Researchers collecting data on collective events have often relied on ad hoc procedures and untested assumptions. A detailed comparison of different sources and data-gathering strategies for one month of events in one city challenges several assumptions made in prior studies. Police agencies and news sources have their own logics of jurisdiction, producing both incomplete and selective records of collective events. Different strategies for searching newspapers for stories about events yield very different results. News items about events do not occur at the same time as the events but are dispersed around them. All three of these patterns vary greatly from source to source and locale to locale. Comparisons of research findings between localities must more explicitly consider these selection issues. All data collection projects need to include methodological analyses of selection structures in their particular cases so that there can be some basis for comparison between studies.

Finding Collective Events

Sources, Searches, Timing

GREGORY M. MANEY
Hofstra University

PAMELA E. OLIVER
University of Wisconsin–Madison

Escaping a tradition of relying on the *New York Times* and other major newspapers, researchers have recently tapped into a wide variety of media and police records to construct data sets for quantitative studies of the shifting patterns of protest across time and locale. The computerization of media and police records has made it

AUTHORS' NOTE: *We thank Lt. Michael Smith of the Madison Police Department; Chief Michael Metcalfe, Ms. Sue Barica, and Mr. Walter Peterson of the Capitol Police Department; Ms. Mikele Stillman of the Street Use Committee, Lt. Glen Miller, Mr. Jeffrey Bender, and Ms. Pat McGuire of the University of Wisconsin Police Department; and Mr. Frank Denton and Mr. Dave Zweifel, editors of the* Wisconsin State Journal *and* The Capital Times*, respectively, for their generous help and cooperation with this research project. We also thank Daniel J. Myers, who collected the Madison Police Department data; Winn Collins, Elizabeth Engstrand, Sheridan Bearheart, Maureen MacDonald, and Scott Zdrazil for their diligent work as undergraduate research assistants; and Ivan Ermakoff, John D. McCarthy, and the anonymous reviewers for their comments in the process of revision. This research was funded by National Science Foundation grants SBR-9511748 and SBR-9819884.*

SOCIOLOGICAL METHODS & RESEARCH, Vol. 30 No. 2, November 2001 131-169

much easier to accumulate large amounts of data about public events. However, as in any new mode of inquiry, initial research procedures have been ad hoc and have varied tremendously between projects. Often, studies have relied on untested assumptions about the relationship between the data and the underlying pool of events they are supposed to represent. Methodological studies are needed to identify these relationships and the implications of using different sources of event information as well as different strategies for searching these sources. This study reports the results of one such study using multiple records of public events for one month in a small U.S. city. Each record source and search strategy identifies a different subset of events that produce different views of "what happened" in that month. By comparing these different sources and search strategies, we are able to assess their selection structures. Such assessments are essential if we wish to meaningfully use event data collected in separate locales using different sources and different search strategies.

Our study yields three clear results. First, no single media or police source captures all, or even a representative sample, of the events of interest. Each source has a distinct orientation in its coverage. Media analysts have shown that different news sources cover different kinds of events at different rates and in different ways (e.g., McCarthy, McPhail, and Smith 1996; Hertog and McLeod 1995). We find that this also applies to police sources. Police permit records are different from logbooks in the manner and consistency of their creation. Moreover, police agencies differ dramatically in the type of information they record and the consistency with which they record it.

Second, alternative methods of searching for descriptions of collective events in newspapers locate different subsets of the event descriptions contained in these sources. Reading the text of newspapers, electronically searching papers using keywords, and electronically searching papers using event-specific information gleaned from extra-media sources present three alternative methods for locating media coverage of events. Each method locates event coverage not found using the other two. Our results cast serious doubt on the assurances that past studies found "all" coverage of relevant events in their media sources. The findings also make us suspicious of any search procedure that is vaguely specified or not validated against other

search strategies and point to a need for more research on methods for collecting event data.

Finally, our findings challenge the (usually tacit) assumption in prior research that newspaper coverage of an event closely coincides with the event. We find that although news coverage of an event is typically concentrated in the few days around the event, a significant proportion of news coverage occurs weeks or even months before and after the event. These results point to problems in prior media sampling designs.

PRIOR RESEARCH

Before presenting the findings of our research, we review the assumptions made by past studies of collective events. In particular, we focus on the choices other researchers have made regarding their sources of event data, their methods for searching these sources, and the time intervals within which they assume event records will appear.

SOURCES

Until very recently, most research on collective events has relied on newspaper archives because they are readily accessible and regularly record events across time and space. But it has long been recognized that newspapers neither cover "all" events nor sample randomly from the universe of events. As a result, claims about the relative preponderance of certain kinds of actions or about changes in repertoires of contention across time or space cannot easily be empirically distinguished from claims about the kinds of actions newspapers choose to report or the changes in media reporting behavior through time. For this reason, scholars have agreed in principle that it is important to compare events reported in the media to some standard as a way of assessing media event selection structures.

The problem, of course, has been finding appropriate standards. Earlier research asked news editors how they chose stories (White 1950; Breed 1955; Sigelman 1973; Wolfsfeld 1984) or compared media sources with one another (Sande 1971; Jenkins and Perrow 1977; Arrendell 1972). Although much research has relied on the

national "media of record," it is well established that the national media overlook large numbers of events covered in local or regional media. Snyder and Kelly (1977:119) found that the *New York Times* covered 18.3 percent of protests selected by local newspapers in 43 major U.S. cities over a five-month period.[1]

Because of the obvious limitations of using media sources to discover what types of events the media do and do not cover, other studies have compared media sources with external, nonmedia sources. Rosengren (1973) assessed the international news coverage of 272 parliamentary elections from 1961 through 1970 identified through official sources. Weimann and Brosius (1991:337) drew on the RAND corporation chronology of international terrorism to identify the attributes of terrorist events that affected media coverage. Unfortunately, these external sources often were created using media sources. Danzger (1975:571) found that most congressional records of racial conflict were based on local and/or national newspaper stories. Similarly, according to Weimann and Brosius, the RAND chronology uses "newspapers, journals, and periodicals, as well as governmental, intelligence, and military publications" (p. 337). So, more recent studies have turned toward a more truly independent external source of event data—police records.

Event selection in police records. McCarthy et al. (1996) pioneered the use of police records to assess the event selection structures of media sources. The authors collected permit records from three police agencies to identify more than 3,000 "demonstration" events taking place in Washington, D.C. during 1982 and 1991.[2] Their study has spawned a small industry of police-media research in several countries (Rucht, Koopmans, and Neidhardt 1998).

The first assumption in police-media research was that police records could be treated as a close approximation to the universe of collective events actually occurring (McCarthy et al. 1996; Hocke 1998). However, police records may be just as subject as any other source to variability in their comprehensiveness and accuracy (Rucht and Neidhardt 1998:72). Recent studies have identified a substantial proportion of events covered by newspapers that are absent from police records (Beissinger 1998; Gentile 1998). While police records remain useful for validating news sources, the selection logic of the

police records themselves must be interrogated if the full potential of this comparative methodology is to be realized.

Researchers using police records should understand the logic of jurisdiction. Police agencies have different legal and spatial jurisdictions in different institutional contexts. While the police in Germany and France sometimes maintain logs devoted explicitly to recording protest events (Hocke 1998), political criteria for record keeping are more suspect in the United States. Even these European protest records were less reliable for some event forms than for others (Hocke 1998). Differences also exist between the types of events recorded in logs versus permit records. McCarthy et al. (1996) report for their research on Washington, D.C. in 1982 and 1991 that "only a very small number of demonstrations that appear in the media record are not in the permit record" (p. 484). However, small events are especially likely to be omitted from both news accounts and permit records. The Metropolitan Police of the District of Columbia (MPDC) does not require permits for protest events that do not block traffic, and the National Park Service does not require permits for events that have fewer than 25 people. Events recorded only in the MPDC desk sergeant's logbook accounted for the majority of demonstrations in the agency's jurisdiction, amounting to 13 percent of the total number of demonstrations included in their multijurisdictional analysis for 1982 and 16 percent for 1991. Thus, record-keeping regulations resulted in event permit records missing a significant proportion of protest events taking place.

Whether selectivity in media and extra-media sources presents data validity problems depends on the research question. A researcher measuring levels of protest in terms of the number of people participating in protest events can be reasonably assured that police and media records provide adequate data, as the largest events are likely to be recorded in both. If, however, the researcher wishes to study the total number of distinct protest events, the early or declining stages of a social movement, or small or marginalized movements, then the systematic exclusion of smaller events by both sets of records makes relying on either of these sources alone problematic. In all cases, the limitations of available data sources must be identified as part of determining the appropriateness of a given research design for a particular research question.

SEARCHES

Just as sources do not capture all collective events, searches do not gather all of the events that are, in fact, contained in the source. Despite the impact of recent technological advances on the ways sociologists conduct research, no one to our knowledge has yet to consider the methodological ramifications of new techniques for collecting media event data. Several media studies neglect to even mention how the researcher gathered media items included in the analysis. Yet search techniques have major implications for the validity of analyses of media coverage of collective public events. Search techniques vary in the total volume, chronological range, and types of articles captured on events, as well as in their sensitivity to coverage of different types of events.

Index, full-text, and electronic database searches. Until recently, methodological debates about searching media archives centered on the efficiency and ease of using indices versus the completeness of scanning hard copies of newspapers and television scripts (Snyder and Kelley 1977; Eisinger 1973; Danzger 1975; McAdam 1982; Everett 1992; McCarthy et al. 1996; Hocke 1998). Most researchers have concluded that the *New York Times* index is fairly complete, while indexes for other newspapers are not. Now that archives of newspapers are being searched electronically, however, scholars have increasingly recognized the limitations of relying on indexes. So, researchers have begun to use electronic databases of media stories. For instance, McCarthy et al. (1996) conducted a supplemental search of the NEXIS database to identify the amount of news coverage a few top issues had received. Oliver and Myers (1999) used NEXIS searches to analyze news stories collected about events. Despite their increasing popularity, little methodological information has been published on electronic searches.

Comparative dimensions of media data searches. Electronic databases would appear to combine the best of both indexes and full readings of the text of media records. Anyone who has spent bleary-eyed hours reading newspapers on microfilm understands that fatigue and boredom diminish human performance. Electronic keyword searches

promise to reduce the number of irrelevant stories that coders need to read. Most search engines highlight the keywords in the text, permitting the reader to rapidly determine whether the highlighted section is relevant. These benefits can be obtained, however, only if adequate search strategies are developed to include all the relevant stories while excluding many of the irrelevant stories. Staring at a computer search screen can be as tedious as reading microfilm.

Apart from potential savings in reading time, electronic searches have several additional advantages over reading through newspapers and scripts. First, more than one source can be searched simultaneously. Second, while printed newspapers differ greatly in their formats in ways that affect readability, items from different media sources will appear in the same style on the computer monitor. Third, electronic searches eliminate distortions associated with the practice of "scanning" texts. To deal with time, money, and attention constraints, researchers often direct their assistants to examine only certain sections, then the headlines, captions, and first and last sentences of the media item (Rosengren 1973; Rucht and Neidhardt 1998). Electronic searches can efficiently find relevant text in nonnews sections and small mentions of events appearing deep in stories that would otherwise be missed by scanning. We now explore various dimensions of electronic data collection. Since, to our knowledge, no methodological studies have addressed this topic, we draw primarily from our own research as a basis for discussion.

Electronic database search strategies. Not all electronic database searches will yield the same data.[3] Each search consists of a set of parameters whose specifications are of considerable importance. Since only articles and graphics captions containing the key terms entered will be yielded by the search engine, what terms are used and how they are combined will greatly affect the subset of the population of event-related items in the database yielded. The objective is to find a set of terms and search strings that will produce as many media items on events of interest as possible while minimizing the amount of irrelevant items to review.

Through discussions with other researchers and in the course of our own work, we have identified two distinct electronic search strategies. The first involves entering key terms that describe the general type of

event included for analysis. We refer to this as a "generic event descriptor" search strategy. For instance, a researcher might specify a search for coverage on protest events using the following terms: "demonstrato*, demonstrati*, rall* or vigil* or picket*, strike*, sit-in*, marche*, marchi*, protest*, riot*."[4] This strategy assumes that each story on a protest event will describe its specific form at least once with commonly used terms.

Another approach uses both extra-media and inter-media data to generate search terms that describe the event in question. We label this as an "event-specific protocol" search strategy. First, extra-media data on specific events of interest are culled for information regarding the names of individuals and organizations participating, date, type of event, actions taking place during the event, messages or issues raised, and location. Next, the event-related stories yielded using this event-specific information are read for additional event-related search terms. The process continues until no further event specific terms are discovered.

Each strategy possesses strengths and weaknesses. The generic event descriptor strategy offers the advantage of identifying a broad range of events of interest without prior knowledge of their existence. Thus, for research in which extra-media sources of data are unavailable or for studies of the event selection structures of extra-media sources, this may be the only feasible *initial* electronic search strategy.[5] On the other hand, because of its narrow focus on specific characteristics of events likely to be reported in stories, the event-specific protocol search strategy should yield a higher level of information on events listed in extra-media sources than the generic event description approach. The event-specific strategy also is less sensitive to variations in signifiers deployed to describe events. Not every story describing a protest event uses generic terminology. In fact, journalists sometimes describe events without using any event-type word at all. Moreover, because some of the event descriptors are words frequently used to describe nonmessage events (i.e., the use of the term *rally* in a description of the ebb and flow of a basketball game), generic event descriptor searchers can yield very high proportions of false hits while still missing a significant proportion of the stories of interest.

Validity, efficiency, and Boolean search strings. The number and type of articles and stand-alone graphic captions yielded by electronic searches depend not only on the key terms selected but also on how the researcher combines these terms in a search string. Searches generally deploy Boolean logic to connect defining terms. The sequence of connecting terms used affects the size and characteristics of the subset of the population of items captured by the search. The use of "OR" to connect two key terms produces items containing *either* search term, while the use of "AND" to join the same terms generates items containing *both* search terms. Unless the presence of either key term requires the presence of the other, the first search will yield a greater number of stories than the second. Hence, the use of OR connectors widens searches, while the use of AND connectors narrows it. A search connecting all terms by OR maximizes validity at the expense of efficiency. More relevant stories and graphics are found but at the price of wading through a larger number of irrelevant items. On the other hand, a search connecting all key terms by AND maximizes efficiency at the expense of validity. There will be fewer irrelevant media items to wade through, but relevant coverage will frequently be missed.

We were unable to construct an event-specific search protocol that would efficiently find all stories about a given event within an 18-month time range. Instead, as we worked on searches, we developed a standardized procedure that would balance completeness with feasibility. Research assistants followed the rule that articles and graphic captions would be read only if a search string yielded 100 or fewer "hits." First, each keyword from the police record was searched one at a time. The articles and stand-alone graphics yielded by these searches were read if 100 or fewer were captured. Next, Boolean searches were constructed, placing the keyword judged most likely to be included in an event-related item after AND, while the other terms that had not yet been searched (because they had yielded more than 100 hits) were connected by OR. This procedure was repeated until there were 100 or fewer stories to be read.

Often, the search procedure required flexibility and intelligence in identifying the best way to organize the search string. Our

undergraduate research assistants varied in their ability to follow the standard sequence and to guess right about the best ordering of keywords in a search. Even after extensive practice, assistants would find considerably different numbers of items when conducting searches on the same event. As a rough measure of intercoder agreement on the presence or absence of event coverage, we took the number of events for which all assistants captured one or more media items (i.e., articles or stand-alone graphics) and divided it by the total number of events for which at least one assistant captured one or more items. Values for different event samples ranged from a low of 63 percent agreement to a high of 90 percent. These variations stemmed from different levels of concentration, attention to detail, and comprehension of the protocol. To increase reliability, assistants' search logs were checked by a supervisor. To increase validity, two assistants were assigned to search for the same events. In retrospect, searching the entire 18-month interval at one time also contributed to fatigue. In the future, we would conduct standardized searches on smaller windows of time. For example, assistants would first search the 2 weeks before and after the event. Next, they would search the month before and after that block, then the month before and after those, and so on.

All of this suggests that electronically searching media sources is not a panacea. There is often too little congruence between the words used in police reports and the words used in newspaper accounts to permit any simple matching. Only by thinking of synonyms, bringing in local knowledge, and wading through dozens of unrelated stories was it possible to locate some of the coverage there to be found. The event-based search protocol should capture media coverage of events that would be missed using other search strategies. The difficulty of conducting these searches reliably, however, might suggest that this approach is a dead-end. Unfortunately, our analysis of alternate approaches suggests that they also have their problems.

TIMING

Regardless of which search method a researcher chooses to employ, search yields will be affected by what dates within the data source are searched. Scholars have generally assumed that media

coverage of an event occurs immediately after the event and accordingly have searched police and media sources for roughly the same date intervals. The rationale for PRODAT's (Documentation and Analysis of Protest in the Federal Republic of Germany) procedure of sampling Monday issues and the remaining issues from every fourth week (Rucht and Neidhardt 1998) is the claim that most movement-relevant events occur on weekends, coupled with the assumption of next-day coverage.[6] McCarthy et al. (1996) only read newspapers and permit records for the same-year intervals during which the events of interest took place (1982, 1991).

Our findings from pilot searches of Madison-area newspapers, along with our understanding of the ways that social movement organizations and social media outlets use each other for their own particular purposes, made us wary of such assumptions. Coverage of permitted events often occurs *before* the event. Event organizers use the media to notify potential participants in advance. So, major planned events frequently receive most of their coverage before the event. On the other hand, a particularly disruptive "surprise" protest event may generate coverage and commentary for weeks, even months, after its occurrence. Some events are barely mentioned when they occur but escalate in public importance through time, becoming referred to again and again. Reporters use anniversaries as "hooks" to revisit events. Stories on annual events frequently discuss what happened at last year's affair. Searching media and police records for the same set interval orients the media search in favor of events in the middle of the interval, as well as toward preevent coverage of events late in the year and postevent coverage of events early in the year. This orientation will produce distorted comparisons of media coverage for events occurring at different times of the year. Virtually all previous research comparing police and media records is at risk of reporting biased results because of this problem.

Event-centered searches avoid this difficulty by searching equal-length intervals for media coverage around each separate event. Equal-interval searches are essential for providing the necessary data for assessing chronological patterns of media coverage and for permitting the creation of sampling designs that avoid timing selectivity.

METHOD

This study is based on a compilation of protest events listed in several types of records kept by four public agencies. The agencies are the Capitol Police Department, the City of Madison Street Use Committee, the Madison Police Department, and the University of Wisconsin Campus Police Department. Each source was compiled according to a different logic, requiring different rules for selecting events.

IDENTIFYING EVENTS IN POLICE RECORDS

Our own research efforts illustrate the significant variations in record keeping and the difficulties these variations pose to extracting event data from police sources. Only through coming to an understanding of the different logics used in constructing various records were we able to locate the event data available and identify their limitations. We had access to the complete permit records for the Madison Police Department and the Capitol Police Department. So, we treated these permit records as our universe of interest, applying almost no prescreening to permit records. Parade or street use permits are required for any event that disrupts traffic. We coded the file folder full of these paper permits available from the traffic sergeant in the Madison Police Department. From the Capitol Police Department, we were able to download the complete files of all permits. A permit is required for any activity on state property, so these records capture most events taking place at the capitol. We retained all permitted events as "public" events, excluding only weddings, photography sessions, and a few meetings about employee-related issues held in meeting rooms, assuming that these were not "public" events per se. A city street use coordinator and the street use committee administratively review and authorize permits to close a street for any activity. Permits, committee agendas, and the coordinator's daily planner were all coded as sources of information about additional public events.

Wisconsin's Capitol Police Department permit records are fairly comprehensive for events at the capitol and contain standardized information, although less of it than the Washington, D.C. permit records used by McCarthy et al. (1996). Madison Police Department permit records are considerably less standardized and less reliable

than the capitol records. The parade permits contain some standardized information fields, but the forms are not always completely filled out. Furthermore, maintenance of the permit file is unsystematic, with old permits being discarded when they are no longer "needed."

Regardless of the police agency, log records are even less systematic and complete than the permit records. The Capitol Police Department, the Madison Police Department, the Campus Police Department, and the City of Madison Street Use Committee all provided logs produced during or after the events recorded. Officers at all the police agencies, however, reported that they would often not generate any log or report record for peaceful protests or other gatherings they observed if no laws were being violated. They also made it clear that they had no interest in maintaining anything resembling the kind of detailed protest log that Hocke (1998) relied on in Freiburg, Germany. At the Madison Police Department, officers stated that *not* writing reports about legal protests was part of a policy of reducing needless paperwork. At the Campus Police Department, officers indicated that legal protests were rarely recorded in earlier years but that they were attempting to systematize such recording after the installation of keypad terminals in squad cars that allow five-line "dispositions" to be typed while on patrol. At the Capitol Police Department, officers reported a general sense that they should report peaceful protests but that the consistency of reporting ebbed and flowed. All officers stated that they could spend several hours patrolling a large peaceful rally without generating any log entry or report about the event. Any protest that resulted in an arrest or a situation where the police might be accused of misconduct did, however, result in a careful written report giving the officer's side of the story.

Among the protests actually recorded in police logs, some official reports give full and detailed information about the sequence of events, while most made available to us do not.[7] The purpose of a report is mainly to justify an arrest or provide the officer's side of the story in cases where there is a potential or actual allegation of police misconduct. The emphasis in the log is on policing. Many entries about protests contain no mention at all of the issue involved and only vague descriptions of the number of participants and their activities. Often, the only clues distinguishing disruptive protests from a few peaceful pickets are "information" calls in the log about officers who

have been called in early or asked to work overtime. The log often either contains no record at all of large permitted rallies (for which we have permit records) or provides only a cursory mention, such as "rally over" or "handled lost child at the rally."

The categorization of protest events recorded in police logs was equally unsystematic. Incident codes in all police logbooks are organized to meet Uniform Crime Report (UCR) requirements, but there is no UCR code for collective events generally or any specific type of protest. As a result, there is no standardized coding of such events even when they are reported.[8] Capitol and Madison Police Department officers told us that dispatch calls about protests were often coded by the dispatchers as "information"—a catch-all code that is used dozens of times a day for all sorts of miscellaneous comments. Current University of Wisconsin Campus Police protocol calls for protests to be coded as "special events" (along with sporting events, concerts, and the like), but this was not consistent even for recent years.

As a result, different selection procedures were used to obtain what log records on collective events did exist at each of these agencies. In general, these procedures were ad hoc and judgmental, requiring assistants to decide whether an event should be coded. In all cases, assistants were instructed to err on the side of inclusion—recording events that might later be deemed irrelevant. Anything that was labeled by the police as a "protest," "rally," or "demonstration" was coded. Anything that referred to a public gathering was coded.

The search logic and categories in the Madison Police Department computerized records were untenable for this project. A member of the research team read every one of the 130,000 entries in the 1994 dispatch log, looking for any event that could be relevant to the project. This log lists a few lines of text typed by the dispatcher who takes the 911 or other calls along with an event code and the names of the officers assigned. For potentially relevant entries, the researcher pulled the full report file and typed coded information into a laptop computer. Only 113 events were found this way after 200 hours of coding: 64 of these events were fights involving 10 or more people.

From the University of Wisconsin Campus Police, we received a download of the log entries that had been given event codes of possible relevance. Assistants read paper or computerized event files for further information to determine whether the event had a public or

collective character. Working with the downloaded case numbers and event codes on a laptop computer, assistants were instructed to annotate each record to indicate the type of event it was and, if potentially relevant, to type in any information in the record about the details of the event. For 1994 and later, the downloaded records contained up to five lines of "disposition" text describing the event, making it possible to restrict the searching of reports only to those events appearing to have the potential to be some sort of collective event. This permitted us to identify, for example, a two-person protest outside a meeting in a building on the University of Wisconsin campus from among the 1,431 "check area" and "check person" calls that we downloaded. Eventually, we concluded that it would not be worth the effort to pull all the "check" event report files for earlier years that lacked disposition text in the downloads.

Not all the "potentially relevant" events in the University of Wisconsin Campus Police logs were searched for media coverage. Many of the collective events in the records were regular university athletic events. We excluded these events from the analysis, knowing that virtually all these events receive newspaper coverage and not wanting to spend resources studying sports coverage. We also excluded social dances at the student union because they were usually social events for particular groups of students on campus. However, in both cases, if the logbook indicated that a protest or collective disruption occurred at the event, the newspapers were searched for coverage.

From the Capitol Police Department, we received a download of all log entries categorized as "crowd control," "rally/riot," "disorder," "battery," or "bomb threats." Preliminary investigation revealed that many log entries about protest-relevant events had other incident codes, so the records officer also used the keywords "protest*," "rally," and "demonstra*" to search for other log entries and printed out the list of event numbers. The second author then used the Capitol Police Department computer to locate the full dispatch entries or reports for these events and print those pertinent to the study's aims, following the rule "when in doubt, keep the event." Some events were also described in longer reports. We were permitted to print out these longer logs and reports so they would be available for later coding; officers monitored this work to ensure that we were not obtaining unauthorized information. Except for the regular Saturday farmers'

market on the capitol grounds where no protests or speeches took place, all events located in this manner were searched for media coverage.

Overall, our search for protest records in the police logbooks was necessarily unsystematic and somewhat subjective. It appears likely that we missed some events of interest. Wading through the logbooks of three different police agencies has led us to doubt the feasibility of creating systematic, replicable procedures for identifying collective events in the logs of diverse police agencies in the United States. Because of the lack of UCR codes for the events of interest to social movements scholars, recording of such events is ad hoc and inconsistent. This might suggest that scholars should retreat into the more routine and systematic permit records. Unfortunately, such a retreat would lose information precisely about the disruptive edge of protests—the radical remnants and the smaller, more marginal groups who do not work with the authorities or obtain permits. The police logs turned up dozens of smaller protests that did not appear in permit records. We would not have learned about the full range of movement events if we used only the permits. This is especially important given that unpermitted protests are also less likely to be covered by the media.

IDENTIFYING EVENTS IN NEWSPAPERS

Once we completed a data set of collective events based on information provided in the police records, two Madison newspapers—the *Wisconsin State Journal* and the *Capital Times*—were searched for all articles, announcements, and stand-alone graphics (collectively referred to as media items) about collective events. The core media-searching strategy of the project involved following an event-specific protocol where keywords from a particular police record of an event were used to search the NEXIS database for an 18-month interval (6 months before and 12 months after the event). To provide an assessment of the effects of different search strategies, two additional media searches were conducted for May 1994. The first was "a full-text read" in which an assistant read all May 1994 issues of the two newspapers and photocopied any story judged to be about a collective event. The second was an electronic keyword search for the

May 1994 issues using generic event descriptors such as protest* and demonstra*.

Frequently, media items refer to an event in passing, mentioning only the name of the event or providing a vague allusion to the type of activity and the issue involved (i.e., "people protesting about welfare at the capitol last week"). In this study, we set minimum criteria for determining whether a particular news story or stand-alone graphic caption contained enough information to be considered to be referring to a particular event. Events were considered to have been mentioned specifically by a media item if the item (1) referred to the event by name or (2) referred to the type of activity, issue addressed, individual(s) involved, or organization involved along with a reference to the location and approximate date of the event.

For the present analysis, we recorded the total number of articles, stand-alone graphics, and event-relevant words. We measured the number of event-relevant words by counting the combined number of words in sentences containing information on the name of the event, the type of event, activities taking place during the event, responses to the event, location, issue(s) addressed, event date(s), or individuals/organizations participating in the event.

RESULTS

Through comparisons of different sources, different search methods, and different time intervals for searching our record of collective events taking place in Madison in May 1994, we explored some of our concerns regarding the validity of data used in event analyses to date. The results are categorized by our three main areas of concern: sources, searches, and timing.

SOURCES

Event selection. The best way to assess event selection structures of the media is to compare media sources with nonmedia sources. One should not infer from this, however, that the media are poorer data sources than those with which they are being compared. Besides ask-

ing what events recorded in police records were covered by the media, we can also ask what events covered by the media were also recorded by the police.

Before we compare police and media sources to determine their respective selection structures, we must decide what types of events to include in our study. Researchers have mainly limited protest event analyses primarily to rallies, marches, vigils, pickets, and leafleting. Since the 1970s, however, in the United States and Western Europe, these forms have become legal and even normative ways to express opinions. As a result of this routinization, the boundaries around "protest" have become permeable and fuzzy. Protests are often symbolic statements with important elite or institutional support, not disruptive challenges to public order. Some protest messages are delivered through nonprotest forms such as ceremonies, speeches, displays, or press conferences. As the standard protest forms become legal and normative, they can carry nonprotest educational or awareness content. In short, protest *forms* and protest *content* cannot be directly equated, nor can it be assumed that protests are disruptions of normal institutional processes. Foreseeing the need to expand event analyses to include less standard forms of expressing protest claims, we include other message forms as well as institutional message events. Other message forms include ceremonies, displays and exhibits, speeches, fund-raisers, petition drives, and celebrations. Institutional message events include bill signings; forums, workshops, and symposiums; community service events; large or political meetings; public hearings; large or political conferences; and press conferences.

Table 1 shows all of the message events that we identified in any police archive or in newspaper stories using every search methodology at our disposal. As the table indicates, different types of events varied in their representation in police records. Local newspapers covered many events not listed in police records. Organizations held conferences and forums to discuss issues, met with targeted actors to present demands, launched petition drives, announced campaigns at press conferences, sponsored fund-raising events to generate resources and publicity, and conducted ceremonies honoring past leaders and present members. Police records offered relatively little information on these nonprotest message events, although there were some such events that only police records mentioned.

TABLE 1: Coverage of Event Types in Police and Media Records

	Numbers				*Proportions*				
Type of Event	*Police Only*	*Media Only*	*Both*	*Total*	*Police Only*	*Media Only*	*Both*	*Police Total*	*Media Total*
Protest forms									
Demonstration or rally	3	1	4	8	0.38	0.13	0.50	0.88	0.63
March	1		1	2	0.50	0.00	0.50	1.00	0.50
Picket, vigil		1		1	0.00	1.00	0.00	0.00	1.00
Other message forms									
Ceremony (award, memorial, etc.)	6	18	5	29	0.21	0.62	0.17	0.38	0.79
Display, exhibit	3	5	2	10	0.30	0.50	0.20	0.50	0.70
Speech, Q&A		1	2	3	0.00	0.33	0.67	0.67	1.00
Fund-raiser	3	12	1	16	0.19	0.75	0.06	0.25	0.81
Petition drive		2		2	0.00	1.00	0.00	0.00	1.00
Celebrations		2		2	0.00	1.00	0.00	0.00	1.00
Institutional message events									
Bill signing		1		1	0.00	1.00	0.00	0.00	1.00
Forum, workshop, symposium		3		3	0.00	1.00	0.00	0.00	1.00
Community service	1	1		2	0.50	0.50	0.00	0.50	0.50
Large or political meeting		4		4	0.00	1.00	0.00	0.00	1.00
Public hearing		3		3	0.00	1.00	0.00	0.00	1.00
Large or political conference		4		4	0.00	1.00	0.00	0.00	1.00
Press conference		11	1	12	0.00	0.92	0.08	0.08	1.00
Summary of forms									
Protest forms	4	2	5	11	0.36	0.18	0.45	0.82	0.64
Other message forms	12	40	10	62	0.19	0.63	0.17	0.37	0.80
Institutional message events	1	27	1	29	0.03	0.93	0.03	0.07	0.97
Total	17	69	16	102	0.17	0.67	0.16	0.33	0.83

Despite the greater overall number of events recorded by the media, the police were just as likely as the newspapers to record protest forms and somewhat more likely to record information about demonstrations and rallies. This is because most protest events occurred at the state capitol, where the police records (especially permit records) are most comprehensive. As a result, the Madison Police Department's policy of not recording legal protests encountered on the streets did not result in the exclusion of a considerable number of protests.

Nonetheless, of the 11 cases of standard protest forms in our data set, police records failed to record 2 (18 percent). Both involved protests accompanying event forms not expressly convened for that purpose. In one instance, a small group of environmentalists protested Norwegian whaling outside of a university graduation ceremony where Norway's prime minister was the keynote speaker. In the other instance, a nurses' union, disgruntled over stalled contract talks, picketed the Wisconsin governor's reelection kickoff. Police records did not reveal either the kickoff or the picket. Similarly, only newspapers provided information on protest activities *within* event forms not expressly convened for that purpose. For instance, during May 1994, a series of public hearings took place in Madison regarding a proposal to ban certain assault weapons. Opponents filled the audience. Testimony often involved vivid symbolic resistance, including one man dressed up as a Nazi thanking the ban's legislative sponsors. Another speaker flashed two fluorescent colored guns concealed in his pants to make the point that most handguns can be hidden regardless of barrel length. The police records again failed to mention these activities.

In addition, the two newspapers were, on the whole, considerably more likely to cover other message forms such as ceremonies and fund-raisers as well as institutional message events such as political meetings and press conferences. Police records did not cover a large number of these events. For example, police records did not cover the following: a petition drive by the Parents-Teachers Organization (PTO) at a local elementary school protesting the rumored transfer of its principle, a fund-raiser for pro-choice Democrats, and a press conference held by two state legislators and 43 environmental groups calling for a special legislative session to consider anti-mining legislation. Our results suggest that researchers interested in small groups (such as the PTO) or the range of activities engaged in by larger movements

TABLE 2: Coverage of Event Attributes in Police and Media Records

	In Police Records		*In Media Records*		*Agreement*	
Attribute	*Frequency*	*Percentage*	*Frequency*	*Percentage*	*Frequency*	*Percentage*
Date	33 of 33	100	80 of 85	94	14 of 16	88
Location	33 of 33	100	77 of 85	91	14 of 15	93
Time of day	18 of 33	55	48 of 85	56	4 of 6	67
Size	8 of 33	24	37 of 85	44	1 of 1	100

(such as the pro-choice and environmental movements) would gain more useful information by searching the local newspapers rather than by searching police records.

Neither police records nor newspapers provided exhaustive inventories of protests or other message events. The incompleteness of both sources points to the obvious existence of the fourth cell in a two-by-two table—events recorded by *neither*.

Factors affecting event descriptions. It had been hoped that external police records could be used to check the accuracy of news accounts of events. Table 2 lists the rates of event attribute coverage by police and media sources as well as the levels of agreement in their descriptions. Both sources were likely to provide the date and location of events. Their rates of offering the time of day that the event occurred were also comparable. Newspapers were more likely to have information on the size of the event. Both record types provided more information on the location and date of events than on the time of day or size. With the possible exception of the time of day the event occurred, police and news records generally agreed when they both provided descriptive information on the same event attribute.

Factors predicting presence in police records. The factors predicting the presence of an event in police records were linked to the jurisdiction and the bureaucratic needs of the police agency maintaining the records. These practices can be seen in the patterns in Table 3. Location served as the major factor predicting whether an event was included in police records. At one extreme, all known events occurring outside on the capitol grounds were in the police records, as were all

TABLE 3: Location of Events and Presence in Police and Media Records

	Frequency	*Proportion*	*Proportion in Police Records*	*Proportion in Media Records*
Specific location				
Capitol square	13	0.13	1.00	0.54
Inside capitol	9	0.09	0.67	0.67
Inside other government building	7	0.07	0.43	0.57
Outside institution of higher learning	5	0.04	0.40	1.00
Inside institution of higher learning	10	0.10	0.30	1.00
Streets	2	0.02	1.00	0.00
Other outside	15	0.15	0.27	0.80
Other inside	39	0.38	0.00	1.00
Varied nongovernmental or no information	2	0.02	0.00	1.00
Total	102	1.00		
Location group[a]				
Inside, public	36	0.35	0.36	0.83
Outside, public	31	0.30	0.58	0.68
Inside, private or uncertain	30	0.29	0.00	1.00
Outside, private or uncertain	5	0.05	0.40	0.80
Total	102	0.99[b]		

a. Events that were both inside and outside are classed as outside. Events whose exact location is uncertain are classified as private.
b. Rounding errors.

events taking place in public streets. At the other extreme, none of the many events that occurred indoors in private spaces were in police records. In general, the police records were less likely to cover events taking place inside than outside and events taking place in private rather than public spaces.

Examining more closely the 59 cases where location did not perfectly determine the presence in a police record (e.g., public indoors and outdoors other than streets), we found further jurisdictional patterns (see Table 4). The largest events, involving more than 500 people, were likely to be found in both police and news records. It is here that, newsworthiness and policing concerns about possible disruption

TABLE 4: Appearance in Police and Media Records by Estimated Size, Day of Week, Time of Day, and Type of Event (excluding locations with perfect determination)

	Proportion of All Cases		*Proportion of Those With Media Records*		
	In Police Record	*In Media Record*	*Frequency*	*In Police Record*	*Frequency*
Size categories					
1-25 people	0.82	0.41	17	0.57	7
26-100 people	0.31	0.81	16	0.15	13
101-500 people	0.45	0.80	20	0.31	16
> 500 people	0.83	1.00	6	0.83	6
Total			59		42
Day of the week					
Monday	0.64	0.82	11	0.56	9
Tuesday	0.50	0.67	6	0.25	4
Wednesday	1.00	0.00	2		0
Thursday	0.40	0.60	5	0.00	3
Friday	0.64	0.79	14	0.55	11
Saturday	0.67	0.60	15	0.44	9
Sunday	0.00	1.00	4	0.00	4
Unknown	0.00	1.00	2	0.00	2
Total			59		42
Time of day					
Morning	0.45	0.73	11	0.25	8
Afternoon	0.64	0.91	11	0.60	10
Evening	0.00	1.00	2	0.00	2
All day	0.90	0.40	10	0.75	4
Unspecified	0.48	0.72	25	0.28	18
Total			59		42
Type of event					
Protest forms	0.82	0.64	11	0.71	7
Other message forms	0.59	0.68	37	0.40	25
Institutional	0.18	0.91	11	0.10	10
Total			59		42

coincide. The very smallest events, involving fewer than 26 people, were more likely to be found in police records than in the newspapers. These events were mostly exhibits or displays but also included ceremonies and disruptive protests. Police knew of the protests if they were called. All the displays and ceremonies in police records came from permit data. Thus, while small events were deemed unworthy of

coverage by the newspapers, legal considerations warranted their inclusion in police permit records.

Moderate-sized events (26-100 people) were much more likely to be found in newspapers than in police records. These events tended to occur in places where police had formal legal jurisdiction (i.e., in public places) but informally lacked jurisdiction because of the nature of the event. Newspapers were far more likely to cover institutional types of events than were police records. Even though a permit is required for every event occurring in the capitol, this rule did not always apply to routine institutional events, including a bill signing by the governor and press conferences by elected officials.

The patterns presented in Tables 3 and 4 enable us to specify how jurisdiction affects police records of public events. Few, if any, police records of events exist for events that take place indoors on private property unless the property owner requests police assistance. In our data set, these private spaces included hotels, churches, private residences and businesses, restaurants and clubs, a private college, and a mall. Apparently, then, it is not possible to use police records to provide any kind of check on the comprehensiveness of newspaper coverage of events occurring indoors on private property. Newspapers are by far the superior record source for such events, although newspapers may also be incomplete in their coverage of events in private places. At the other extreme, police records will be relatively comprehensive for locales where laws require police permission for all events—in Madison, public streets and state property. For such places, police records can be used to validate news coverage. Police records for other public places where peaceful assembly is lawful depend on the institutional practices of a particular agency. As a result, record keeping may vary greatly across time and across personnel. The generation of police records is likely to depend on assessments of the disruptive potential of an event by the police or by citizens who call the police. These assessments may differ somewhat from newspapers' assessments of newsworthiness. The police may record disruptive actions by relatively small numbers of people around an issue not deemed newsworthy by newspapers, while newspapers are more likely to record events around issues they deem important regardless of whether they are disruptive.

TABLE 5: Presence and Level of Media Coverage on Police-Recorded Events by Search Method (n = 33 events in police records)

All Event Types	*Any Media Coverage*		*Number of Articles*	*Number of Graphs*	*Number of Event-Relevant Words*
	Frequency	*Percentage*			
Search method					
Full-text reading of microfilm (May only)	11	33	20	8	4,976
Generic event descriptor NEXIS search (May only)	8	24	32	18	8,724
Event-specific protocol NEXIS search (May only)	12	36	40	16	12,246
Event-specific protocol NEXIS search (18 months)	15	45	61	26	17,829

SEARCHES

We compared the search yields of three different search methods: (1) a full-text read of the microfilm copies of *The Capital Times* and *The Wisconsin State Journal* for May 1994, (2) searching the electronic copies of these same newspapers in the NEXIS archive using generic event descriptor keywords, and (3) using our event-specific protocol to search the same NEXIS archive for events listed in police records.

We evaluate the relative merits of different search strategies in several ways. First, we consider their ability to locate news coverage of the specific set of events recorded in the police records. Table 5 indicates that the event-specific protocol was more effective than either of the other two using this criterion. Considering only stories printed in May, the event-specific search found newspaper mentions of 12 of the 33 events in the police records, while the full-text read found 11 and the generic event descriptors found 8 events. Correspondingly, the event-specific protocol overall found more articles and more event-related text. Perhaps because brief graphic captions are more likely to refer to events in generic terms, the generic event descriptor search yielded more graphics than the other two methods.

TABLE 6: Presence and Level of Media Coverage on Police-Recorded Events by Search Method and Type of Event (n = 33 events in police records)

	Any Media Coverage				
All Event Types	*Frequency*	*Percentage*	*Number of Articles*	*Number of Graphs*	*Number of Event-Relevant Words*
Protest forms (n = 9)					
Full-text reading of microfilm (May only)	3	33	6	2	1,564
Generic event descriptor NEXIS search (May only)	2	22	6	1	1,695
Event-specific protocol NEXIS search (May only)	3	33	7	2	2,242
Event-specific protocol NEXIS search (18 months)	5	56	12	5	5,121
Other message forms (n = 22)					
Full-text reading of microfilm (May only)	7	32	13	6	3,376
Generic event descriptor NEXIS search (May only)	5	23	23	14	6,636
Event-specific protocol NEXIS search (May only)	8	36	31	13	9,289
Event-specific protocol NEXIS search (18 months)	9	41	47	20	11,993
Institutional forms (n = 2)					
Full-text reading of microfilm	1	50	1	0	36
Generic event descriptor NEXIS search (May only)	1	50	3	3	393
Event-specific protocol NEXIS search (May only)	1	50	2	1	715
Event-specific protocol NEXIS search (18 months)	1	50	2	1	715

Table 6 compares the three search strategies for different event types. The event-specific protocol searches of coverage in May

captured more nonprotest message events than the other two methods. The efficacy of the event-specific protocol search, however, is offset by the underrepresentation of other message events in police sources upon which such searches rely (see Table 1). Full-text reads were just as likely to identify protest events. All three search methods were equally successful at locating institutional events.

In terms of the frequency and prominence of event coverage, the event-specific searches yielded the most articles and event-relevant words for all event types, except for institutional forms in which the generic event descriptor searches yielded more articles and event-related text. While the generic event descriptor strategy picked up more graphics than the other two methods for other message and institutional events, the other methods yielded more graphics for standard protest forms. For all three event types, the generic event descriptor searches yielded higher levels of event-related text than the full-text reading of microfilm.

Overall, the evidence suggests that if police records are to be compared with news records, an event-specific protocol search is most effective for finding news coverage of those events. At least for our news sources, reading the text of newspapers appeared to be equally adequate for collecting data for analyzing media selection of standard types of protest events, but it was less adequate for other event forms. We believe that the other message event forms are also important for mobilizing constituencies, expressing dissent, and generating pressure for social change. At a minimum, our data suggest the importance of checking search procedures with cross-validation pretesting, and we believe that it is most appropriate to use a search method that adequately captures a full range of message event forms.

Nonetheless, as we established above, newspapers reported on many more events than were present in the police records. It is appropriate, therefore, to compare full-text reads to generic descriptor searches for all the events located, not just those in the police records. Table 7 shows the search results for all events located in the newspapers. As compared with Table 5, these results give a more favorable impression of the generic descriptor searches. The generic event descriptor search captured coverage on 69 events compared with 55 events identified through reading fully the text of newspapers on microfilm. The generic descriptor search also captured a greater

TABLE 7: Comparison of Search Methods for Events Receiving News Coverage in May (*n* = 85)

All Event Types	*Frequency*	*Percentage*	*Number of Articles*	*Number of Graphs*	*Number of Event-Relevant Words*
Events in any media records	85	100			
Search method					
Full-text reading of microfilm (May only)	47	55	70	27	27,370
Generic event description NEXIS search (May only)	59	69	105	56	36,855
Ratio of GED to FTR	1.26		1.50	2.07	1.35
Event-specific protocol NEXIS search (May only)	12	14	40	16	12,246
Event-specific protocol NEXIS search (18 months)	15	18	61	26	17,829

NOTE: FTR = full-text read; GED = generic event descriptor search.

number of articles, graphics, and event-related words than reading microfilm.

Overall, we conclude that neither the full-text read nor the generic descriptor search is ideal. Of the 85 events located in the newspapers by either search method,[9] 27 percent were found by the full-text read and the generic event descriptor search, and 27 percent were found by the full-text read but not by the generic event descriptor search. Conversely, 46 percent were located by the generic descriptor search but not the full-text read. Thus, the full-text read found several events that were missed by the generic event descriptor search and vice versa.

Some stories describing collective events did not use generic descriptors. While especially acute for the large and diverse group of "other message events," this problem is also common for protest events. A few examples illustrate the problem of language and writing styles and the ways that a journalist can provide clear descriptions of protests without using generic event descriptors. In one instance, a group of activists protested the lack of consideration of bicyclists'

safety and access to Madison streets by taking up an entire lane of traffic during rush hour. One newspaper printed a picture of the bikers with a caption saying the following:

> Local bicycle advocates took to the road trying to block traffic Friday afternoon in an effort to show that bikes deserve respect and access to the streets. These pedalers managed to take up a whole lane. . . . ("Rush-Hour Ride"—graphic by Mike Devries appearing in *The Capital Times* on May 7, 1994, p. 3A)

Interestingly, the police record refers to this event as a "bike rally," a term used both for recreational biking events and for several protest-type events involving either bicycles or motorcycles. We were not sure whether this event was a protest form or a recreational event until we located the news item about it. In another instance, an editorial mentioned and supported a picket with a political message but did not use the words *picket*, *protest*, or *demonstration*:

> Hooters may bill itself as a "neighborhood restaurant," but we don't know of many corner diners where "trainers" tighten the T-shirts of waitresses to "make sure as much is showing as possible." And we don't know of any eateries that feel they have to call the police to keep concerned feminists away from their front doors—as Hooters did Monday. . . . Hooters may say it's all about owls. But it's really about sexual exploitation. ("Making Sleaze Mainstream"—staff editorial appearing in *The Capital Times* on May 10, 1994, p. 6A)

In summary, while generic event descriptor searches of newspapers capture more events and larger amounts of event coverage in the source, they consistently miss event coverage that describes events without reference to conventional terminology.

Search method selection structures. Beyond the terminology used in media items to describe events, are there other factors that make it more likely that one search method will identify coverage rather than the others? The selection structure of the event-specific protocol search (using both 1-month and 18-month intervals) mirrors that of the police records (see Tables 3 and 4). So, we focus on the differences between the full-text read and generic event descriptor methods. Table 8 presents information about the 81 events identified by one or both of

TABLE 8: Yields of Full-Text Read (FTR) and Generic Event Descriptor Search (GED) by Type of Event, Levels of Coverage, and Estimated Size

	FTR Only		*GED Only*		*Both*	
Type of Event	*Frequency*	*(Percentage)*	*Frequency*	*(Percentage)*	*Frequency*	*(Percentage)*
Protests	1	(20)			4	(80)
Other message	17	(35)	23	(48)	8	(17)
Institutional	4	(14)	13	(46)	11	(39)
Total	22	(27)	37	(46)	22	(27)
Levels of Coverage		*Mean*		*Mean*		*Mean*
Articles		1.27		1.36		4.17
Words		397		499		1,604
Graphics		0.55		0.92		1.65
		Mean Number of Words		*Mean Number of Words*		*Mean Number of Words*
Protests		114				1,160
Other message		364		473		1,578
Institutional		608		544		1,785
Size Categories	*Frequency*	*(Percentage)*	*Frequency*	*(Percentage)*	*Frequency*	*(Percentage)*
1-25	2	(18)	3	(27)	6	(55)
26-100	8	(26)	14	(45)	9	(29)
101-500	9	(29)	15	(48)	7	(23)
500+	3	(38)	4	(50)	1	(13)
Total	22	(27)	36	(44)	23	(28)

the search methods. The biggest factor predicting whether the full-text read or generic event descriptor methods found an event was the sheer amount of material available to be found. Events found by both methods on average had three times as many articles, three times as many words, and nearly twice as many graphics as those found by only one of them.

Guided by exploratory multivariate analyses not shown, we determined some clear patterns distinguishing full-text reads from generic descriptor searches for those events not captured by both. Although the small number of protest events identified by both search methods combined prevents anything but tentative conclusions, the two search methods appear to differ in terms of the type of events for which they were likely to yield coverage. The generic descriptor search was more likely to find other message and institutional events. Both methods appear about as likely to identify protests. To the extent that one can reach a conclusion based on five cases of protest, it does appear that analyses limited to standard protest forms lose little but also gain little by searching electronic media databases. Analyses inclusive of nonprotest forms frequently deployed by social movement organizations, however, appear to benefit significantly from electronic searches.

Location also affected the likelihood of different search methods capturing the same event (see Table 9). Net of other factors, the full-text read was better at locating events taking place outside, especially at the capitol. In contrast, the generic descriptor search was better at locating events taking place indoors (except at the capitol). In terms of the interaction between the type of event and its location, other message events at the capitol were captured more often by the full-text read, while other message events located elsewhere were identified more by the generic descriptor search.

The two searches also varied in their ability to identify events of different sizes. Neither method differed significantly in its ability to capture small events (fewer than 26 people) or large events (involving more than 500 people). Events of 26 to 500 people, however, were more often located by the generic event descriptor search. In general, then, the reading of microfilm copies of the two newspapers examined favored outdoor events occurring in central public locations. Searches using generic event descriptors as keywords were better at picking up

TABLE 9: Search Method Yields of Media-Recorded Events by Estimated Size, Day of Week, Time of Day, Type of Event, and Location

	Proportion of All Cases Located in Media Records (n = *85*)				
	Yielded by FTR	*Yielded by GED*	*Yielded by ESP (May)*	*Yielded by ESP (18 months)*	*All Media-Recorded Events*
Size categories					
1-25 people	0.73	0.82	0.45	0.45	11
26-100 people	0.55	0.70	0.03	0.06	33
101-500 people	0.49	0.67	0.06	0.12	33
> 500 people	0.63	0.63	0.50	0.50	8
Day of the week					
Monday	0.64	0.71	0.36	0.36	14
Tuesday	0.73	0.64	0.00	0.09	11
Wednesday	0.25	0.75	0.00	0.00	4
Thursday	0.56	0.89	0.00	0.00	9
Friday	0.50	0.69	0.25	0.38	16
Saturday	0.53	0.71	0.18	0.18	17
Sunday	0.38	0.63	0.00	0.00	8
Unspecified	0.67	0.50	0.00	0.00	3
Time of day					
Morning	0.67	0.56	0.11	0.22	9
Afternoon	0.69	0.62	0.38	0.46	13
Evening	0.56	0.75	0.00	0.00	16
All day	0.40	1.00	0.40	0.40	5
Unspecified	0.50	0.69	0.10	0.12	42
Type of event					
Protest forms	0.71	0.57	0.43	0.71	7
Other message forms	0.52	0.62	0.16	0.18	50
Institutional	0.57	0.86	0.04	0.04	28
Location group					
Inside, public	0.47	0.80	0.17	0.20	30
Outside, public	0.76	0.52	0.29	0.38	21
Inside, private or uncertain	0.43	0.73	0.00	0.00	30
Outside, private or uncertain	1.00	0.50	0.25	0.25	4

NOTE: FTR = full-text read; GED = generic event descriptor search; ESP = event-specific protocol.

moderate-sized, indoor, other message events, and institutional message events.

Aesthetic variations between sources. While differences in journalistic writing styles affect the effectiveness of generic event descriptor

searches, aesthetic differences *between* newspapers can result in a disproportionate number of events and event coverage being selected from more visually pleasing sources. To test this hypothesis, we examined the number of event-related articles and stand-alone graphics appearing in the *Capital Times* versus the *Wisconsin State Journal* located through the three different search methods. We expected that, because of the *Capital Times's* larger font size, larger point spacing, and smaller sized editions, the full-text read search method would capture items from the *Times* to a greater extent than the other techniques. Our results confirm this expectation: Of all the media coverage located by any method in the *Capital Times*, 49 percent was found in the full-text read, while only 22 percent of the total coverage located by any method in the *Wisconsin State Journal* was located in the full-text read. The possibility of such aesthetic effects on coder accuracy, therefore, should be evaluated in any comparisons between newspapers based on full-text reads. Searches of electronic databases appear less vulnerable to this problem, although generic event descriptor searches are affected by variations in the writing practices of different news organizations, as we discussed earlier.

TIMING

Above, we challenged the assumption that the relevant span for news coverage coincides with the date range for the events of interest. To provide some data on this point, we compared the results of an event-specific protocol search for items appearing in an 18-month interval around the date of the event (6 months before and 12 months after the event) with the same search for items appearing in the same month as the event. As Table 5 shows, 3 of the 33 events recorded in the police records received their only media coverage outside the month of May. Of these additional events, 2 were protest rallies.[10] As expected, searching for media coverage only in May disproportionately excluded events taking place during the first and last weeks of the month. Of the 3 protest and other message form events receiving coverage only outside the month of May, 1 took place in the first week of May and 2 took place in the last week of May. All such events taking place in the middle of the month received coverage that appeared in May.

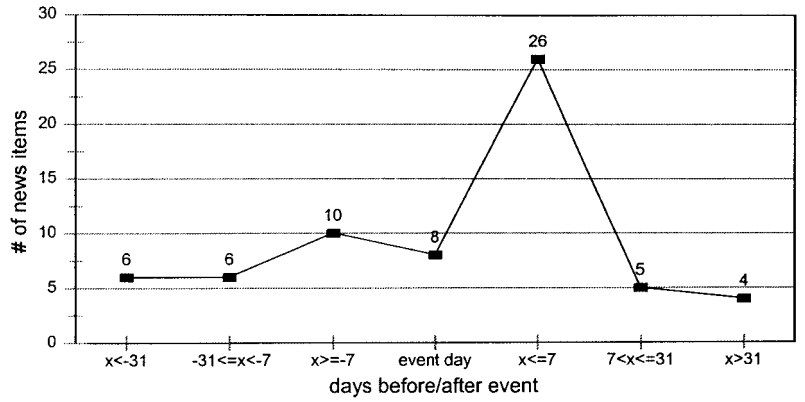

Figure 1: Distribution of Media Items on Protest and Other Message Events

Even for those events receiving media coverage during May, additional coverage could be found in other months. Of the 65 news items (articles or stand-alone graphics with captions) captured by the event-specific protocol search, 29 percent ran outside of May. Figure 1 presents the distribution of media items on protest and other message events captured using the event-specific protocol method. Almost one-third of the total number of media items appeared over a week before or after the date of the event. Of the 65 total items, more than one-third (26) appeared before the date of the event. Sometimes, the only media coverage appeared before the event happened, such as when a rally was held by students in support of a summer camp for children infected with HIV. The large number of items previewing events reflects efforts by event organizers to use the media as a tool for mobilizing potential participants. Overall, our results suggest that if researchers are going to rely on newspapers as data, they must stop equating the timing of news coverage with the timing of events and broaden newspaper searches to wider intervals around the period within which they wish to identify events.

CONCLUSION

There are no perfect records of collective events, nor are there perfect methods for gathering all of the collective events in any given

source. Each record source and data collection method has its own logic and selectivity. Nevertheless, our data offer some preliminary conclusions that can guide further research.

First, different police agencies and media sources cover distinctively different types of events. Reliance on only one source tends to weight the research toward some event types and against others. Focusing only on permit records eliminates information about many small unpermitted protests. Relying on official compilations of "protests" unquestioningly accepts an official definition of a "protest," which will vary between jurisdictions and is unlikely to coincide exactly with the researcher's intent. In the United States, official definitions of protest often neglect events taking place in privately owned spaces, disorderly events whose political meaning is contested (so they are categorized as simple disorder rather than protests), and orderly events that fall outside standard forms of protest. Moreover, our findings suggest that police records frequently omit other message and institutional events that, while not in the form of traditional protests, have come to play an increasingly important role in social movement mobilization. On the other hand, focusing only on the media makes it impossible to determine whether the media are ignoring certain classes of events. In particular, small events organized by "outsiders" who lack routine access to reporters will be missed in an analysis based on media sources alone. If one is interested in compiling the most detailed possible inventory of events occurring within a particular time-space boundary, particularly if one is especially interested in the activities of small or marginal groups, searching and merging multiple record sources is the best approach for gaining information on the most diverse sets of events.

Second, the information contained in police and media records varies by police jurisdiction and media source. In the Madison data, newspapers were more likely to have information on the size of the event. When they both provided descriptive information on the same event, police and news records generally agreed.

Third, researchers must avoid the naive assumption that "all" events of interest can be located in news sources through either reading microfilm or conducting keyword searches of electronic databases. This is the only research we are aware of that has systematically compared search strategies. Our work strongly suggests that all

methods find some events of interest and miss others. For searches of media sources for message events that can draw on external, nonmedia event records, an event-specific protocol search yields a greater number of events and more event coverage than other data collection methods. For searches that cannot draw on external sources for event information, the generic event descriptor searches proved relatively more adept in our data at picking up small events, indoor events, other message events, and institutional events, while full-text reads disproportionately captured large events, events at the state capitol, and outdoor events. It seems likely that the findings for other news sources would differ, but the general caution is warranted that full-text reads and electronic searches capture different subsets of the stories available to be captured in a news source.

Just as the full-text read and generic event descriptor search strategies possess their relative strengths, they also have their specific weaknesses. Because of its reliance on terms conventionally used to describe specific types of events, the generic event descriptor method frequently misses coverage not making recourse to conventional terminology. On the other hand, full-text reads miss event coverage occurring later in a news story and are more subject to aesthetic effects. So, the validity of any given data collection strategy needs to be cross-checked with alternate search strategies before its results can be taken as given.

Fourth, regardless of the strategy used, searches of media records for event data need to be grounded in a correct understanding of the distribution of media coverage around an event and the distribution of events through time. In our study, much of the event coverage appeared weeks, if not months, before and after an event. As a result, searches for media items that are limited to the interval during which the events take place will systematically underrepresent events taking place at the beginning and end of the periods of interest. The time interval of a media search, therefore, should be centered on each event and should be longer than the immediate period during which events of interest take place.

Given the small number of cases in our data set and variations in sources across locale, our findings should be viewed as suggestive rather than conclusive. Further research is needed to ensure that event

researchers look before they leap into previously untapped sources of data and newly available methods for searching them.

NOTES

1. Our own findings suggest that the *New York Times* covered less than 5 percent of Madison, Wisconsin-area message events as recorded in two local newspapers.

2. The authors refer to the following types of events collectively as demonstrations (p. 483, n. 11): "placarding, leafleting, petitioning, tabling, as well as picketing, vigiling, rallying, marching, chanting, gesturing, and the like."

3. Differences between electronic databases will also affect the number of event-related coverage captured and thus data validity. Potentially relevant considerations include differences in (1) what parts of the media sources are included (i.e., whether advertisements, graphics, letters to the editor, etc. are entered) and (2) features of the search engine (i.e., the ability to truncate words, use complex sequences of Boolean variables, specify dateline location, etc.). We leave a comparison of different databases for subsequent research and focus here on the formulation of search strategies when using NEXIS—the most widely available and commonly known electronic media database in the United States. The NEXIS search engine was significantly changed after we completed this data collection.

4. John McCarthy suggested this search string developed by a former graduate student of his, John Crist.

5. We say initial search strategy because stories yielded using generic event descriptive terms can, in turn, be used to generate event-specific search terms.

6. None of the newspapers used to construct the PRODAT data set had Sunday editions. Thus, Monday editions served as "next-day coverage" for events taking place on either the previous Saturday or Sunday.

7. It became clear to us that all agencies maintained intelligence records. These records, however, were not made available to the research team.

8. Although all agencies had a "riot" and/or "rally" incident code dating from prior decades, the codes had fallen into disuse, and none was consistently used for the events of interest.

9. The 18-month event-specific protocol search captured eight events not identified by either the full-text read or generic event descriptor search.

10. The other event was an exhibit.

REFERENCES

Arrendell, Charles. 1972. "Predicting the Completeness of Newspaper Election Coverage." *Journalism Quarterly* 49:290-95.

Beissinger, Mark R. 1998. "Event Analysis in Transitional Societies: Protest Mobilization in the Former Soviet Union." Pp. 284-316 in *Acts of Dissent: New Developments in the Study of Protest*, edited by Dieter Rucht, Ruud Koopmans, and Friedhelm Neidhardt. Berlin: Edition Sigmas Rainer Bohn Verlag.

Breed, Warren. 1955. "Social Control in the Newsroom: A Functional Analysis." *Social Forces* 33:326-55.

Danzger, M. Herbert. 1975. "Validating Conflict Data." *American Sociological Review* 40:570-84.

Eisinger, Peter K. 1973. "The Conditions of Protest Behavior in American Cities." *American Political Science Review* 67:11-28.

Everett, Kevin. 1992. "Professionalization and Protest: Changes in the Social Movement Sector, 1961-1983." *Social Forces* 70:957-76.

Gentile, Pierre. 1998. "Radical Right Protest in Switzerland." Pp. 199-226 in *Acts of Dissent: New Developments in the Study of Protest*, edited by Dieter Rucht, Ruud Koopmans, and Friedhelm Neidhardt. Berlin: Edition Sigmas Rainer Bohn Verlag.

Hertog, James K., and Douglas M. McLeod. 1995. *Anarchists Wreak Havoc in Downtown Minneapolis: A Multi-Level Study of Media Coverage of Radical Protest*. Journalism & Mass Communication Monographs No. 151. Columbia, SC: Association for Education in Journalism and Mass Communication.

Hocke, Peter. 1998. "Determining the Selection Bias in Local and National Newspaper Reports on Protest Events." Pp. 131-63 in *Acts of Dissent: New Developments in the Study of Protest*, edited by Dieter Rucht, Ruud Koopmans, and Friedhelm Neidhardt. Berlin: Edition Sigmas Rainer Bohn Verlag.

Jenkins, J. Craig, and Charles Perrow. 1977. "Insurgency of the Powerless: Farm Worker Movements (1946-1972)." *American Sociological Review* 42:249-68.

McAdam, Doug. 1982. *Political Process and the Development of Black Insurgency, 1930-1970*. Chicago: University of Chicago Press.

McCarthy, John D., Clark McPhail, and Jackie Smith. 1996. "Images of Protest: Dimensions of Selection Bias in Media Coverage of Washington Demonstrations, 1982 and 1991." *American Sociological Review* 61:478-99.

Oliver, Pamela E., and Daniel J. Myers. 1999. "How Events Enter the Public Sphere: Conflict, Location, and Sponsorship in Local Newspaper Coverage of Public Events." *American Journal of Sociology* 105:38-87.

Rosengren, Karl E. 1973. *International News: Methods, Data and Theory*. Lund, Sweden: Department of Sociology.

———. 1974. "International News: Methods, Data and Theory." *World Politics* 36:145-56.

Rucht, Dieter, Ruud Koopmans, and Friedhelm Neidhardt. 1998. "Introduction: Protest as a Subject of Empirical Research." Pp. 7-32 in *Acts of Dissent: New Developments in the Study of Protest*, edited by Dieter Rucht, Ruud Koopmans, and Friedhelm Neidhardt. Berlin: Edition Sigmas Rainer Bohn Verlag.

Rucht, Dieter, and Friedhelm Neidhardt. 1998. "Methodological Issues in Collecting Protest Event Data: Units of Analysis, Sources and Sampling, Coding Problems." Pp. 65-89 in *Acts of Dissent: New Developments in the Study of Protest*, edited by Dieter Rucht, Ruud Koopmans, and Friedhelm Neidhardt. Berlin: Edition Sigmas Rainer Bohn Verlag.

Sande, Oystein. 1971. "The Perception of Foreign News." *Journal of Peace Research* 8:221-37.

Sigelman, Lee. 1973. "Reporting the News: An Organizational Analysis." *American Journal of Sociology* 79:132-51.

Snyder, David, and William R. Kelly. 1977. "Conflict Intensity, Media Sensitivity and the Validity of Newspaper Data." *American Sociological Review* 42:105-23.

Weimann, Gabriel, and Hans-Bernd Brosius. 1991. "The Newsworthiness of International Terrorism." *Communication Research* 18:333-54.

White, David Manning. 1950. "The 'Gatekeeper': A Case Study in the Selection of News." *Journalism Quarterly* 27:383-90.

Wolfsfeld, Gadi. 1984. "The Symbiosis of Press and Protest: An Exchange Analysis." *Journalism Quarterly* 61:550-56.

Gregory M. Maney is an assistant professor of sociology at Hofstra University. His research interests include social movements, ethnic conflicts, and globalization. His work on transnational dimensions of social movements has recently appeared in Social Problems, Mobilization *and* Research in Social Movements, Conflicts and Change.

Pamela E. Oliver is professor of sociology at the University of Wisconsin–Madison. She has chaired the American Sociological Association's sections on collective behavior and social movements and political sociology.