No News is Good News?: Mark and Recapture for Event Data When Reporting Probabilities are Less than One.

Cullen S. Hendrix
Korbel School of International Studies
University of Denver
cullen.hendrix@du.edu

Idean Salehyan*
Department of Political Science
University of North Texas
idean@unt.edu

Abstract. We discuss a common, but often ignored, problem in event data: underreporting bias. When collecting data, it is often not the case that source materials capture all events of interest, leading to an undercount of the true number of events. To address this issue, we propose a common method first used to estimate the size of animal populations when a complete census is not feasible: mark and recapture. By taking multiple sources into consideration, one can estimate the rate of missing data across sources and come up with an estimate of the true number of events. To demonstrate the utility of the approach, we compare Associated Press and Agence France Press reports on conflict events, as contained in the Social Conflict in Africa Database. We show that these sources capture approximately 76% of all events in Africa, but that the non-detection rate declines dramatically when considering more significant events. We also show through regression analysis that deadly events, events of a larger magnitude, and events with government repression, among others, are significant predictors of overlapping reporting. Ultimately, the approach can be used to correct for undercounting in event data and to assess the quality of sources used.

* We would like to thank Christian Davenport, James Nichols, and James Hines for comments on and assistance with earlier drafts. Earlier versions were presented at the Peace Science Society and International Studies Association meetings. This material is based upon work supported by, or in part by, the US Army Research Laboratory and the US Army Research Office under contract/grant number W911NF-09-1-0077.

Social science methodologists have spent a considerable amount of time and effort developing new estimators for analyzing data. However, less emphasis is usually placed on how data are collected and compiled from source material. This is especially true for conflict studies, where users have gravitated toward a standard set of variables. Typically, scholars reach for "off the shelf" data on civil or international conflict and pay little attention to issues of data quality and reliability. Even if our statistical approaches are sound, the inferences we make about the world are only as good as the data we use.

The last few years have witnessed tremendous growth in the number of data resources on conflict events, actors, and processes (see a recent special issue of *International Interactions*, with a useful overview by Schrodt (2012)). In addition to long-standing sources, such as the Correlates of War, the Uppsala Conflict Data Program, and Minorities at Risk, new data collection projects on civil conflict, international conflict, ethnic violence, terrorism, protest, human rights, peacekeeping, and so on, have emerged at a rapid pace. Some of these resources use human researchers to sort through a large corpus of qualitative information, and often make complex judgments about the nature of actors or events. Other data projects rely on computer-generated information through natural language processing, text mining, and related techniques (King and Lowe 2003). Both methods have virtues and present their own challenges and limitations.

In general terms, all conflict data projects have a similar data generation pipeline, and as such, face a common set of challenges. An idealized account of this process can be summarized as follows (see Figure 1). An event occurs, and is ultimately the "thing" we would like to study. It is detected and reported by primary and secondary sources (e.g. news articles, NGO reports, books). The researcher then compiles reports of said event or unit, at which point information on actors, targets, dates, etc., is extracted from the corpus of text and turned into quantitative indicators (i.e. coding). At this point, the data may or may not be validated through procedures such as inter-coder reliability tests and comparisons against similar datasets. Finally, the data are compiled and ready for analysis/dissemination.

--FIGURE 1 HERE--

We focus on the first link in the chain: whether an event is reported upon by existing sources, a problem to which conflict researchers have paid scant attention (but see: Davenport and Ball 2002; Drakos and Gofas 2006; Hug 2003). Our focus here is on the news media, which has been frequently relied upon for coding conflict data. Though all the subsequent steps in the data generation pipeline have important implications for the inferences we draw, non-detection is perhaps the most vexing problem. Zeroes (or non-events) can be the result of two data-generating processes: 1) nothing happened, or 2) something happened but it was not reported, either because there was no reporter present or because reporters chose not to cover it. Event data are often aggregated into counts, which are then treated as the key outcomes of interest: battle deaths, the number of protests and riots, the number of terrorist attacks, the number of militarized interstate disputes, etc. To the extent that non-detection is present and non-random, inferences based on event count analyses are biased, and the nature of this bias is not obvious unless we have accurate information on journalistic processes.[1] In some arenas, false negatives are likely to be more prevalent than others. Full scale wars with large numbers of battle deaths are unlikely to escape the attention of primary sources.

The issue of non-detection is more challenging with respect to phenomena other than large-scale armed conflict, such as protests, riots, and episodic violence such as coups or communal conflict (Buhaug and Urdal 2013; Powell and Thyne 2011; Salehyan et al. 2012). Such smaller-scale and less deadly events may simply slip through the cracks. Moreover, disaggregating large-scale civil and international wars into battle events (Raleigh et al. 2010; Sundberg and Melander 2013) also faces the non-detection challenge, as it is unlikely that every battle is reported. Reporting effort is non-constant across time and space, meaning that many events are likely to be missed. Therefore it is difficult to know whether event counts represent the actual number of event occurrences.

In this paper, we investigate event detection/non-detection in the Social Conflict in Africa Database (SCAD), analyzing data on 1,443 events across Africa in 2012 (Salehyan et al. 2012). Borrowing from practices in field ecology, we use a mark and recapture approach (Lincoln 1930;

[1] Some sources, such as NGO or government reports, may have a strategic incentive to under- or over-count information. For example, activists may have an incentive to overstate the number of protesters and governments may have an incentive to claim that the opposition is using violent tactics. Such strategic misrepresentation is another potential source of bias, but not one we address here as we focus on journalistic accounts.

Nichols 1992; Williams et al. 2002) to estimate the “true” number of social conflict events by leveraging independent sources of news reports: the Associated Press (AP) and Agence France Presse (AFP). Our preliminary analysis suggests these two sources detected approximately 76 percent of the social conflict events that occurred, but that detection rates were significantly higher for events which results in large numbers of deaths (>=10), involved more than 1,000 participants, and in which the government responded with repression. Subsequently, we present an analysis of the covariates of event detection for each of the news sources. We find several significant covariates of double detection and some systematic differences between the two news agencies. Our approach, we believe, offers a useful method for detecting rates of missingness in event data due to reporting effort, and a generalizable diagnostic for the whether the sources used by a data project capture a large proportion of world events. It is an approach that is simple and flexible enough to be applied to numerous other data projects.

The remainder of the paper is as follows. The next section introduces the basic logic of the mark and recapture approach and presents a way of generalizing the method to estimating the “true” number of events (i.e., the estimated number of false negatives) in an event dataset based on more than one reporting source. The following section presents the results of the analysis of the SCAD events for 2012 and then investigates the covariates of coverage by particular media sources. The final section offers our conclusions.

II. Mark and recapture

Mark and recapture (also known as capture-recapture and multiple recapture) is a method used in ecology to estimate population sizes when it is unfeasible to count the entire population of interest. Mark and recapture begins with the capture and tagging a sample of the animal one wishes to count, which are then released back into the population. Subsequently, a second sample of the animal is captured. By assumption, the proportion of marked animals in the second sample should be proportional to the number of marked individuals in the whole population (Lincoln 1930), providing an estimate of the likelihood that any given animal will be captured. This probability is then used to estimate the size of the entire population, within a confidence interval. The technique, developed first for estimating populations of fish and migratory waterfowl, is now widely used in estimating animal population dynamics in situations where it would be wholly impractical – due to costs or logistics – to count each individual animal.

Epidemiologists have also applied this method to estimate the prevalence of disease, when it is impossible to count the true number of infections (Chao et al 2001).

The basic estimator is rather simple and intuitive, and is represented as $N = (M*C)/R$, where N is the estimated population size, M is the number of animals (or people, or events) captured and "marked" on the first attempt, C is the total number of animals captured on the second attempt, and R is the number of animals captured on the first attempt that were subsequently recaptured. While this is the most basic form and requires several critical assumptions, there are numerous extensions to the basic estimator, which account for factors such as seasonal migration; differences in capture probability due to factors such as sex and age; and death or loss in the population. In order to convey the basic logic of mark-recapture techniques, which are foreign to most social scientists, we do not delve into these extensions here. There exist entire textbooks and software packages available for analyzing more complex models for those so inclined (see Williams, Nichols, and Conroy 2002).

Typically, scholars have ignored the possibility that event data are incomplete (for exceptions on media bias, see Earl et al. 2004; and Davenport 2010). When using observed counts as if they represent the true number of events, results can potentially be misleading. This is especially true if underreporting is systematic rather than random, and if a key variable of interest is correlated with both reporting probabilities and the dependent variable. Li (2005) and Drakos and Gofas (2006) discuss this problem with respect to terrorism. The level of democracy could lead a country to be more susceptible to terrorist attacks for theoretically interesting reasons; for example, democratic governments may be more sensitive to the death of their citizens and thus a more attractive target. However, it is also likely that democracies have a free press, leading to a higher probability that a given event is reported. All else being equal, we would expect a relatively small terrorist attack in, say, the United States to be more likely to be captured by the news media than a similar attack in Cuba, which is more restrictive with information. Therefore, results may be driven in whole or in part by different reporting environments and non-random undercounting.

Mark and recapture approaches to underreporting hold great promise for event count data, or other types of data for which it is reasonable to assume that the detection probability is less than one. In the social sciences, some have applied the method to estimating civilian deaths due to conflict and human rights abuse (Lum et al. 2013; Manrique-Valler et al. 2012). Human rights groups, particularly the Human Rights Data Analysis Group, have been using an extension of mark and recapture, multiple systems estimation, in their advocacy work.[2] Multiple systems estimation uses the degree of overlap in often incomplete reporting of human rights violations – such as lists of civilian deaths during civil wars – to derive estimates of the total numbers of violations. Both techniques can be used with stratified samples in order to address differences in reporting across different regions, time periods, etc. (International Working Group for Disease Monitoring and Forecasting 1995).

For our purposes we examine conflict data derived from newswire agencies. Conflict event data often rely upon news sources such as the Associated Press, the Agence France Presse, and Reuters, among others. These sources can be thought of as independent surveys of the population of interest. Although it is not possible to "mark" and "recapture" a protest event or riot, one can observe how many independent events were reported by sources 1, 2, 3,…, and how many were reported by multiple agencies. This "double reporting" is akin to multiple captures of the same event. Of course, there is a critical assumption to using this method. One must assume that news agencies are looking for the same thing and have similar standards for "newsworthiness". That is, if news sources 1 and 2 both witness an event, they both will report upon in. Non-detection implies just that: a reporter did not witness an event or missed the story, rather than not reporting because they found it to be unimportant. For some types of events, as we discuss below, this is not a plausible assumption. News agencies are likely to devote less effort to some parts of the world, operate in different commercial environments, and may have different standards for what they deem interesting.

III. Mark and recapture analysis of the SCAD dataset

SCAD is an event data source for researchers studying social conflict, a broad category that encompasses several forms of contentious collective action, including protests, riots, strikes, and armed attacks that do not fit the convention definition of armed conflict. The data are human coded from keyword searches of Associated Press (AP) and Agence France Presse (AFP) news

[2] See Amelia Hoover Green, "Multiple Systems Estimation: the Basics," *Human Rights Data Analysis Group*, March 11, 2013. Accessible online at: https://hrdag.org/mse-the-basics/ . (Access date, August 21, 2014).

wires, and cover all countries on the African continent with a population greater than 1 million. Each record in SCAD refers to a unique event, though a single event can occur in multiple locations. Events can last for a single day to several weeks or months. To define an event, our research team determined the particular actor(s) involved, their target(s) and the issue(s) at stake. A conflict is classified as a single event if the issues, actors, and targets are the same and there is a distinct, continuous series of actions over time; for a full description of the data, see Salehyan et al. (2012).

Although the data span the period, 1990-2012, we only began collecting systematic information on the reporting news agency for the 2012 data. Events are tagged as being covered by the AP, the AFP, or both. For the 1,443 events in 2012, the basic descriptive statistics indicate AFP is the more authoritative source for Africa. Table 1 indicates that of the total number of events, the AFP was the sole reporting agency for 45.88%, the AP was the sole reporting agency for 22.52% of events, while both sources covered 31.6% of the events. We consider this last category as being “recaptured,” or more appropriately, “double reported”.

--TABLE 1 HERE--

Using these simple descriptive statistics, we can calculate Lincoln-Petersen estimates of the “true” number of events, based on the mark and recapture rate. AFP detected 1,118 events (662 detected by AFP + 456 detected by both) over the period. Of these, the AP detected 456 (those detected by both), resulting in a detection probability for the AP of 456/1,118 = 0.41. AP detected 781 events, of which AFP detected 456, resulting in a detection probability for the AFP of 456/781 = 0.58. Thus, the AFP is somewhat less likely to miss an event that was captured by the AP. The combined non-detection rate is simply the product of the two non-detection rates $(1 - p_{AP})*(1 - p_{AFP})$, or 0.24. This method suggests the two news sources detected 76 percent of all social conflict events.

What does this suggest about the “true” number of events? Per the Lincoln-Peterson method, total population size is: $N^* = S/R$, where N^* is the estimate of total population size; S is the total sample size (i.e., all detected individuals), and R is the combined detection rate. This suggests a total population of events, Africa-wide, of 1,443/.76 = 1,899, with a 95% confidence interval of [1855, 1943].[3] Clearly, this is higher than the 1,443 events that were observed. Whether this is good or bad is hard to know. To our knowledge, other event data sources do not report similar statistics, so the degree of underreporting in their datasets is unknown.

One restrictive assumption is that both agencies have the same standards for what gets reported. It is problematic if the AP is looking for toads while the AFP is looking for frogs. For instance, given that the AP did not report on 59% of the stories that were included in AFP reports, this should only be because AP did not observe the event and not because they deemed the story to be too insignificant to write about (or possibly let the AFP take the lead). Given that U.S. news outlets largely consume AP newswires, we find this to be an overly restrictive assumption. As an aside, we note that preliminary data on Latin America suggest that the AP is the better news source for the Western Hemisphere, ostensibly because the U.S. public is more interested in that region than in Africa. For our current example, many of the incidents in the data are very small protests, which attract few people and cause little damage to people or property. It is likely the case that for some of the events in the data, one or the other news agency *skipped* reporting rather than *missed* the event.

However, we can stratify the sample into more or less significant events, and estimate detection rates for those that are more socially or politically important. Combined detection rates for some subsets of the data are significantly higher. The adage "If it bleeds, it leads" receives strong support. Events for which any deaths were reported have a combined detection rate of 0.86; if ten or more deaths were reported, the detection rate increases to 0.98. In contrast, for events where no deaths were reported, the detection rate falls to 0.69. Combined detection rates were higher for events with more participants. For events with greater than 100/1,000/10,000 participants, the combined detection rate is 0.85/0.92/0.98 (see Figure 2). The lowest detection rates were for events on which reporting was scant and characterized by missing information. For events for which the number of participants is unknown, the detection rate is 0.70, while for events for which the location is unknown (<5% of all events), the detection rate is only 0.16. Thus, the sources used in SCAD are quite good at capturing events that are deadly and/or attract

[3] Variance is calculated as follows:

$$\mathrm{var}(N) = [(n_{ap}+1)(n_{afp}+1)(n_{ap}-n_{ap,afp})(n_{afp}-n_{ap,afp})]/[(n_{ap,afp}+1)(n_{ap,afp}+1)(n_{ap,afp}+2)].$$

a large number of people. If researchers are concerned about event non-detection, they may use this information to restrict their sample to more significant events.

--FIGURE 2 HERE--

Which news agencies cover what?

The basic analysis above indicates that AFP is a more comprehensive news source for the African continent, but are there specific traits of events or countries that make their detection by one source – or both sources – more or less likely? To pursue this question, we conducted a bivariate probit regression analysis with two equations for detection by the AP and detection by the AFP. Bivariate probit regression is similar to two distinct regression models, but accounts for correlated outcomes and calculates a rho parameter, which is the correlation coefficient between the residuals of the two probit models, which tests whether the residuals are correlated across models and whether simultaneous estimation is preferable to running independent probits. We run a separate logit model with a binary dependent variable for whether or not an event was jointly detected by the sources.

The independent variables are drawn from SCAD event attributes and country-level attributes from a variety of sources. The event-level variables correspond to types of contention, target choice, government responses, magnitude (in terms of participants and numbers killed), and the nature of actor demands. *Protest or riot* is an indicator for whether the event was a protest or riot, the most public forms of contentious challenge.[4] *Violent attacks* are those events – ranging from government-initiated attacks to anti-government violence, communal violence, and

[4] Corresponding to SCAD etypes 1-4. "1. Organized demonstration: demonstrations planned and conducted by a formal group or organization. 2. Spontaneous demonstration: mass demonstrations without formal organization.
3. Organized riot: rioting planned by a formal group or organization. Participants intend to cause injury to people or property. 4. Spontaneous riot: rioting without formal organization" (Salehyan et al. 2012, 506).

intrastate violence, which are primarily armed in nature.[5] We also include indicators for whether the event occurred in an *urban area* (population greater than 100,000) as reporters may congregate in major cities. We include information on whether the central or regional government was a target (*central gov. target* and *regional gov. target*), and the *duration* of the event (in days). We also include *government repression*, which takes on a value of 1 if non-lethal repression used (e.g. tear gas, arrests, etc.) or 2 if lethal repression was used. We include indicators for whether the events were large in magnitude, having *>1,000 participants*, and if the event resulted in *>=10 deaths*. We include two indicators for whether the aims of the dissidents/actors related to *political* or *ethnic/religious issues* as these might imply greater salience.[6]

The country-level variables are relatively standard. *Democracy* is an indicator that takes on a value of 1 if the Polity2 score for the country is greater than or equal to 6, per Marshall, Jaggers and Gurr (2013). We include the Reporters Without Borders *Press Freedom Index*, which measures the degree to which the press is hindered in the collection and dissemination of information, taking into account the legal environment for the media and its independence from government (higher numbers indicate worse conditions).[7] These two measures may influence the quality and availability of information about dissent. We include *(ln) population*, on the supposition that larger countries should receive more media attention and that events therein

[5] Corresponding to SCAD etypes 7-10. "7. Pro-government violence: repression against individuals or groups, initiated by government actors or violence by private actors in support of the government. 8. Anti-government violence: violence against government targets by permanent or semi-permanent militias. These events do not meet the inclusion criteria for the Uppsala University Armed Conflicts Dataset. 9. Extra-government violence: violence involving at least one permanent or semi- permanent militia group. Government forces are not included as an actor or a target (e.g. communal conflict). As opposed to rioting, extra-government violence involves an organized militia or criminal group. 10. Intra-government violence: fighting between two armed factions within the government (e.g. mutinies or coups)" (Salehyan et al. 2012, 506-507).

[6] Per the issue1, issue2 and issue3 variables in SCAD.

[7] The actual questionnaire on which the codings are based is available at http://www.rsf.org/index/qEN.html.

would be more likely to be detected (World Bank 2013). Finally, we include three variables that capture regional and linguistic variation. *North Africa* takes on a value of 1 for the countries Algeria, Egypt, Libya, Morocco, and Tunisia; given attention to Arab Spring, these countries may experience more media exposure. *Anglophone* refers to English-speaking countries, where we presume AP reporting would be more focused (AP is US-based firm), and *Francophone* refers to French-speaking countries, where presumably AFP would focus more attention.

--TABLE 2 HERE—

Results are reported in Table 2. Confirming the earlier results of the mark and recapture analysis, large numbers of fatalities and participants were the strongest predictors of dual detection (Model 2), as was the use of repression. For the AP (Model 1A), events were more likely to be reported if they were deadly, had a large number of participants, and repression was used by the state. Democracy and population size were also positively associated with reporting, and reporting was less common in francophone countries. For the AFP (Model 1B), deadly events and use of repression are positively associated with reporting, although protest size just misses the threshold for significance (p=0.08). Interestingly, political and ethnic issues are more likely to be reported by the AFP, but not the AP. Reporting was more common in French-speaking countries, as we would expect, be we were surprised to find that the AFP was less likely to report in democratic countries. The rho parameter was significant and negative, indicating that accounting for the specified covariates, there was still substantial non-correspondence between the two reporting agencies.

Looking at substantive effects, an event that was met with repression was approximately 43% more likely to be double-reported; lethal repression increased the probability of double reporting by 86% relative to no repression (estimates from Model 2). An event with more than 1,000 participants was almost twice (96%) more likely to be double-reported, and an event with 10 or more deaths was over twice (148%) more likely to be double-reported. These findings confirm the conjecture that “if it bleeds it leads.” Double reporting was less than half as likely (-61%) likely in French-speaking countries, as the AP is less likely to devote effort to such countries. These event and country attributes are akin to animal and survey site characteristics in field ecology, where mark and recapture methods were first developed. Although we do not discuss such extensions here, it is possible to include such covariates into the basic model when accounting for detection rates.[8]

IV. Conclusion

Conflict researchers rarely consider the accuracy and reliability of the sources they consult. Quantitative analyses tacitly assume that the data represent the universe of cases, but with conflict data this assumption is most likely false. There is likely to be systematic undercounting in events data, due to the fact that reporting agencies cannot observe each and every incident of interest. This is especially true in low-information environments as typically found in conflict zones. Ignoring this undercounting can lead to systematic bias in the data, which poses empirical and theoretical conundrum for researchers. Mark and recapture methods can help to identify the magnitude and sources of bias or underreporting in the data.

Armed with this knowledge what should be done? One approach is to replace the observed event count with an estimated event count, bootstrapping along the confidence interval. However, this is not feasible for SCAD. The estimates above are for all countries in Africa in 2012, producing a relatively large sample size. Typically, regressions are run on a country/year basis with the count of events as the dependent variable; for many countries included, the number of events per year is very small (as low as 1-2 events). This small per-country sample size makes it difficult to reliably estimate population counts for some country-years.

Another approach is to use the mark-recapture method as a diagnostic tool when selecting sources and collecting data in the first place. For SCAD this enabled us to think much more critically about the sources used and the process by which reports are generated. Through this procedure we were able to judge the rate of missingness from AFP and AP reports, which is a useful tool for assessing the quality of a source. Other data projects can use similar methods when trying to arbitrate among the many news outlets that are readily available online. If one source is nested within another (i.e. one source captures a subset of another's stories, with very little independent information), then it can be safely discarded. In addition, for human coding, there is a trade off between the number of sources used and the labor it takes to sift through another source. Using this method, one can gauge the marginal benefit of adding another source,

[8] See Program Mark for software and extensions: http://warnercnr.colostate.edu/~gwhite/mark/mark.htm (accessed August 21, 2014)

given time constraints. In other words, is the time and effort expended in using an additional source worth the cost given the increase in detection probabilities?

We have also shown that by stratifying the data into more and less newsworthy categories, we significantly alter detection/reporting probabilities. For events with a large number of participants and those that cause fatalities, we are confident that SCAD approaches the universe of such events. By dissecting the data in this way, we provide a diagnostic method to ascertain which slices of the data are of higher quality than others, allowing researchers to make their own judgments about how to use this information.

Ultimately, using mark-recapture methods requires a different way of thinking about the corpus of text that is used when coding data. For many data projects, sources are selected in an ad hoc manner depending what is readily available. Often this means inconsistency in the sources used, with some countries and time periods having rich information, while others receive scant attention. This approach can severely confound empirical analyses if reporting varies across cases. This is especially true with the advent of the Internet and digital archives. Throughout the 1990s and 2000s, more and more news sources began to appear in online repositories such as Lexis-Nexis and Factiva. In collecting data on say, protest, simply incorporating new sources as they become available can lead to what falsely appears to be an upward trend in the aggregate data and underreporting on countries that are less digitally connected.

The method we propose for collecting data is quite flexible and intuitive. Mark and recapture methods simply require researchers to be consistent with the sources they consult across cases and note which source reported on which event(s). We do not pretend that multiple recapture methods will be useful for all data projects. However, when feasible, researchers should collect their data in a manner that approximates multiple independent attempts to sample a population, and indicate which sources where used for particular data points. This allows others to interrogate the data, assess its quality and reliability, and potential sources of bias.

References

Buhaug, Halvard, & Urdal, Henrik. 2013. An urbanization bomb? Population growth and social disorder in cities. *Global Environmental Change*, 23 (1):1-10.

Chao, Anne, P.K. Tsay, Sheng-Hsiang Lin, Wen-Yi Shau, and Day-Yu Chao. 2001. The Applications of Capture-Recapture Models to Epidemiological Data. *Statistics in Medicine.* 20(2): 3123-3157.

Davenport, Christian. 2010. *Media Bias, Perspective, and State Repression: The Black Panther Party.* New York: Cambridge University Press.

Davenport, Christian and Patrick Ball. 2002. Views to a Kill: Exploring the implications of Source Selection in the Case of Guatemalan State Terror, 1977-1995. *Journal of Conflict Resolution*, 46 (3): 427-450.

Drakos, Konstantinos and Andreas Gofas. 2006. The Devil You Know But Are Afraid to Face: Underreporting Bias and its Distorting Effects on the Study of Terrorism. *Journal of Conflict Research.* 50(5): 714-735.

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

Hendrix, Cullen and Idean Salehyan. 2012. Climate Change, Rainfall, and Social Conflict in Africa. *Journal of Peace Research*, 49 (1):35-50.

Hoover Green, Amelia. 2013. Multiple Systems Estimation: the Basics. *Human Rights Data Analysis Group*, March 11, 2013. https://hrdag.org/mse-the-basics/ . (Access date, August 21, 2014).

Hug, Simon. 2003. Selection Bias in Comparative Research: The Case of Incomplete Datasets. *Political Analysis.* 11(3): 255-274.

International Working Group for Disease Monitoring and Forecasting. 1995. Capture-Recapture and Multipe-Record Systems Estimation I: History and Theoretical Development. *American Journal of Epidemiology* 142 (10): 1047-1058.

King, Gary and Will Lowe. 2003. An Automated Information Extraction Tool for International Conflict Data with Performance as Good as Human Coders: A Rare Events Evaluation Design. *International Organization.* 57(3): 617-642.

Li, Quan. 2005. Does Democracy Promote or Reduce Transnational Terrorist Incidents? *Journal of Conflict Resolution.* 49(2): 278-297.

Lincoln, F. C. 1930. Calculating Waterfowl Abundance on the Basis of Banding Returns. *United States Department of Agriculture Circular*, 118, 1–4.

Lum, Kristian, Megan Emily Price and David Banks. 2013. Applications of Multiple Systems Estimation in Human Rights Research. *American Statistician.* 67(4):191-200

Manrique-Valler, Daniel, Megan Price, and Anita Ghodes. 2013. Multiple Systems Estimation Techniques for Estimating Casualties in Armed Conflicts. In, *Counting Civilian Casulaties: An Introduction to Recording and Estimating Nonmilitary Deaths in Conflict.* Taylor Seybolt, Jay Aronson, and Baruch Fischoff, eds. Oxford: Oxford University Press

Marshall, Monty, Keith Jaggers, & Tedd Gurr. 201). Polity IV Project: Political Regime Characteristics and Transitions, 1800-2012. Center for International Development and Conflict Management. University of Maryland-College Park.

Nichols, James. 1992. Capture-Recapture Models: Using Marked Animals to Study Population Dynamics. *BioScience.* 42(2): 94-102

Powell, Jonathan and Clayton Thyne. 2011. Global Instances of Coups from 1950 to 2010. *Journal of Peace Research.* 48(2): 249-259.

Raleigh, Clionadh, Andre Linke, Havard Hegre, & Joakim Karlsen. 2010. Introducing ACLED: An Armed Conflict Location and Event Dataset. *Journal of peace Research*, 47 (5): 651-660.

Reporters Without Borders. 2014. *2011/2012 Press Freedom Index*. Paris, France. Accessed at https://en.rsf.org/press-freedom-index-2011-2012,1043.html, March 17, 2014.

Salehyan, Idean, Cullen Hendrix, Jesse Hamner, Christina Case, Christopher Linebarger, Emily Stull, and Jennifer Williams. 2012. Social Conflict in Africa: A New Database. *International Interactions*, 38 (4), 503-511.

Schrodt, Philip. 2012. Precedents, Progress, and Prospects in Political Event Data. *International Interactions*, 38(4): 546-569.

Sundberg, Ralph, & Melander, Erik. 2013. Introducing the UCDP Georeferenced Event Dataset. *Journal of Peace Research*, 50 (4), 523-532.

Williams, Byron, James Nichols, and Michael Conroy. 2002. *Analysis and Management of Animal Populations.* San Diego, CA: Academic Press

Work Bank. 2014. *World Development Indicators*. Washington, DC: World Bank.

Figure 1. The Process of Data Collection.

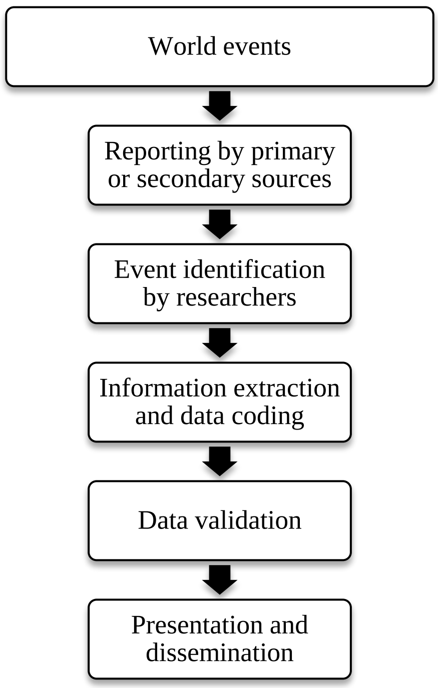

Figure 2: Percent of Social Conflict in Africa Database Events Detected by Source and Number of Participants, 2012. Source: Salehyan et al. (2012).

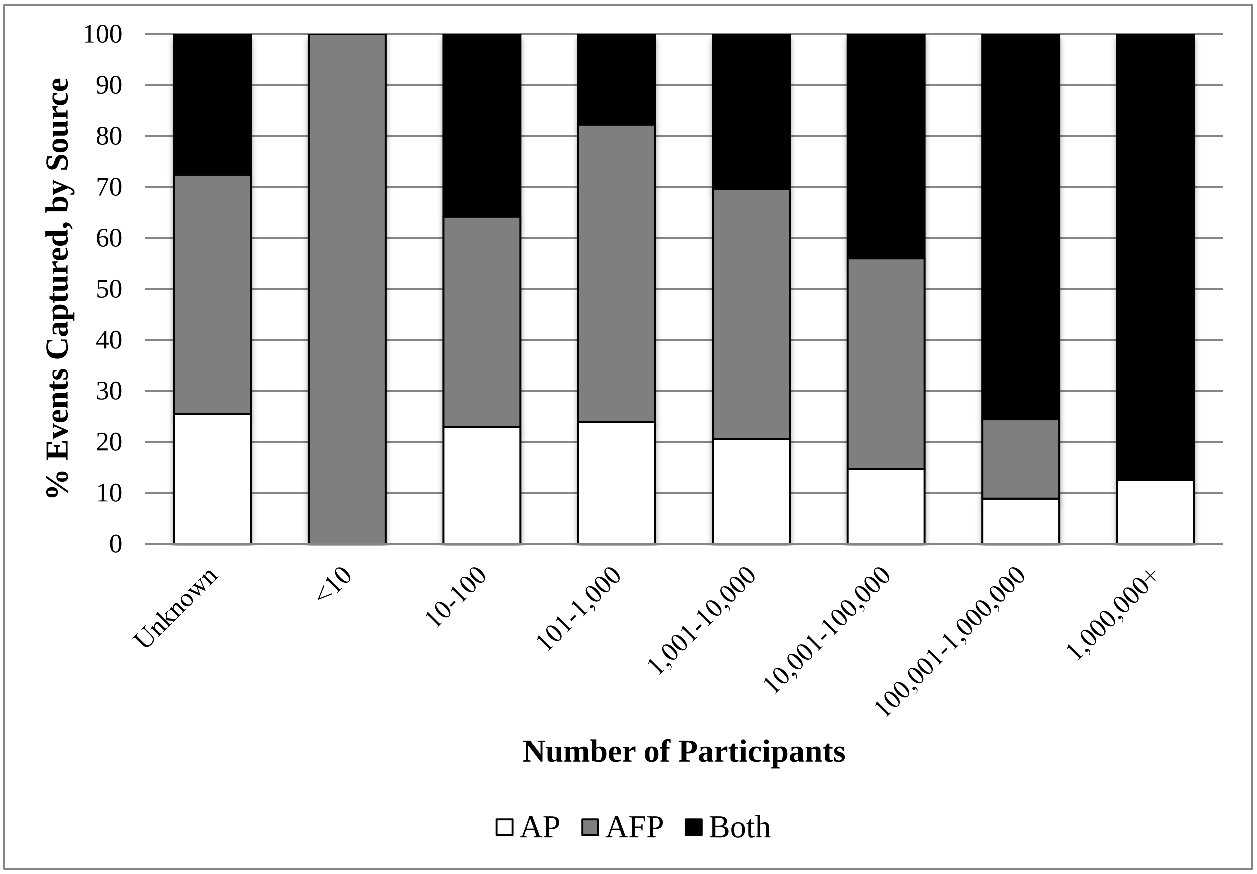

Table 1. Event Detection by News Source in the Social Conflict in Africa Database, 2012.

News Source	Frequency	Percent	Cumulative Percent
AFP Only	662	45.88	45.88
AP Only	325	22.52	68.40
Both	456	31.60	100.00
Total	*1,443*	*100.00*	

Table 2. Regression Analysis of Reporting by the AP and AFP

	1. Bivariate Probit		2. Logit
Dependent Variable:	**A. AP Report**	**B. AFP Report**	**Both Report**
Protests and Riots	**0.133**	**-0.368**	**-0.22**
	(0.257)	(0.253)	(0.296)
Violent	**0.311**	**-0.302**	**0.316**
	(0.312)	(0.294)	(0.326)
Urban	**0.067**	**0.04**	**0.31**
	(0.102)	(0.102)	(0.166)
Central Government	**0.083**	**-0.189**	**-0.083**
	(0.109)	(0.113)	(0.131)
Regional Government	**-0.130**	**0.139**	**0.101**
	(0.159)	(0.142)	(0.225)
Duration	**0.005**	**-0.003**	**0.005**
	(0.004)	(0.004)	(0.005)
Repression Used (0 = None, 1 = Nonlethal, 2 = Lethal)	**0.162***	**0.225*****	**0.611*****
	(0.066)	(0.059)	(0.115)
>1000 Participants	**0.289*****	**0.392**	**1.220*****
	(0.084)	(0.222)	(0.181)
>10 Deaths	**0.765*****	**0.726***	**1.869*****
	(0.195)	(0.305)	(0.365)
Political Issue	**-0.114**	**0.176***	**0.075**
	-0.096	-0.08	(0.146)
Ethnic Issue	**-0.04**	**0.261****	**0.282**
	(0.116)	(0.081)	(0.146)
Democracy	**0.303****	**-0.281****	**0.204**
	(0.094)	(0.108)	(0.198)
Press Freedom	**0.003**	**-0.004**	**-0.002**
	(0.005)	(0.004)	(0.006)
Population	**0.117***	**-0.066**	**0.112**
	(0.059)	(0.055)	(0.065)
North Africa	**0.017**	**0.245**	**0.511**
	(0.211)	(0.239)	(0.276)
Anglophone	**-0.438**	**0.29**	**-0.144**
	(0.279)	(0.253)	(0.317)
Francophone	**-0.747*****	**0.375***	**-0.982*****
	(0.185)	(0.166)	(0.244)
Constant	**-0.637**	**1.255****	**-1.834*****
	(0.380)	(0.385)	(0.429)
Rho	**-0.932*****		--
	0.200		
N	1371		1371

* $p<0.05$, ** $p<0.01$, *** $p<0.001$. Standard errors, clustered by country, in parentheses.