Annu. Rev. Sociol. 2004. 30:65–80
doi: 10.1146/annurev.soc.30.012703.110603

First published online as a Review in Advance on January 7, 2004

THE USE OF NEWSPAPER DATA IN THE STUDY OF COLLECTIVE ACTION

Jennifer Earl,[1] Andrew Martin,[2] John D. McCarthy,[2] and Sarah A. Soule[3]

[1]*Department of Sociology, University of California, Santa Barbara, California 93106-9430; email: jearl@soc.ucsb.edu*
[2]*Department of Sociology, Pennsylvania State University, University Park, Pennsylvania 16802-6207; email: awm127@psu.edu, jxm516@psu.edu*
[3]*Department of Sociology, University of Arizona, Tucson, Arizona 85721; email: soule@u.arizona.edu*

Key Words protest, selection bias, description bias, news source, social movements, protest event analysis

■ **Abstract** Studying collective action with newspaper accounts of protest events, rare only 20 years ago, has become commonplace in the past decade. A critical literature has accompanied the growth of protest event analysis. The literature has focused on selection bias—particularly which subset of events are covered—and description bias—notably, the veracity of the coverage. The "hard news" of the event, if it is reported, tends to be relatively accurate. However, a newspaper's decision to cover an event at all is influenced by the type of event, the news agency, and the issue involved. In this review, we discuss approaches to detecting bias, as well as ways to factor knowledge about bias into interpretations of protest event data.

INTRODUCTION

Scholarship in collective action and social movements has developed a rich research tradition that uses data culled from newspaper reports of these events, among other sources.[1] Subsequent research using a variety of sources has largely validated the central findings of many projects that initially relied on newspaper data.

This research tradition developed because of the numerous opportunities that newspaper-based event data open to scholars both theoretically and methodologically. Event data allow researchers to examine multiple kinds of collective action, from racial violence (Olzak 1989b, 1992; Bergesen & Herman 1998), to agrarian protest and rebellion (Paige 1975), to various types of conventional and nonconventional social movement protests (Earl et al. 2003, Kriesi et al. 1995). In doing

[1]We refer to protest and collective action events interchangeably as "events" in this review.

so, it moves toward a stress on action and away from purely organizational views of social movements (see Melucci 1994 for a discussion of the theoretical importance of this change in conceptualization of movement activity). Newspaper data also facilitate both comparative and historical research (Tarrow 1996) and make quantitative research on social movements more viable (Olzak 1989a). Event data allow researchers more leverage over processual and mechanistic elements in causal explanations (Olzak 1989a). In addition to these numerous advantages, there is often no other alternative available for researchers interested in work beyond case studies of particular movements (Franzosi 1987).[2]

Despite the centrality and analytic utility of newspaper-based event data, recent research critically evaluates this data source. This work, which uses and reflexively assesses newspaper event data, offers a renewed opportunity to evaluate the utility of newspaper data in the study of collective action and social movements.

FROM CLASSIC TO CONTEMPORARY: NEWSPAPER EVENT RESEARCH

There are a number of reviews of existing research using newspaper event data (Franzosi 1987; Koopmans & Rucht 1999; Olzak 1989a, 1992, Rucht et al. 1999). Rather than repeating these, we summarize the major contributions of this research and then turn to an assessment of the criticisms of the approach.

Most major research traditions in social movements have benefited from analysis of newspaper event data. The development of the political process model depended heavily on newspaper event data (Jenkins & Perrow 1977, McAdam 1982), as has related work on political opportunities (Eisinger 1973, Koopmans 1995, McAdam 1982, Tarrow 1989) and protest cycles (Koopmans 1993, Tarrow 1989).[3] Some aspects of resource mobilization have been investigated using these data (Jenkins & Perrow 1977, Jenkins & Eckert 1986, Soule et al. 1999). New social movement (NSM) scholars have used newspaper event data to evaluate their claims (Koopmans 1995, Kriesi et al. 1995, Rucht et al. 1999).[4] Finally, newspaper data have been used to study more spontaneous forms of collective behavior,

[2]Important questions in the study of social movements certainly arise that cannot easily be addressed using newspaper event data. For instance, scholars interested in the internal organizational dynamics of a movement—which may not be evident from data on event occurrence and frequency—are less likely to benefit from newspaper event data.

[3]While research using the *World Handbook Data* (Taylor & Jodice 1983) has faced more criticism (Rucht and Olemacher 1992), political process research relying on other newspaper data has faced less criticism (for a review of political process research see Jenkins & Schock 1992).

[4]Most of the criticism directed at NSM theory has surrounded what is "new" about these movements (Pichardo 1997), as opposed to the veracity of empirical NSM research.

including ethnic violence (Olzak 1989b, 1992), rioting (Danzger 1975), and global collective action (Taylor & Jodice 1983).

Not only has newspaper data contributed to the development of major theories in this field, but it has also illuminated ongoing debates over particular social movement processes. For instance, scholarship on strategic and tactical development, such as the study of repertoires of contention (Tilly 1979, 1995), tactical innovation and diffusion (McAdam 1983, Soule 1997), and tactical overlap (Olzak & Uhrig 2001), has been facilitated by the use of newspaper data. Research on repression and protest control has relied on this kind of data (Earl et al. 2003, Koopmans 1993), as has research assessing the relationship between various social movements and legislative votes and action (McAdam & Su 2002, Soule et al. 1999).

The use of this type of data is likely to continue, adding to the prominence of newspaper data in the study of collective action and social movements. Four major projects that used newspapers as a data source on U.S. and European protest events have contributed to this trend. Kriesi and his collaborators (1995) gathered extensive data on new social movements in Europe, while Rucht and collaborators' (1999) PRODAT project added to the stock of European newspaper-based protest event data. In the United States, a team led by McAdam, McCarthy, Olzak, and Soule collected extensive data on U.S. protest using the *New York Times* (see McAdam & Su 2002 and Earl et al. 2003 for initial discussions of these data). Bond and Jenkins led a team of researchers who compiled two datasets that rely on automated content coding of news articles from Reuters, an electronic news wire service (Jenkins & Bond 2001, Bond et al. 1997). These datasets are called Kansas Events Data System (KEDS) and Protocol for the Analysis of Nonviolent Direct Action (PANDA). Taken together, these four newspaper event projects are likely to offer opportunities for scholars to address diverse research questions in both U.S. and Western European settings.

CURRENT CONTROVERSIES REGARDING NEWSPAPER EVENT DATA

Although newspapers have provided data for some analyses of collective action and social movements, researchers have issued a number of criticisms of the quality of such data, and these criticisms suggest possible limitations on the utility of these data. One set of criticisms addresses researchers' collection practices. Another suggests that newspaper data may have flaws, regardless of the collection method used. In particular, some scholars argue that newspapers selectively report events ("selection bias") or that they erroneously report information on events they cover ("description bias") (McCarthy et al. 1996, 1999).[5]

[5]Others have also discussed "researcher bias," which is introduced through coding and data entry errors (Franzosi 1987). We do not address that source of bias here.

Criticisms of Data Collection Schemes

Some early newspaper research relied on indexes prepared by the newspaper publisher or a private source to identify articles on collective action and social movements (e.g., Jenkins & Eckert 1986, McAdam 1982). However, indexing systems can vary in many ways, including inclusiveness (whether articles that are not primarily about protests, but that touch on protest events, are included in the index), thoroughness (how thorough indexers are in reading articles and compiling the indexes), and consistency (whether consistent terms are employed within an index over time). Researchers therefore argue that indexing does not generate either the total population of protest events reported in newspapers or a representative sample. Rather, it produces a sample of newspaper accounts that is structured according to the indexing methodology.

Recent research has responded to this criticism by moving away from the use of indexes to locate articles on events. For instance, the U.S. protest event project mentioned above uses daily scans of the *New York Times* to locate articles on events. The European counterpart to that dataset, PRODAT, also avoids the use of indexes (Rucht & Ohlemacher 1992), as does the dataset constructed by Van Dyke (2003), which uses student newspapers on nine college campuses in the United States.

Daily newspaper scans are resource intensive, however. Some researchers, hoping to reduce the resources required for data collection, use sampling techniques. For example, one method involves sampling newspapers over time. Some researchers use only Monday editions of European newspapers, arguing that this practice (*a*) captures large events that occurred on weekends (Kriesi et al. 1995), (*b*) is useful in European countries where Sunday papers are not published (Koopmans & Rucht 2002), and (*c*) nets relatively more events because Monday is a slow newsday (Barranco & Wisler 1999, Oliver & Myers 1999). However, this technique reduces the likelihood that labor-related events (e.g., strikes) or student protests will be included in the sample, thus introducing its own set of biases into analyses (Koopmans & Rucht 2002, Rucht & Ohlemacher 1992). Other sampling techniques involve sampling small segments of time, such as weeks, or months, or randomly sampling days in a year.

Whatever the choice of sampling strategies, scholars using purposive sampling schemes should be aware of the possible biases that may be introduced, as well as the undercounting that occurs when one uses any sampling strategy. Because of these possible limitations, some projects (e.g., the U.S. project led by McAdam, McCarthy, Olzak, and Soule, and used by Earl et al. 2003 and McAdam & Su 2002) have opted not to sample days at all, gathering instead the entire population of events reported in a newspaper.

Selection Bias

Some critics argue that, regardless of the sampling technique used, newspaper data suffers from selection bias because news agencies do not report on all events

that actually occur.[6] Critics claim that the sample of events on which newspapers do report is not representative but is instead structured by various factors such as competition over newspaper space, reporting norms, and editorial concerns.

In historical context, this concern over selection bias is ironic. Event research initially gained popularity partly because earlier research designs had sampled on the dependent variable, creating severe sample selection bias. Olzak's review of event research makes this argument, noting, "[S]ocial scientists commonly study only the places that have collective events in studies that try to learn the causes of events. . . . Any attempt to infer causal relationships will be crippled by sample selection bias" (1989a, p. 121). Historically situated, the debate should revolve around how much event analysis and newspaper data represent relative improvements over prior research strategies.

This is not to suggest that concerns about newspaper data are entirely new. Molotch & Lester (1974) were among the first to raise such concerns with newspaper data more generally, and these concerns were soon applied to social movement and collective action data (Danzger 1975, Snyder & Kelly 1977). More recently, Franzosi (1987), Huxtable & Pevehouse (1996), McCarthy et al. (1996, 1999), Hug & Wisler (1998), Koopmans (1999), Rucht & Neidhardt (1999), Barranco & Wisler (1999), Oliver & Myers (1999), and Oliver & Maney (2000) have examined the validity and/or reliability of newspaper protest event data. These researchers have focused on three sets of characteristics in predicting selection biases: (*a*) event characteristics, (*b*) news agency characteristics, and (*c*) issue characteristics.

Some events, it is argued, are seen as more "newsworthy" by the press, and thus are more likely to be reported (Barranco & Wisler 1999, Hocke 1999, McCarthy et al. 1996, Oliver & Myers 1999).[7] Factors that influence judgment of an event's newsworthiness include the proximity of the event to the news agency (regionally, see McCarthy et al. 1996, 1999; internationally, see Mueller 1997), the size of the event (Barranco & Wisler 1999, Hug & Wisler 1998, McCarthy et al. 1996, 1999; Oliver & Myers 1999, Oliver & Maney 2000), the intensity of the event (Mueller 1997), violence at the event (Barranco & Wisler 1999), the presence of counterdemonstrators or police, sponsorship by social movement organizations (SMOs), or the use of sound equipment (Oliver & Maney 2000).

Some scholars also examine how the structure and process of news agencies may independently affect the selection of events. For example, the presence of a wire service in a city can increase the probability of reporting (Danzger 1975), news routines and the production process can affect the quality of news reports

[6]Assessing selection bias requires an independent source of evidence about the population of events on which a newspaper selectively reports.

[7]Lester (1980) argues that newsworthiness is not an event characteristic but is instead a constructed characteristic that is produced through interactions between reporters and editors. Similarly, Davenport & Ball (2002) describe how reporters focused on the context within which the political killings in Guatemala took place, rather than characteristics of the events.

(Gamson et al. 1992, Molotch & Lester 1974), and reporter routines and "beats" may cause selective reporting (Oliver & Myers 1999).

Finally, some critics argue that when events resonate with more general social concerns, they are more likely to be reported. They refer to this as the "issue attention cycle" (Downs 1972) or the media attention cycle (McCarthy et al. 1996, 1999). For example, if general social concern is triggered by legislative conflict over a particular issue, newspapers are more likely to report on events that relate to that conflict (Oliver & Maney 2000).

STABILITY OF SELECTION BIAS To assess how representative of the population a subset of events is, it is critical to understand the stability of selection bias over time and between newspapers. Several research projects report fairly consistent selection processes over time and within papers (Barranco & Wisler 1999, McCarthy et al. 1996, 1999). However, selection bias seems to vary according to the type of paper: local, regional, national, or international.

In contrast to findings on the relative stability of selection bias, one study argues "[t]he verdict is in. . .it is simply not possible to assert, in the absence of data, that the patterns of selection in news coverage of protest events should be assumed to be relatively stable across time or locale or issue" (Oliver & Maney 2000, pp. 494–495). It is important to note, though, that the research upon which this conclusion is based relies on local and regional newspapers, which have been found to be more biased (Barranco & Wisler 1999) than national newspapers in terms of movement-specific reporting biases (e.g., some movements have more events covered than others).

As a final note on sample selection bias, some point to the percentage of events that are eventually reported in newspapers as an aggregate measure of the selectivity of newspapers, and hence selection bias. Several research projects have found average reporting rates approaching one half of all events (Barranco & Wisler 1999 for four Swiss cities, Oliver & Maney 2000 for Madison, Wisconsin), whereas others have recorded substantially lower reporting rates (McCarthy et al. 1996, 1999 for Washington D.C., Titerenko et al. 2001 for Minsk). However, using coverage rates as indicators of selection bias conflates two distinct issues: whether newspapers are a population or a sample, and whether, if a sample, the sample is unrepresentative of the population of events. Thus, it is difficult to evaluate how serious selection bias is from coverage rates because a sample of even 5% of events would not be problematic if it were truly representative.

EVALUATING THE IMPACT OF SELECTION BIAS Although many of the studies reviewed above might seem to offer reasons to be cautious about the use of newspaper event data, several considerations should be taken into account. First, most of this research focuses on locations or time periods in which protests are more institutionalized and, hence, less "newsworthy." For instance, research on selection bias in the United States has examined reports on protest events that occurred in Washington, D.C. (McCarthy et al. 1996), where protest is highly institutionalized

(McCarthy & McPhail 1998) and thus less newsworthy relative to other kinds of events. Other selection bias research in the United States focuses on periods in which protest was argued to have been institutionalized (Oliver & Myers 1999, Oliver & Maney 2000). In fact, Oliver & Maney (2000) base their theoretical concern over selection bias on the claim that since the 1970s, protest has become increasingly institutionalized and thus has decreased in newsworthiness.

The flip side of Oliver & Maney's (2000) claim, though, is that when or where protest is less institutionalized, protest will be more newsworthy, will likely garner more coverage, and may be less susceptible to selection bias. To the extent that Oliver & Maney (2000) and others have argued that institutionalization is responsible for media inattention, their findings may not be generalizable to situations in which protest is less institutionalized. Even Molotch (1979), a critic of newspaper data, acknowledges that critiques of newspapers cannot be "specified in a *general*, that is, ahistorical, acontextual, way" (pp. 91–92, emphasis in original).

Second, researchers should consider how valid generalizations may be from one type of news source to another. Researchers primarily study regional and national newspapers. For instance, while McCarthy et al. (1996, 1999) study national newspapers, Oliver & Myers (1999) and Oliver & Maney (2000) study local newspapers, and Barranco & Wisler (1999) study both national and local newspapers, and the findings are different in important ways. McCarthy et al. (1996, 1999) find that biases were fairly constant between 1982 and 1991. Barranco & Wilser (1999) also find great consistency over time in the sources of selection bias for national newspapers. Myers & Caniglia (2003) report that the direction of coefficients of the correlates of riots were remarkably parallel in equivalent analyses of aggregated local newspaper coverage and *New York Times* and *Washington Post* coverage, although the magnitude of the coefficients were not. Oliver & Maney (2000) find that sources of biases changed from 1993 to 1996 in the local newspapers they examined.

Although Oliver & Maney (2000) imply that these results can be generalized to other newspaper data, including national newspapers, the findings of McCarthy et al. (1996) and Barranco & Wisler (1999) suggest caution in doing so. In fact, although Barranco & Wisler found that local papers had smaller biases toward large and violent demonstrations, they found that "these advantages over national press in terms of validity are somewhat overshadowed by the existence of a strong bias with regard to types of movements" (1999, p. 308).[8]

Third, differences in coding criteria and procedures may account for some of what appears to be selection bias, as suggested by Jackman & Boyd (1979) in their study of collective violence and coups in 30 black African nations. Although

[8]Research on the stability of selection bias across diverse newspapers has been less extensive and has produced mixed results. Martin (2005) finds that the overlap in domestic strike selection by the *New York Times* and *Daily Labor Report* was over 40%, suggesting comparable selection mechanisms. But, Davenport & Litras' (2003) study of Black Panther Party coverage in Oakland finds larger differences between mainstream and advocacy newspapers.

they had anticipated serious selection bias, they found that once they controlled for coding differences between data sources, the remaining selection bias had no impact on the estimates produced in their analyses. Differences in coding criteria are evident in recent work, too. For example, Oliver and her collaborators collect data on public events writ large, including such events as public parties, outdoor concerts, and individual leafleters (Oliver & Myers 1999). Many of the events that Oliver and her collaborators include in their dataset would be *intentionally* excluded from other datasets that focused on public protest events or public acts of collective violence (e.g., McAdam & Su 2002, Earl et al. 2003). In turn, the bias produced by selective reporting on these events would also be excluded. In other words, if one were to examine selective reporting for all the events that Oliver & Myers code, it would overstate the selectivity of reporting for protest events, but this overstatement would be due to differences in coding, not selective reporting.

Finally, the substantive effect of selection bias on estimated regression coefficients remains in question. Hug & Wisler (1998) attempt to address this issue by comparing different corrections for possible selection bias in regional and national newspapers. Using a Heckman two-step procedure (see below for more discussion), they attempt to extract the effects of selection before estimating the theoretically relevant coefficients. However, in a model of the degree of violence at a protest, with protest size and the city in which the protest occurred used as independent variables, the uncorrected coefficient was actually more conservative than the corrected coefficient, suggesting that selection bias might be responsible for a false negative in this case, but not a false positive. Their population level model also shows that the uncorrected estimates are more conservative.

Description Bias

Description bias concerns the veracity with which selected events are reported in the press, and is also a concern when using newspaper data. Findings suggest that although "hard news" (i.e., the who, what, when, where, and why of the event) is mostly subject to errors of omission, "soft news" (i.e., impressions and inferences of journalists and commentators) is subject to multiple sources of bias (McCarthy et al. 1999).

Researchers interested in description bias pursue two tactics. Some explore newspaper description bias across an identified set of events. McCarthy et al. (1998) examine print and electronic coverage of protests that occurred in 1982 and 1991 in Washington, D.C., identifying three dimensions of description bias: (*a*) omission of information, (*b*) misrepresentation of information, and (*c*) framing of the event by the media. They find that newspapers tend to provide more accurate coverage of events than electronic sources, although the print media are beginning to resemble electronic media in terms of soft news coverage. When hard news on protests is reported in newspapers, it is presented reasonably accurately. Smith et al.'s (2001) follow-up research on these events finds that event characteristics related to controversy (e.g., arrests, violence) drew attention to the event and away

from the protest issues, reducing the utility of newspaper data when studying protest claims.

Researchers also explore newspaper coverage of a single protest event. For example, Richards & McCarthy (2003) compare newspaper coverage of the Promise Keepers "Stand in the Gap" Rally to the Million Man March, both of which occurred in the mid-1990s in Washington, D.C. They find that newspaper stories generally portrayed the Million Man March (and Reverend Louis Farrakan) in a more negative light than the Promise Keepers, although specific features of both events (e.g., mobilization and logistics) were often ignored. McPhail & Schweingruber (1999), who used crowd observers to collect data on the 1995 March for Life in Washington, D.C., find that newspapers are likely to report only the most frequent actions of participants.

In explaining description bias, many turn to the corporate hegemony model, which claims that corporate influence results in the portrayal of challenging groups as "radical" or "fringe" elements, especially those that challenge the power of the state or the hegemony of capitalism (Bagdikian 2000, Gitlin 1980, Herman & Chomsky 1988, Lee & Solomon 1990, Parenti 1993, Ryan 1991). A related body of research examines how the labor movement, in particular, is portrayed in an unfavorable light by the media (Beharrell & Philo 1977; Glasgow University Media Group 1976, 1980, 1982; Puette 1992).

EVALUATING THE IMPACT OF DESCRIPTION BIAS Although newspaper data may ignore key dimensions of a protest (e.g., its purpose), when event characteristics are included, especially hard news items, the reports are, in general, accurate, indicating that missing data may be the most serious form of description bias. That is not to say, however, that distortions never occur in newspaper coverage, and reporting may be particularly susceptible to error when information about an event is collected from other sources (e.g., authorities or participants). Because these actors often have a stake in how the event is portrayed, the information they provide can be biased. In addition, newspaper descriptions of protest apparently are becoming more thematic in nature (i.e., linking stories on individual protests to broader issues rather than providing a detailed description of the event itself). This trend may be especially pertinent to research exploring the "soft" characteristics of events, suggesting that scholars will have to be more sensitive to the shifting context within which events are reported. Finally, despite recent efforts to assess description bias systematically, the topic has been neglected in contrast to selection bias, and hence, remains less fully understood.

Criticisms Beyond Collective Action and Social Movement Research

Scholars from a variety of fields rely upon newspaper accounts of events. Here we offer only a brief discussion of such research, focusing on patterns of results relevant to protests and collective actions. In terms of selection bias, studies suggest

that the media focuses on occurrences that have major social impact, analogous to the patterns of selection bias for events that have newsworthy characteristics (e.g., notorious, unusual, large, violent, dramatic, and/or rare). For instance, newspaper coverage of crime focuses on particularly violent and notorious acts, such as murder, rape, aggravated assault, and offenses with multiple victims (Antunes & Hurley 1977, Johnston et al. 1994). Similarly, newspapers tend to cover diseases associated with high or increasing mortality rates (Adelman & Verbrugge 2000). Research in other areas also suggests that proximity influences selectivity in reporting. Media attention to natural disasters shows local and regional bias (Singer et al. 1991), in addition to a bias toward reporting events with large death tolls (Gaddy & Tanjong 1986, Singer et al. 1991). Although most research on newspaper event coverage has explored how specific characteristics of events affect the likelihood of newspaper coverage, some research suggests that cultural factors are also important in explaining variations in newspaper coverage. For instance, Bunis et al. (1996) found that newspaper coverage of homelessness and famine peaked in the holiday season between Thanksgiving and Christmas because of heightened sympathy.

Studies of newspaper portrayals of criminal activity show that reports tend to portray minorities as criminals (Campbell 1995) and female crime victims as to blame for their own victimization, at least implicitly (Meyers 1997). These findings resonate with soft news description biases reported above, in that the media tends to mirror stereotypes and beliefs commonly held by the larger society in soft news reporting.

ADDRESSING METHODOLOGICAL DILEMMAS

Research has begun to explore different methodologies that allow researchers to take into account the problematic aspects of newspaper data on protest and collective action events. These solutions include the triangulation of media sources, the use of electronic archives, and the incorporation of methods employed in survey research to address nonrandom response rates.

Triangulation of multiple sources is used to ensure a broader range of coverage, which is likely both to capture more events (addressing selection bias) and to provide multiple accounts of each event (addressing description bias). For example, Koopmans & Rucht (2002) note that each of the two newspapers coded by PRODAT contributed 25% of unique events in the dataset (i.e., 50% of the events were covered by both newspapers). Using two sources allows them to capture more events and to assess differences in reporting on the same events that are covered by both newspapers. Researchers have long recognized the benefits of relying on more than one data source to ensure a more accurate representation of the event population (Bessinger 1998, Jenkins & Perrow 1977, Khawaja 1994, Myers 1997, Tilly 1995). As discussed above, research on the utility of employing multiple sources for event data has produced mixed results. Although many (Barranco &

Wisler 1999, Jackman & Boyd 1979, McCarthy et al. 1996, Martin 2005) find little variation in bias across sources and time, others (Davenport & Litras 2003, Oliver & Maney 2000) conclude that variation in bias presents a major obstacle when using newspaper data to study events. In light of these mixed results, we recommend that scholars who employ multiple newspapers in an effort to attain more detail on a greater number of events be sensitive to possible causes of variation across the sources used (e.g., time, region, type of newspaper, and political biases).

A second strategy that is closely related to the use of multiple sources is the increasing popularity of electronic archives as sources of data on protest events.[9] These archives consist of either print media (LEXIS-NEXIS, ProQuest Historical *New York Times*, and Newslibrary) or broadcast media (*Vanderbilt Network News Index and Abstract*). These databases allow the researcher to search multiple sources—sometimes hundreds at a time—using keyword search strings. Researchers have begun using these archives (McCarthy et al. 2002, Soule 1997), and studying them (Martin 2005, Oliver & Myers 1999, Richards & McCarthy 2003). For instance, Maney & Oliver (2001) compare police coverage, hard copy newspaper coverage, and NEXIS coverage of collective events occurring in Madison, Wisconsin, in May 1994. They argue that drawbacks exist for both hard copy and electronic searches: Electronic searches may miss events that are framed in unusual ways, whereas hard copy searches may omit events embedded late in the story and risk error because of coder fatigue. However, using both types of search strategies would allow researchers to capture a fuller range of events.

Third, Hug & Wisler (1998) draw upon strategies developed by survey researchers for handling missing cases (i.e., selection bias). They propose either weighting cases on key protest characteristics (e.g., size) or employing statistical corrections for the selection mechanism. Weighting strategies are especially appropriate if bias is constant, but weighting is not as robust a correction in multivariate analyses of variables that are not included in the weighting scheme. For multivariate analyses, they suggest using the Heckman two-step method, which estimates a selection equation (to correct for selection bias) and then estimates the substantive equation in which researchers are interested. Just as was the case with weighting, researchers must assume that the selection mechanism is stable over time. Researchers must also have a population of events available so that they can estimate the selection model.

Although the strategies proposed by Hug & Wisler (1998) can work well for addressing issues of selection bias, they are less appropriate for dealing with description bias. One dimension of description bias identified by McCarthy et al. (1998) is the omission of information in a media account of an event. When studying protest, most scholars are interested in such characteristics as size, location, length, tactic(s), claim, response of authorities, and target. It is very likely that news reports ignore at least some of these characteristics for many protests. In survey

[9]See Kaufman et al. (1993) and Saracevic & Kantor (1991) for a general discussion of the issue of using electronic media archives to study social phenomena.

research, partial missing information is referred to as item nonresponse, which is different from total nonresponse (i.e., selection bias). Kalton (1983) describes common methods of dealing with item nonresponse, such as grouping cases on some variable that is not missing (size, in the case of protest) and then assigning the group mean for the variable to the cases with the missing value. Imputation methods are an increasingly common strategy for dealing with missing data in survey research and could also be applied to missing data in media records of protest events (Allison 2002, Little & Rubin 2002, Rubin 1987).

CONCLUSION: THE CONTINUING CONTRIBUTION OF NEWSPAPER DATA

The use of newspaper data on protest events has helped to further the study of social movements, from the development of theory to the testing of empirical questions that are central to the field. Of course, there are many important social movement questions that cannot be addressed with data on protest events because data on the rate and/or characteristics of protest events cannot shed light on the less public face of social movement organizing. Questions regarding internal organizational dynamics, movement decision making, and leadership, for instance, might be better addressed with other types of data. Nonetheless, as we indicated in the introduction, many questions are made more accessible using newspaper event data.

Event data, for example, allow for the examination of multiple types of collective action, facilitate longitudinal research, and make quantitative research on social movements more viable. And, for many historical and comparative research designs, newspapers remain the only source of data on protest events.[10] Finally, some biases that *have* been found in this source may actually be advantageous for researchers. For example, if researchers are interested in the actual effect of a movement on some outcome, it stands to reason that those covered in newspapers are the relevant events. As Lipsky (1968, p. 1151) notes, "If protest tactics are not considered significant by the media. . .protest organizations will not succeed. Like a tree falling unheard in the forest, there is no protest unless protest is perceived and projected."

Our review has demonstrated that despite growth in research on the quality and possible limitations of newspaper event data, firm conclusions on its strengths and weaknesses remain elusive. Indeed, research on both selection and description bias shows that even though some aspects of event data (e.g., soft news) may be affected by bias, other aspects have withstood a great deal of critical evaluation.

In fact, the evidence suggests that social movement researchers face the same question that almost all other social scientists face: Are the best available, yet

[10]Often newspaper data are employed when government data on events, such as police records, are unavailable. However, scholars have asserted that even when "official" data do exist, they often provide an incomplete record of events (Maney & Oliver 2001, Shalev 1978).

imperfect, data worthy of analysis? We argue that researchers can effectively use such data and that newspaper data does not deviate markedly from accepted standards of quality. For instance, when considering selection bias, newspaper data compare favorably to bias from nonresponses in surveys. The 78% reporting rate for rallies reported by Oliver & Myers (1999) is similar to acceptable response rates in national surveys. Similarly, the coverage rate of protest events in newspapers is probably much higher than the rate of reporting for certain crimes, and yet criminologists continue to study crime rates *because it is often the best data available* (Hindelang et al. 1979). Indeed, Berk's (1983) review of the topic asserts that "the prospect for sample selection bias is pervasive in sociological data" (p. 390). Argues Berk, "Studies of classroom performance of college students rest on the nonrandom subset of students admitted and remaining in school. Studies of marital satisfaction are based on the nonrandom subset of individuals married when the data are collected. Studies of worker productivity are limited to the employed. And, potential problems are complicated by inadequate response rates" (Berk 1983, p. 391). The pervasiveness of bias led Berk to argue that the question should not be whether bias exists, but "whether the bias is small enough to be safely ignored" (Berk 1983, p. 392). To this we would add that the more we know about the structure and sources of bias in our data, the better prepared we will be to avoid erroneous interpretations of its patterns. We conclude that researchers must approach newspaper data with a humble understanding that although not without its flaws, it remains a useful data source. Thus, researchers should avoid both the unexamined use of newspaper data as well as blanket condemnations of its use.

The *Annual Review of Sociology* is online at http://soc.annualreviews.org

LITERATURE CITED

Adelman RC, Verbrugge LM. 2000. Death makes news. *J. Health Soc. Behav.* 41:347–67

Allison PD. 2002. *Missing Data*. Thousand Oaks, CA: Sage

Antunes GE, Hurley PA. 1977. The representation of criminal events in Houston's two daily newspapers. *Journal. Q.* 54:756–60

Bagdikian BH. 2000. *The Media Monopoly*. Boston: Beacon

Barranco J, Wisler D. 1999. Validity and systematicity of newspaper data in event analysis. *Eur. Sociol. Rev.* 15:301–22

Beharrell P, Philo G, eds. 1977. *Trade Unions and the Media*. London: Macmillan

Bergesen AJ, Herman M. 1998. Immigration, race and riot: the 1992 Los Angeles Uprising. *Am. Sociol. Rev.* 63:39–54

Berk RA. 1983. An introduction to sample selection bias in sociological data. *Am. Sociol. Rev.* 48:386–98

Bessinger M. 1998. National violence and the state. *Comp. Polit.* 30:401–22

Bond D, Jenkins JC, Taylor C, Schock K. 1997. Mapping mass political conflict and civil society. *J. Confl. Res.* 4:553–79

Bunis WK, Yancik A, Snow DA. 1996. The cultural patterning of sympathy toward the homeless and other victims of misfortune. *Soc. Probl.* 43:387–402

Campbell CP. 1995. *Race, Myth, and the News*. Thousand Oaks, CA: Sage

Danzger MH. 1975. Validating conflict data. *Am. Sociol. Rev.* 40:570–84

Davenport C, Ball P. 2002. Views to a kill: exploring the implications of source selection in the case of Guatemalan state terror, 1977–1995. *J. Confl. Res.* 46:427–50

Davenport C, Litras MFX. 2003. *Rashomon and repression: A multi-source analysis of contentious events*. Work. Pap., Dep. Polit. Sci., Maryland Univ.

Downs A. 1972. Up and down with ecology—the issue attention cycle. *Publ. Int.* 28:38–50

Earl J, Soule SA, McCarthy JD. 2003. Protests under fire? Explaining protest policing. *Am. Sociol. Rev.* 69:581–606

Eisinger PK. 1973. The conditions of protest behavior in American cities. *Am. Polit. Sci. Rev.* 67:11–68

Franzosi R. 1987. The press as a source of socio-historical data. *Hist. Methods* 20:5–16

Gaddy GD, Tanjong E. 1986. Earthquake coverage by the western press. *J. Commun.* 36:105–12

Gamson WA, Croteau D, Hoynes W, Sasson T. 1992. Media images and the social construction of reality. *Annu. Rev. Sociol.* 18:373–93

Gitlin T. 1980. *The Whole World is Watching*. Berkeley: Univ. Calif. Press

Glasgow Univ. Media Group. 1976. *Bad News*. London, UK: Routledge & Kegan

Glasgow Univ. Media Group. 1980. *More Bad News*. London, UK: Routledge & Kegan

Glasgow Univ. Media Group. 1982. *Really Bad News*. London, UK: Writers & Readers

Herman ES, Chomsky N. 1988. *Manufacturing Consent*. New York: Pantheon Books

Hindelang MJ, Hirschi T, Weis JG. 1979. Correlates of delinquency: the illusion of discrepancy between self-report and official measures. *Am. Sociol. Rev.* 44:995–1014

Hocke P. 1999. Determining the selection bias in local and national newspaper reports on protest events. See Rucht et al. 1999, pp. 131–63

Hug S, Wisler D. 1998. Correcting for selection bias in social movement research. *Mobilization* 3:141–61

Huxtable PA, Pevehouse JC. 1996. Potential validity problems in events data collection: news media sources and machine coding protocols. *Int. Stud. Notes* 21:8–19

Jackman RW, Boyd WA. 1979. Multiple sources in the collection of data on political conflict. *Am. J. Polit. Sci.* 23:434–58

Jenkins CJ, Bond D. 2001. Conflict carrying capacity, political crisis, and reconstruction. *J. Confl. Res.* 45:3–31

Jenkins CJ, Eckhert CM. 1986. Channeling black insurgency. *Am. Sociol. Rev.* 51:812–29

Jenkins CJ, Perrow C. 1977. Insurgency of the powerless: farm workers movements (1946–1972). *Am. Sociol. Rev.* 42:249–68

Jenkins CJ, Schock K. 1992. Global structures and political processes in the study of domestic political conflict. *Annu. Rev. Sociol.* 18:161–85

Johnston JWC, Hawkins DF, Michener A. 1994. Homicide reporting in Chicago dailies. *Journal. Q.* 71:860–72

Kalton G. 1983. *Introduction to Survey Sampling*. Thousand Oaks, CA: Sage

Kaufman PA, Dykers CR, Caldwell C. 1993. Why going online for content analysis can reduce research reliability. *Journal. Q.* 70:824–32

Khawaja M. 1994. Resource mobilization, hardship, and popular collective action in the West Bank. *Soc. Forces* 73:191–220

Koopmans R. 1993. The dynamics of protest waves. *Am. Sociol. Rev.* 58:637–58

Koopmans R. 1995. *Democracy from Below*. Boulder, CO: Westview

Koopmans R. 1999. The use of protest event data in comparative research. See Rucht et al. 1999, pp. 90–110

Koopmans R, Rucht D. 1999. Protest event analysis—where to now? *Mobilization* 4: 123–30

Koopmans R, Rucht D. 2002. Protest event analysis. In *Methods of Social Movement Research*, ed. B Klandermans, S Staggenborg, pp. 231–59. Minneapolis: Univ. Minn. Press

Kriesi H, Koopmans R, Dyvendak JW, Giugni MG. 1995. *New Social Movements in Western Europe*. Minneapolis: Univ. Minn. Press

Lee MA, Solomon N. 1990. *Unreliable Sources*. New York: Carol Publ. Group

Lester M. 1980. Generating newsworthiness: the interpretive construction of public events. *Am. Sociol. Rev.* 45:984–94

Lipsky M. 1968. Protest as political resource. *Am. Polit. Sci. Rev.* 62:1144–58

Little RJA, Rubin DB. 2002. *Statistical Analysis with Missing Data*. Hoboken, NJ: Wiley-Intersci.

Maney GE, Oliver PE. 2001. Finding event records. *Sociol. Methods Res.* 29:131–69

Martin AW. 2005. Addressing the selection bias in media coverage of strikes: a comparison of mainstream and specialty print media. *Res. Soc. Mov. Confl. Change*. In press

McAdam D. 1982. *Political Process and the Development of Black Insurgency, 1930–1970*. Chicago: Univ. Chicago Press

McAdam D. 1983. Tactical innovation and the pace of insurgency. *Am. Sociol. Rev.* 48:735–54

McAdam D, Su Y. 2002. The war at home: anti-war protests and Congressional voting, 1965–73. *Am. Sociol. Rev.* 67:696–721

McCarthy J, McPhail C. 1998. The institutionalization of protest in the United States. In *The Social Movement Society*, ed. DS Meyer, S Tarrow, pp. 83–110. Boulder, CO: Rowman & Littlefield

McCarthy JD, Martin AW, McPhail C, Cress D. 2002 . *Mixed-issue campus disturbances, 1985–2001: describing the thing to be explained.* Presented at Annu. Meet. Am. Sociol. Assoc.

McCarthy JD, McPhail C, Smith J. 1996. Images of protest: dimensions of selection bias in media coverage of Washington demonstrations, 1982 and 1991. *Am. Sociol. Rev.* 61:478–99

McCarthy JD, McPhail C, Smith J, Crishock LJ. 1999. Electronic and print media representations of Washington D.C. demonstrations, 1982 and 1991: a demography of description bias. See Rucht et al. 1999, pp. 113–30

McPhail C, Schweingruber D. 1999. Unpacking protest events. See Rucht et al. 1999, pp. 164–95

Melucci A. 1994. A strange kind of newness. In *New Social Movements: From Ideology to Identity*, ed. E Laraña, H Johnston, JR Gusfield, pp. 101–30. Philadelphia: Temple Univ. Press

Meyers M. 1997. *News Coverage of Violence Against Women*. Thousand Oaks, CA: Sage

Molotch H. 1979. Media and movements. In *The Dynamics of Social Movements*, ed. MN Zald, JD McCarthy, pp. 71–93. Cambridge, MA: Winthrop

Molotch H, Lester M. 1974. News as purposive behavior: on the strategic use of routine events, accidents, and scandals. *Am. Sociol. Rev.* 39:101–12

Mueller C. 1997. International press coverage of East German protest events, 1989. *Am. Sociol. Rev.* 62:820–32

Myers DJ. 1997. Racial rioting in the 1960s. *Am. Sociol. Rev.* 62:94–112

Myers DJ, Caniglia BS. 2003. *Distance related bias in newspaper coverage of civil disorders, 1968–1969*. Work. Pap., Dep. Sociol., Univ. Notre Dame

Oliver PE, Maney GM. 2000. Political processes and local newspaper coverage of protest events: from selection bias to triadic interactions. *Am. J. Sociol.* 106:463–505

Oliver PE, Myers DJ. 1999. How events enter the public sphere. *Am. J. Sociol.* 105:38–87

Olzak S. 1989a. Analysis of events in studies of collective actions. *Annu. Rev. Sociol.* 15:119–41

Olzak S. 1989b. Labor unrest, immigration, and ethnic conflict in urban America, 1880–1914. *Am. J. Sociol.* 94:1303–33

Olzak S. 1992. *The Dynamics of Ethnic Competition and Conflict*. Stanford, CA: Stanford Univ. Press

Olzak S, Uhrig SCN. 2001. The ecology of tactical overlap. *Am. Sociol. Rev.* 66:694–717

Paige JM. 1975. *Agrarian Revolution*. New York: The Free Press

Parenti M. 1993. *Inventing Reality*. New York: St. Martin's

Pichardo NA. 1997. New social movements: a critical review. *Annu. Rev. Sociol.* 23:411–30

Puette WJ. 1992. *Through Jaundiced Eyes*. Ithaca, NY: ILR Press

Richards A, McCarthy JD. 2003. *Description bias in newspaper coverage of mass gatherings*. Work. Pap., Dep. Sociol., Penn. State Univ.

Rubin DB. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley

Rucht D, Koopmans R, Neidhardt F, eds. 1999. *Acts of Dissent*. New York: Rowman & Littlefield

Rucht D, Neidhardt F. 1999. Methodological issues in collecting protest event data. See Rucht et al. 1999, pp. 65–89

Rucht D, Ohlemacher T. 1992. Protest event data: collection, uses and perspectives. In *Studying Collective Action*, ed. M Diani, R Eyerman, pp. 76–106. London: Sage

Ryan C. 1991. *Prime Time Activism*. Boston: South End Press

Saracevic T, Kantor P. 1991. Online searching: still an imprecise art. *Libr. Journal*. 116:47–51

Shalev M. 1978. Lies, damned lies, and strike statistics: the measurement of trends in industrial conflict. In *The Resurgence of Class Conflict in Western Europe Since 1968*, ed. C Crouch, A Pizzorno, pp. 1–19. New York: Holmes & Meier

Singer E, Endreny P, Glassman MB. 1991. Media coverage of disasters. *Journal. Q*. 68:48–58

Smith J, McCarthy JD, McPhail C, Augustyn B. 2001. From protest to agenda building: description bias in media coverage of protest events in Washington, DC. *Soc. Forces* 79:1397–423

Snyder D, Kelly WR. 1977. Conflict intensity, media sensitivity and the validity of newspaper data. *Am. Sociol. Rev.* 42:105–23

Soule SA. 1997. The student divestment movement in the United States and tactical diffusion: the shantytown protest. *Soc. Forces* 75:855–83

Soule SA, McAdam D, McCarthy JD, Su Y. 1999. Protest events: cause or consequence of state action? *Mobilization* 4:239–56

Tarrow S. 1989. *Democracy and Disorder*. Oxford: Clarendon

Tarrow S. 1996. Social movements in contentious politics. *Am. Polit. Sci. Rev.* 90:873–83

Taylor CL, Jodice DA. 1983. *World Handbook of Political and Social Indicators*. New Haven, CT: Yale Univ. Press

Tilly C. 1979. Repertories of contention in America and Britain, 1750–1830. In *The Dynamics of Social Movements*, ed. M Zald, J McCarthy, pp. 126–55. Cambridge, MA: Winthorp

Tilly C. 1995. *Popular Contention in Great Britain 1758–1834*. Cambridge, MA: Harvard Univ. Press

Titarenko L, McCarthy JD, McPhail C, Augustyn B. 2001. The interaction of state repression, protest form and protest sponsor strength during the transition from Communism in Minsk, Belarus, 1990–1995. *Mobilization* 6:129–50

Van Dyke N. 2003. Crossing movement boundaries: factors that facilitate coalition protest by American college students, 1930–1990. *Soc. Probl.* 50:226–50