distinguish between newspapers and broadcasts. Twitter and Facebook are identifiable with the presence of "Twitter" or "Facebook" in the link or the word "FB" entered as the source.

[13] To investigate the introduction's claim that protests often receive multiple reported size estimates, we model, using CCC, the number of sources per event as a function of the Best Guess variable. Figure A1 shows that there exists no relationship between the size of a protest and the number of sources reporting on it, suggesting that the problem of multiple reported sizes exists for protests regardless of their size.

For Twitter, three approaches approximate the number of protesters.[14] The first counts the number of unique accounts in a city between 10:00 and 17:00 hours local time using one of three common keywords: "womensmarch," "whyimarch," or "imarchfor." Tweets are not resolved to intracity locations for three reasons. First, for the majority of tweets, Twitter aggregates the specific location of a tweet to local polygons while reporting the actual location as the city that contains that polygon. The polygon, however, does not necessarily envelop the location of the actual tweet. For example, Rancho Palos Verdes and Monterrey, two cities near Los Angeles, recorded, based on GPS coordinates, the most number of tweets from January 21, but Twitter labels those tweets as coming from Los Angeles. Intracity location information is therefore unknown for almost all tweets. Second, the Crowd Counting Consortium does not always provide intracity geographic information. A protest it labels as occurring in Seattle would require further research for obtaining the specific location, for example. Third, most cities, even large ones, contain only one protest per day; if they contain multiple events, one protest will contain the vast majority of the protesters (Biggs 2016). We call this measure *Twitter: Text Accounts.*

The second Twitter approach counts the number of faces in any protest photo from a city. A convolutional neural network first identifies protest photos, and then another convolutional neural network identifies faces in that photo. Faces in a protest photo are summed and added per city-day. We call this measure *Twitter: Images Faces.*[15]

The third Twitter approach counts the number of people tweeting a protest photo. This approach does not count retweets, so it assumes that any tweet of a protest photo requires that the Twitter user was actually at a protest. We call this measure *Twitter: Images Accounts.* The Discussion section provides more detail about why these three measures should correlate with the cellphone data estimates. For an explanation of using deep learning and computer vision to extract politically relevant data from images, see Joo and Steinert-Threlkeld (2019) and Steinert-Threlkeld, Chan, and Joo (2020).

For either image approach to be accurate, we assume that Twitter account owners take photographs of protesters, share them on Twitter during the day of the protest, are not strategic about the timing of the sharing (Pfeffer and Mayer 2018), and enable geolocation for the tweet. If these assumptions are true, using Twitter is similar to authorities' estimating procedures (McPhail and McCarthy 2004). While we can think of no reason that accounts would, in aggregate, be biased in sharing protest photos, it is possible these estimates will not be accurate, especially because so many steps are required to enter our data. The results presented in the next section show that, despite the biases that could enter in any of these steps, the Twitter approaches are accurate.

Table 1 shows that size estimates vary substantially across datasets. To compare the estimates, we convert each estimate of size to a per capita measure, log transform that result because it is still very skewed, and take the z-score of the logged per-capita measure (for the remainder of the paper, "z-score" refers to this measure). We prefer per capita to raw estimates because population size will drive the cellphone, CCC, and Twitter estimates. Incorporating it into our measurement therefore removes a possible source of omitted variable bias. The z-score is appropriate because the estimates vary by orders of magnitude across datasets.[16]

Because each source records a different number of protests, Section S8.5 investigates the overlap of the cellphone data and four preferred size estimates in more detail. Manual investigation and Table A20 show no meaningful demographic or socioeconomic difference between the cities in which the measures record protest. Cities that have more people, higher earners, greater income inequality, and more inhabitants with a professional degree are more likely to record protests.

RESULTS

The results indicate that *CCC: News* and *Twitter: Text Accounts* estimate protest size variation with the least error. *Twitter: Images Faces* is next best, while counting the number of accounts that share protest photos is the least accurate. No estimate, however, is inaccurate, and the Discussion section describes when different ones may be preferred.

Unless otherwise stated, all results are for the 155 cities for which each dataset records a protest size. The supplementary materials show results keeping the maximum number of observations each dataset and the cellphone data share do not change.

Figure 1 shows the correlation of the z-score for each of the datasets with the cellphone data. Section S7.1 reports these correlations in a table.[17]

Table 2 shows that *CCC: News* and *Twitter: Text Accounts* produce the closest estimates to the cellphone data. Those two measures are also the closest to the cellphone data the greatest percentage of the time, whether measuring difference (rows 1–5) or the number of observations within 0.1, 0.5, and 1.0 standard deviation of the cellphone data (rows 6–8). *Twitter: Images Accounts* is the least accurate on six of the eight fit measures.

[14] Tweets were collected by connecting to Twitter's streaming application programming interface (API) using the package streamR and requesting only tweets with geographic coordinates (Barbera 2013). For more detail, see Steinert-Threlkeld (2018).

[15] To verify that duplicate faces do not affect the results, we estimated the percentage of duplicate faces per city; Section S3 presents these results, which show that duplicate faces are very rare.

[16] The supplementary materials show that the results hold when not using the per capita or standardized measures.

[17] Figure A20 in Section S11 fits a loess line instead; no nonlinear trends exist in the data.

FIGURE 1. Correlation – Estimates

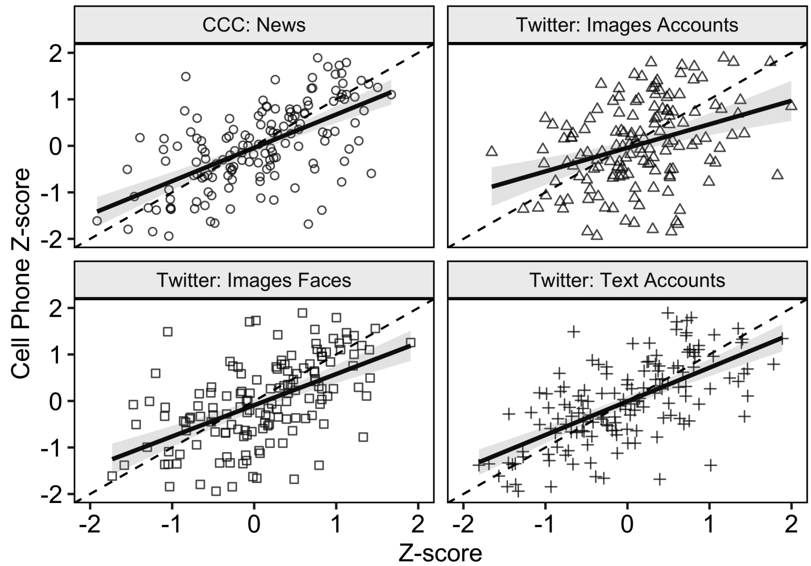


TABLE 2. Measuring Fit

	CCC: News	*Twitter: Images Accounts*	*Twitter: Images Faces*	*Twitter: Text Accounts*
Mean Error	**0.06**	0.10	0.08	0.07
Mean Error (Trimmed)	0.06	0.11	0.14	**0.04**
Mean Absolute Error	**0.58**	0.82	0.71	0.60
Mean Absolute Error (Trimmed)	0.51	0.70	0.59	**0.50**
Closest Estimate %	**0.31**	0.22	0.20	0.27
Percentage Within 0.1 SD	**0.17**	0.08	0.11	0.14
Percentage Within 0.5 SD	0.52	0.38	0.52	**0.54**
Percentage Within 1.0 SD	**0.82**	0.67	0.77	0.81

Note: "(Trimmed)" datasets drop observations where $|z| > 3$. **Bold** is the best measure per row; underline, the worst.

Figures 2 and 3 show the results when each dataset's measure is close to the cellphone data. Figure 2 shows the cumulative distribution of the four datasets' closest estimates across the range of z-scores. This graph reveals any bias in each estimate. For example, if most of a dataset's closest predictions are for small protests, its cumulative distribution function (CDF) will approach 1.0 quickly. If its closest predictions are randomly scattered across the range of protest size, its CDF will track the cellphones' CDF. *Twitter: Text Accounts* has more of its closest estimates occur for small protests, which makes sense: because more tweets contain protest hashtags than protest images, it records small protests better than image estimates. *Twitter: Images Accounts* tracks the cellphone data as protests pass their average size. *Twitter: Images Faces* outperforms *Twitter: Images Accounts* for most of its range except $0 \leq z \leq 1$. CCC is the most consistently accurate of the three, as its CDF closely tracks the cellphone data estimates.

Finally, Figure 3 shows the mean error by z-score decile, with the closest estimate per decile labeled. All measures overestimate the size of small protests and underestimate large ones. *CCC: News* and *Twitter: Text Accounts* perform best, as they each have the smallest mean error three times. *Twitter: Images Faces* is closest twice; *Twitter: Images Accounts*, once.

The supplementary materials present additional analyses confirming these results. To verify that results are not driven by the per capita transformation, we show results measuring the outcome as the z-score of the logarithm of protest size; the subsections presenting them are labeled "Not Per Capita." We also generate protest size estimates using the first principal component of all four measures as well as the two Twitter images measures. Section S4 explains this analysis, and its results are presented in other sections of the Supplementary Materials. The first principal component usually generates the least error.

FIGURE 2. Location of Best Fits per Measure

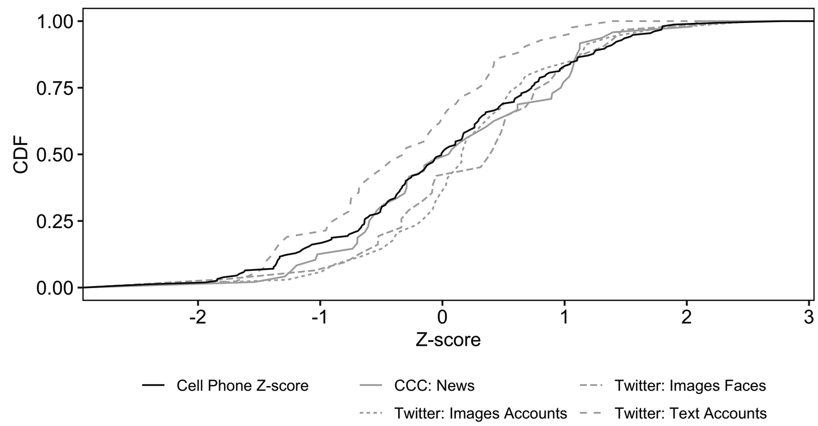
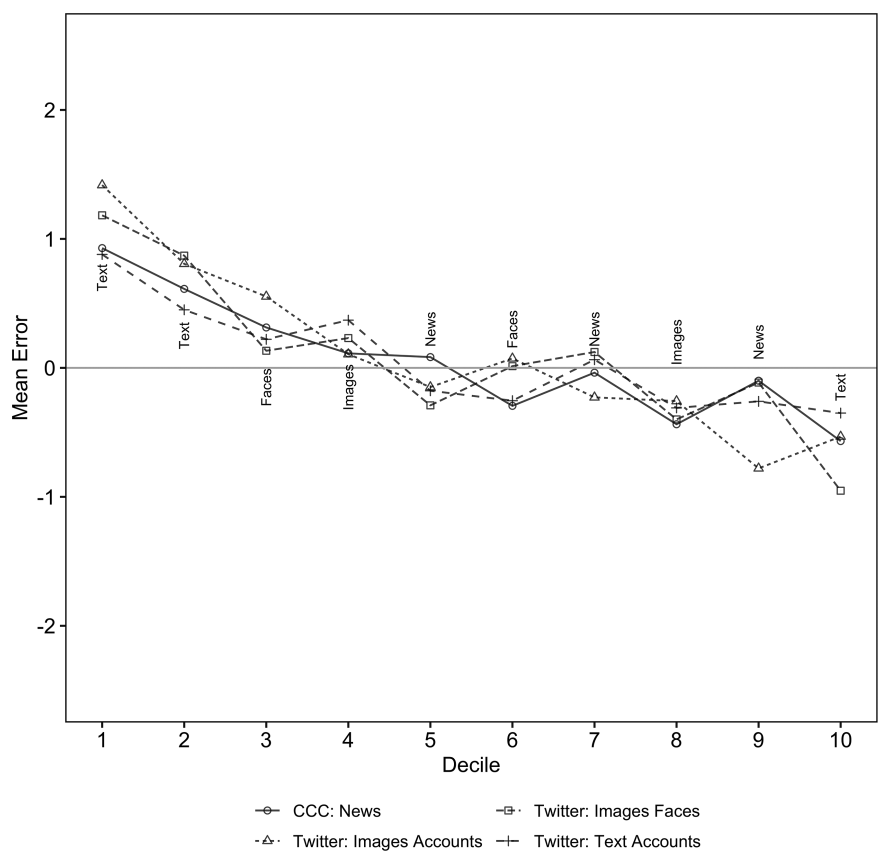


FIGURE 3. Mean Error by Decile



After introducing the principal components analysis, subsequent sections replicate the results above. Section S5 repeats Table 2 using all observations per measure and the "Not Per Capita" measure. Section S6 replicates Figure 3 using all observations per measure and the "Not Per Capita" measure. Both incorporate the principal components results and show results when using the maximum number of observations per dataset.

Sections S7 and S8 provide two sets of additional results. Section S7 shows the Pearson correlation coefficient for each measure, using different subsets of data and operationalizations of protest size. Section S8 introduces a series of models regressing the cellphone data on the four measures, separately and pooled; all models include state fixed effects, socioeconomic controls at the urbanized area level taken from the 2016

American Community Survey, and population. Each table includes a placebo model using only ACS data. *CCC: News* and *Twitter: Text Accounts* feature the highest correlations and best model fit regardless of the subset of data or operationalization of protest size. Wealthier, more Democratic, and more populous areas generated larger protests, matching others' results (McKane and McCammon 2018).

DISCUSSION

Why Does It Work?

News reports rely on others' estimates, and the social media estimates are from models applied to relatively small samples with many entry points for bias. That both sets of estimates work may therefore surprise.

The estimates in *CCC: News* correlate well with the cellphone data for two reasons. First, they tend to come from actors applying methodology similar to that developed in Schweingruber and McPhail (1999). While we can find no reports of how activists generate their estimates, state authorities approximate crowd attendance by estimating the carrying capacity of public areas and the percentage of that area covered with people (McPhail and McCarthy 2004). The estimates in newspapers therefore are not as haphazard as their lack of methodological reporting would suggest. Second, CCC generates its data by accepting submissions of protest reports instead of managing a team of researchers; adopting this open-source approach means it includes a broader range of sources than most event datasets (Herkenrath and Knoll 2011; Nam 2006).

The data generating process for *Twitter: Images Faces* is similar. The protest photos primarily consist of people documenting their immediate surrounding. They contain anywhere from 1 to 48 faces, with an average of 2.81 per photo. So long as people who geotag their photos are randomly distributed across a protest site, then *Twitter: Images Faces* is similar to randomly sampling segments of a protest site (McPhail and McCarthy 2004). See Steinert-Threlkeld, Chan, and Joo (2020) for verification of this approach in Hong Kong, South Korea, Spain, and Venezuela.

Which Dataset Should I Prefer?

We are agnostic about whether or not researchers should prefer news or social media data, as there are advantages and disadvantages to relying on either. Broadly speaking, Twitter requires a greater fixed cost but has lower marginal cost than using newspapers. Newspapers are easier to access than social media, so projects using them can start more quickly. If a project includes local newspapers, they may also provide more geographic coverage than Twitter. One disadvantage is that they are still secondary sources of information, whereas tweets, properly collected, provide primary evidence. A project using newspapers to measure protests also requires more ongoing personnel engagement than one using Twitter.

Though acquiring and processing Twitter data is more technically challenging than doing the same for news articles, it has some advantages over them. Protest images, because they are primary source material, may be closer to the cellphone data in their data generating process than news are. Using the keyword approach to estimate protest size is much less difficult than using images and equally accurate, so estimating protest size variation once a data pipeline is established is relatively easy.

One advantage to using images on Twitter versus tweet text is that text requires subject matter expertise on keywords per protest event, whereas visuals are closer to a universal language (Barry 1997; Graber 1996). Measuring size with text therefore requires more ongoing involvement than image-based estimates. Images, however, are available in fewer circumstances than text (Hawelka et al. 2014).

How Broadly Applicable Are These Results?

Since this paper's validation has only been tested on one event, the scope to which it holds remains to be tested. The results probably hold in other wealthy democracies, though for now that claim remains an assumption. Steinert-Threlkeld, Chan, and Joo (2020), for example, show that *Twitter: Images Faces* and newspaper estimates correlate with each other in four countries about 70% as well as in this paper (see Table A9), though that paper does not have cellphone data with which to verify those estimates. This approach should work in countries where domestic newspapers are not reliable sources, as event datasets rely on foreign newspapers and wire services for size estimates (Clark and Regan 2018; Raleigh et al. 2010; Weidmann and Rod 2018). Twitter penetration correlates strongly with a country's per capita income, so this methodology may also work in wealthy non-democracies (Hawelka et al. 2014). This work cannot test the international robustness of this finding because CCC only focuses on the United States, there is no other dataset of contemporaneous protest size, and we could not obtain cellphone location data in other locations.

Though cellphone location data are probably the best source for capturing variation in protest size, researchers still must rely on media reports or social media because of the difficulty of acquiring them. Orange, the French telecommunications company, made call detail record data available via its Data for Development program in 2012 and 2014 for the Ivory Coast and Senegal, respectively, but those data were not public, and the competition is no longer held. Data brokers such as SafeGraph and Cuebiq work with academics on a case-by-case basis. Otherwise, cellphone data are available based on idiosyncratic partnerships between researchers and private companies.

Concluding Thoughts

These results suggest that news sources provide accurate estimates of protest size variation, as do social media text or images of protesters (Botta, Moat, and Preis 2015). Transforming, and not reporting, raw data throws away information future researchers may find useful. Event datasets that report size should therefore report raw estimates recorded in news. Researchers using those data can then decide whether or not to transform them.

The concern that news and social media estimates of protest size cannot be trusted should be laid to rest. Though others have reported protest size using cellphone records, we are the first, as far as we are aware, to use location data from them as opposed to call detail records collected by transmission towers (Shalev and Rotman 2020; Traag, Quax, and Sloot 2017) and to verify secondary datasets. Whether a researcher prefers news or social media to measure protest size variation is his or her choice, as they are equally accurate. So long as one believes that estimates of protest size variation from cellphone location data are accurate, then estimates of protest size variation from news and social media are trustworthy.

SUPPLEMENTARY MATERIALS

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S0003055420000295.

Replication materials can be found on Dataverse at: https://doi.org/10.7910/DVN/TRLSJA

REFERENCES

Amenta, Edwin, Neal Caren, Elizabeth Chiarello, and Yang Su. 2010. "The Political Consequences of Social Movements." *Annual Review of Sociology* 36: 287–307.

Andrews, Kenneth T., Neal Caren, and Alyssa Browne. 2018. "Protesting Trump." *Mobilization: An International Quarterly* 23 (4): 393–400.

Barbera, Pablo. 2013. "streamR." URL: https://cran.r-project.org/web/packages/streamR/.

Barranco, Jose, and Dominique Wisler. 1999. "Validity Event and Analysis Systematicity of Newspaper Data in Event Analysis." *European Sociological Review* 15 (3): 301–22.

Barry, Ann Marie Seward. 1997. *Visual Intelligence: Perception, Image, and Manipulation in Visual Communication*. SUNY Press.

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.

Beyerlein, Kraig, Peter Barwis, Bryant Crubaugh, and Cole Carnesecca. 2018. "A New Picture of Protest: The National Study of Protest Events." *Sociological Methods & Research* 47 (3): 384–429.

Beyerlein, Kraig, Peter Ryan, Aliyah Abu-Hazeemand, Amity Pauley. 2018. "The 2017 Women's March: A National Study of Solidarity Events." *Mobilization: An International Quarterly* 23 (4): 425–49.

Biggs, Michael. 2016. "Size Matters: Quantifying Protest by Counting Participants." *Sociological Methods & Research* 47 (3): 351–83.

Botta, Federico, Helen Susannah Moat, and Tobias Preis. 2015. "Quantifying Crowd Size with Mobile Phone and Twitter data." *Royal Society Open Science* 2 (5):150162.

Chen, M. Keith, and Ryne Rohla. 2018. "The Effect of Partisanship and Political Advertising on Close Family Ties." *Science* 360 (6392):1020–4.

Chenoweth, Erica, and Jeremey Pressman. 2017. "Crowd Counting Consortium." URL: https://sites.google.com/view/crowdcountingconsortium/home.

Choi-Fitzpatrick, Austin, Tautvydas Juskauskas, and Boby Sabur. 2018. "All the Protestors Fit to Count: Using Geospatial Affordances to Estimate Protest Event Size." *Interface: A Journal for and about Social Movements* 10 (1): 297–321.

Clark, David H., and Patrick M. Regan. 2018. "Mass Mobilization Protest Data." The Mass Mobilization Project. URL: https://www.binghamton.edu/massmobilization/about.html.

Cox, Joseph. 2019. "I Gave a Bounty Hunter $300. Then He Located Our Phone." Vice Media Group. URL: https://motherboard.vice.com/en_us/article/nepxbz/i-gave-a-bounty-hunter-300-dollars-located-phone-microbilt-zumigo-tmobile.

Davidson, Thomas, and Mabel Berezin. 2018. "Britain First and the UK Independence Party: Social Media and Movement-Party Dynamics." *Mobilization: An International Quarterly* 23 (4): 485–510.

de Montjoye, Yves-Alexandre, Sébastien Gambs, Vincent Blondel, Geoffrey Canright, Nicolas De Cordes, Sébastien Deletaille, Kenth Engø-Monsen, *et al.* 2018. "On the Privacy-Conscientious Use of Mobile Phone Data." *Scientific Data* 5: 180286. URL: http://dx.doi.org/10.1038/sdata.2018.286.

Earl, Jennifer, Andrew Martin, John D. McCarthy, and Sarah A. Soule. 2004. "The Use of Newspaper Data in the Study of Collective Action." *Annual Review of Sociology* 30: 65–80.

Fassiotto, Magali, and Sarah A. Soule. 2017. "Loud and Clear: the Effect of Protest Signals on Congressional Attention." *Mobilization: An International Quarterly* 22 (1): 17–38.

Fisher, Dana R., Lorien Jasny, and Dawn M. Dow. 2018. "Why Are We Here? Patterns of Intersectional Motivations across the Resistance." *Mobilization: An International Quarterly* 23 (4): 451–68.

Gerner, Deborah J., and Philip A. Schrodt. 1998. "The Effects of Media Coverage on Crisis Assessment and Early Warning in the Middle East." In *Early Warning and Early Response*, eds. Susanne Schmeidl and Howard Adelman. New York: Columbia University Press.

Graber, Doris A. 1996. "Say It with Pictures." *The ANNALS of the American Academy of Political and Social Science* 546 (1): 85–96.

Hawelka, Bartosz, Izabela Sitko, Euro Beinat, Stanislav Sobolevsky, Pavlos Kazakopoulos, and Carlo Ratti. 2014. "Geo-located Twitter as Proxy for Global Mobility Patterns." *Cartography and Geographic Information Science* 41 (3): 260–71.

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.

Hlavac, Marek (2018). "stargazer: Well-formatted Regression and Summary Statistics Tables. R package version 5.2.2. Available at https://CRAN.R-project.org/package=stargazer.

Joo, Jungseock, and Zachary C. Steinert-Threlkeld. 2019. "Image as Data: Automated Visual Content Analysis for Political Science." Working paper. URL: https://arxiv.org/abs/1810.01544.

Lohmann, Susanne. 1994. "The Dynamics of Informational Cascades: The Monday Demonstrations in Leipzig, East Germany, 1989-91." *World Politics* 47 (1): 42–101.

Mann, Leon. 1974. "Counting the Crowd: Effects of Editorial Policy on Estimates." *Journalism Quarterly* 51 (2): 278–85.

McAdam, Doug, and Yang Su. 2002. "The War at Home: Antiwar Protests and Congressional Voting, 1965 to 1973." *American Sociological Review* 67 (5): 696–721.

McCarthy, John D., Clark McPhail, and Jackie Smith. 1996. "Images of Protest: Dimensions of Selection Bias in Media Coverage of Washington Demonstrations." *American Sociological Review* 61 (3): 478–99.

McCarthy, John D., Clark McPhail, Jackie Smith, and Louis J. Crishock. 1999. "Electronic and Print Media Representations of Washington, D.C. Demonstrations, 1982 and 1991: A Demography of Description Bias." In *Acts of Dissent: New Developments in the Study of Protest*, eds. Dieter Rucht, Ruud Koopmans, and Friedhelm Neidhardt. Lanham, MD: Rowman & Littlefield Publishers, 113–30.

McKane, Rachel G., and Holly J. McCammon. 2018. "Why We March: The Role of Grievances, Threats, and Movement Organizational Resources in the 2017 Women's Marches." *Mobilization: An International Quarterly* 23 (4): 401–24.

McPhail, Clark, and John McCarthy. 2004. "Who Counts and How: Estimating the Size of Protests." *Contexts* 3 (3): 12–8.

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.

Nam, Taehyun. 2006. "What You Use Matters: Coding Protest Data." *PS: Political Science & Politics* 39 (2): 281–7.

Opp, Karl-Dieter, and Christiane Gern. 1993. "Dissident Groups, Personal Networks, and Spontaneous Cooperation: The East German Revolution of 1989." *American Sociological Review* 58 (5): 659–80.

Pfeffer, Jürgen, and Katja Mayer. 2018. "Tampering with Twitter's Sample API." *EPJ Data Science* 7 (50): 1–21.

Raleigh, Clionadh, Andrew Linke, Havard Hegre, and Joakim Karlsen. 2010. "Introducing ACLED: An Armed Conflict Location and Event Dataset: Special Data Feature." *Journal of Peace Research* 47 (5): 651–60.

Salehyan, Idean, Cullen Hendrix, Jesse Hammer, Christina Case, Christopher Linebarger, Emily Stull, and Jennifer Williams. 2012. "Social Conflict in Africa: A New Database." *International Interactions* 38 (4): 503–11.

Schweingruber, David, and Clark McPhail. 1999. "A Method for Systematically Observing and Recording Collective Action." *Sociological Methods & Research* 27 (4): 451–98.

Shalev, Michael, and Assaf Rotman. 2020. "Using Location Data from Mobile Phones to Study Participation in Mass Protests." *Sociological Methods and Research*. Published online 3 April 2020. https://doi.org/10.1177/0049124120914926.

Sobolev, Anton, M. Keith Chen, Jungseock Joo, and Zachary C. Steinert-Threlkeld. 2020. "Replication Data for: News and Geolocated Social Media Accurately Measure Protest Size Variation." Harvard Dataverse, V1, https://doi.org/10.7910/DVN/TRLSJA.

Steinert-Threlkeld, Zachary C. 2018. *Twitter as Data*. Cambridge: Cambridge University Press.

Steinert-Threlkeld, Zachary C., Alexandar Chan, and Jungseock Joo. 2020. "How Violence Affects Protests." Working paper. URL: https://preprints.apsanet.org/engage/apsa/article-details/5e1f9019cc45bb0019b9d894.

Stephan, Maria J., and Erica Chenoweth. 2008. "Why Civil Resistance Works." *International Security* 33 (1): 7–44.

Tarrow, Sidney, and David S. Meyer, eds. 2018. *The Resistance: The Dawn of the Anti-Trump Opposition Movement*. Oxford: Oxford University Press.

Tilly, Charles, and Sidney Tarrow. 2015. *Contentious Politics*. Second edition. Oxford: Oxford University Press.

Traag, Vincent A., Rick Quax, and Peter M. A. Sloot. 2017. "Modelling the Distance Impedance of Protest Attendance." *Physica A: Statistical Mechanics and Its Applications* 468: 171–82.

Uba, Katrin. 2005. "Political Protest and Policy Change: The Direct Impacts of Indian Anti-privatization Mobilizations, 1990-2003." *Mobilization: An International Journal* 10 (3): 1990–2003.

Valentino-DeVries, Jennifer, Natasha Singer, Michael H. Keller, and Aaron Krolik. 2018. "Your Apps Know Where You Were Last Night, and They're Not Keeping It Secret.". URL: https://www.nytimes.com/interactive/2018/12/10/business/location-data-privacy-apps.html?smid=tw-share.

Walgrave, Stefaan, and Joris Verhulst. 2011. "Selection and Response Bias in Protest Surveys." *Mobilization: An International Journal* 16 (2): 203–22.

Walgrave, Stefaan, and Rens Vliegenthart. 2012. "The Complex Agenda-setting Power of Protest: Demonstrations, Media, Parliament, Government, and Legislations in Belgium, 1993-2000." *Mobilization: An International Journal* 17 (2): 129–56.

Walgrave, Stefaan, and Ruud Wouters. 2014. "The Missing Link in the Diffusion of Protest: Asking Others." *American Journal of Sociology* 119 (6): 1670–709.

Walgrave, Stefaan, Ruud Wouters, and Pauline Ketelaars. 2016. "Response Problems in the Protest Survey Design: Evidence from Fifty-one Protest Events in Seven Countries." *Mobilization: An International Journal* 21 (1): 83–104.

Weidmann, Nils B., and Espen Geelmuyden Rød. 2018. "Coding Protest Events in Autocracies." In *The Internet and Political Protest in Autocracies*. Oxford: Oxford University Press.

Woolley, John T. 2000. "Using Media-Based Data in Studies of Politics." *American Journal of Political Science* 44 (1): 156–73.

Wouters, Ruud, and Kirsten Van Camp. 2017. "Less than Expected? How Media Cover Demonstration Turnout." *The International Journal of Press/Politics* 22 (4): 450–70.

Wouters, Ruud, and Stefaan Walgrave. 2017. "Demonstrating Power: How Protest Persuades Political Representatives." *American Sociological Review* 82 (2): 361–83.

Yip, Paul S. F., Ray Watson, K.S. Chan, Eric H. Y. Lau, Feng Chen, Ying Xu, Liqun Xi, Derek Y. T. Cheung, Brian Y. T. Ip, and Danping Liu. 2010. "Estimation of the Number of People in a Demonstration." *Australian & New Zealand Journal of Statistics* 52 (1): 17–26.