



Using Media-Based Data in Studies of Politics

John T. Woolley

American Journal of Political Science, Vol. 44, No. 1 (Jan., 2000), 156-173.

Stable URL:

<http://links.jstor.org/sici?sici=0092-5853%28200001%2944%3A1%3C156%3AUMDISO%3E2.0.CO%3B2-%23>

American Journal of Political Science is currently published by Midwest Political Science Association.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/mpsa.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact jstor-info@umich.edu.

Using Media-Based Data in Studies of Politics

John T. Woolley University of California, Santa Barbara

Political scientists have made broad use of data that are drawn from media sources. These data tend to be of two kinds: measures of the volume of media attention to an issue, or counts of events reported in the media for which there is no other systematic data source. With increasing access to full-text data archives, there will be many more opportunities for scholars to create media-based data. This paper reviews a number of issues having to do with potential bias and error in such data. Some practical solutions are described, and a media-based data series having to do with monetary policy is used to illustrate how to probe for the validity of a media-based events count.

Political scientists have often used media reports—usually print media—as the basis for data on important aspects of politics and the policy process. Many scholars drawing on data sets for measures of coups, riots, demonstrations, elections, and other political events (e.g., Polity III and the *World Handbook of Political and Social Indicators*) are utilizing data originally drawn from newspaper accounts. We rely extensively on “media-based event counts.” With the growing availability of computerized text retrieval systems, one can anticipate even more use of such indicators in the future. It is a propitious moment to review the issues raised by these data.

We use media-based counts extensively, and we probably should do so more often, especially in studies of public policy. For example, many conceptions of the policy process suggest that changing beliefs, expectations, or information of critical actors may be associated with changes in policy outcomes and outputs. Mobilizing support often depends on generating favorable media attention. Media accounts provide information on these variables over time and over political jurisdictions.

I examine problems scholars confront when using media-based event counts. Scholars may easily go astray exploiting these opportunities, and using online sources does not appear to be a panacea. Many problems do not have a simple and direct solution; for others, additional empirical research may help. Here I draw broadly on research dealing with media data (see also Franzosi 1987 and Kaufman, Dykers, and Caldwell 1993). Although scholars in foreign policy and comparative politics have often examined the problems of media-based event counts, few scholars in American politics and public policy refer to that work. I shall also analyze a data series involving U.S. monetary policy, showing with a real example how scholars can probe the validity of media-based data.

Terms of Reference

The analytical problems discussed here involve a few essential elements (Schrodt 1994). The basic assumption is that some underlying process gen-

John T. Woolley is Professor of Political Science, University of California, Santa Barbara, Santa Barbara, CA 93106 (woolley@alishaw.sscf.ucsb.edu).

I gratefully acknowledge research support provided by the School of Business and the Center for Business-Government Relations, Georgetown University; the Academic Senate, University of California Santa Barbara; and the research assistance of Dirk Nowack and Rafe Slattery. I especially appreciate the helpful suggestions and comments of M. Kent Jennings, Lorraine McDonnell, and Peter Merkl.

American Journal of Political Science, Vol. 44, No. 1, January 2000, Pp. 156–173

©2000 by the Midwest Political Science Association

erates “real” events. The record for analysis—whether media reports, government documents, private papers, or prior scholarly work—is a “coding” of this universe of events. Compared to this universe, every record is necessarily incomplete. Each coding involves more or less explicit selection rules that may be biased and involve some error. Coding rules and error rates may be unstable over time. Below, I will use the term “record-coding” to indicate this process of creating the event record.

We rarely know with confidence the true universe of events, so it is difficult to identify precisely the selection mechanisms of a record-coding process. It is, however, often possible to compare two or more different codings of the same event stream and thus to identify some of the sources of bias and error in record-codings.

Scholars extract information from and further reduce and simplify (i.e., code) the available record. This “data coding” is in our control, and we may be able to correct for record-coding bias and error. While data coding may introduce errors and biases, in contrast to record coding the data-coding process is usually relatively well documented and examined for bias.¹

Commonly, there is another layer between the original record and scholarly data coding. This coding layer involves the creation of periodical indexes for nonscholarly purposes. Index record coding is another source of bias, error, and instability.

Two Common Kinds of Media-Based Data Studies of Media Focus

In studies of media focus, the variable of interest is media attention per unit of time—number of stories, number of column inches, amount of television time—per month or year. Typically this is taken as a measure of “the agenda” or perhaps more precisely, the media agenda or of the stimuli media present to citizens and elites (e.g., Brody 1991; Fan, 1994; Edwards, Mitchell, and Welch, 1995; Zaller 1992). These studies often rely on a single source, usually *The New York Times Index* or the *Reader's Guide to Periodical Literature*, and a reference to print media is often implicit. This approach characterizes a broad literature stimulated by Cohen's (1963) claim that

the media tell voters what to think about, but not what to think (Kosicki 1993; Gilens 1996).

Scholars in public policy have used measures of media focus to represent a wide variety of variables in addition to the media agenda. Sometimes media focus is taken to represent the “broad public agenda” or the “systemic agenda” (Flemming, Bohte, and Wood 1997). Measures of media focus have been interpreted as measuring “public concern” for a policy problem (Walker 1977); environmental group activity (Wood and Waterman 1993); and public awareness of a policy issue (Wood and Anderson 1993).

Typically, such studies have not devoted much time to discussing the measurement issues involved (exceptions include Baumgartner and Jones 1993; McAdam 1982; Hill, Hanna, and Shafqat 1997; Baumgartner, Jones, and MacLeod 1998). The “real event” is the concept the scholar is trying to measure, and the “coded record” is the original set of publications that reported or commented on the issue in question. Scholars have sometimes been imprecise about the underlying events of interest.

Event Count Studies

Less commonly, scholars in American politics use media sources as the basis for event counts when other sources are unavailable. Scholars simply count the frequency of events as reported in some coded record(s). An event is counted only one time, no matter how many stories are written about it. Such data have long been important in international relations and comparative politics, and substantial progress has been made in developing data on international interactions by using computer programs for text analysis (Schrodt and Gerner 1994; Schrodt, Davis, and Weddle 1994). Many characteristics of the event may be coded, providing a rich data set.

Students of American and comparative politics have used media accounts to construct data sets on strikes and violence, urban protests, riots, and urban civil rights conflict (see extensive citations in Franzosi 1987). In an important example, McAdam (1982) measured “black insurgency” by counting events reported in the *New York Times Index*. Rosenstone and Hansen (1993) used McAdam's data as the basis for a measure of political mobilization.

In studying U.S. monetary policy, Havrilesky (1993) constructed an event data set to measure “signaling” from the Administration to the Federal Reserve. This is a genuinely novel contribution to the study of monetary policy. It is a count of the number of *Wall Street Journal* reports of Administration dissatisfaction with current monetary policy. The series was carefully constructed

¹In one study, coders in different years, coding *the same source* with the same rules, agreed exactly on the frequency of coups only 54 percent of the time (Taylor and Jodice 1983, 195). The highest rates of agreement were for assassinations (100 percent), executions (92 percent), and failed coups (92 percent). Bond et al. 1997 suggest that correct codings (event recognition plus correct classification) in the 50–60 percent range are probably “good.”

and cross-checked by independent coders, is accessible to all researchers, and is easy to replicate. The basic source, *The Wall Street Journal*, is unmatched in the caliber of its financial reporting and the degree of its access to relevant policy-makers, lending special plausibility to the series. Significant work has already been based on these data (Havrilesky 1993; Froyen, Havrilesky, and Waud 1997; Chappell, Havrilesky, and McGregor, 1993). The main finding is of a systematic relationship between Administration signaling and future monetary policy. Signals for "ease" are followed by lower interest rates; signals for "tightness" are followed by higher interest rates.

Users of event counts should consider how biases and errors in record coding might affect their data. What events are underrepresented or overrepresented? Are these biases stable over time? To what extent does observed variation reflect media practices as opposed to the underlying event stream?

Media Counts: Bias and Error

Media-based event counts present a familiar epistemological problem: How do we know that we know the "fact" revealed through the record? Why did a newspaper publish a story on this topic? Why did it publish six such stories this month and only two last month? Does the different publication rate mean there were three times as many real events this month as last month? Could it mean that something external to the event series drew the newspaper's attention to these events more frequently this month than last month?

Record Coding: Source Biases

Merritt (1994) is typical in proposing to test the validity of an event series based on a particular record coding by comparing it with an event series from another source to see if they are highly correlated. This approach, however, might only indicate similar record-coding biases, so we need to probe further for biases. Doran, Pendley, and Antunes (1973) warn that even when media volumes are correlated over time, we need to consider the implications of large absolute differences in media volumes. Further, high correlations for some broad topic or issue may be consistent with substantial differences on more narrowly defined aspects of a topic.

Scholars assembling event data in international relations and comparative politics have invested considerable effort in comparing various print media to identify their coding biases and limitations (e.g., Sigler 1972; Doran, Pendley, and Antunes 1973; Burrowes 1974; Taylor and

Jodice 1983; Lichbach 1984; Howell and Barnes 1993). Among their conclusions are these: most media exhibit significant regional biases, disproportionately cover large urban areas or areas with wire service offices,² and report events with large numbers rather than small.³ Thus, more events will be reported for any institution or process that is covered by more reporters independent of the underlying frequency. This implies "censored samples" in addition to any ideological or substantive biases the media may have. When possible, scholars should use multiple sources to create event counts, striving to offset known or suspected biases. Prudence suggests using more than one highly specialized medium when possible (Burrowes, 1974; Buckman 1993).

Another conclusion is that "hard facts" are less subject to bias than are interpretations of the meaning of the event or the motives of participants (McAdam 1982; Franzosi 1987). This would include, for example, that a mass protest took place in some location, its relative size, and its degree of violence. Another conclusion is that more significant events—or at least large and violent events—are more likely to be reported and without bias (Snyder and Kelly 1977; Franzosi 1987; Lichbach 1984; Schrodtt 1994). Thus, although we are dealing with censored data, we have some basis for prior beliefs about the nature of bias and may be able to make some appropriate adjustments. These are positive conclusions for users of some event data, but not for scholars like Havrilesky who are dealing with neither hard facts nor public events of obvious importance. Nor can it reassure the users of media focus measures based on a narrow range of media or a single publication.

If these conclusions hold true, then for certain kinds of events, cross-record agreement and event-count correlations should be high. Those events are infrequent, irregular, and of great importance, such as "coups and earthquakes" (Taylor and Jodice 1983, 178; Rosenblum 1981) or "civil wars, coups d'état, and mass arrests" (Franzosi 1987).⁴

² Danzger (1975) shows that reporting on riots or civil rights conflicts reflects the geographical distribution of wire service offices. Snyder and Kelly (1977) disagree, arguing that above a certain size event, media report "intense," i.e., violent, events without bias. Put differently, the media are biased against reporting small and less violent events that may nonetheless be important and may provide variation necessary to test some theories.

³ These biases may not be stable over time. Riffe, Aust, and Lacy 1993 document a substantial drop in international news items in the *New York Times* over the period 1969–90.

⁴ This suggests higher correlations for conflictual than cooperative events in the Schrodtt-Gerner 1994 KEDS/WEIS comparison; they found the reverse. Rosenblum (1981) cites coups and earthquakes to show how reporting about poor countries is biased by emphasizing "bad" news.

TABLE 1 **Media Based Counts of Singular and Important Events**
Comparison of Categories of Events Occurring in Eleven Caribbean Countries,
1948-64 by Data Coding

Event	Total By Data Source			
	Doran ¹	Feierabend ²	Taylor & Jodice ³	Banks ⁴
Coups	66	53		
Successful			25	20
Unsuccessful			62	
Demonstrations	511	74	270	59
Assassination	117	26	105	28
Election	67	38	60	63*
Revolts	60	36		69
Riots			629	125

Notes:

1. Doran et al. (1973). Record sources: *Hispanic American Report*; *Tiempo*; *Vision*.

2. Feierabend and Feierabend, "Cross-National Data Bank of Political Instability Events," as reported in Doran et al.; Record sources: *Deadline Data* and *Yearbook of the Encyclopedia Britannica*.

3. Taylor and Jodice (1983). Record sources: *New York Times Index*; *Keesings Contemporary Archives*

4. Banks (1997). Record source: "Daily files of the *New York Times*"

*This is the sum of two variables, the number of legislative elections for the lower house held in a given year, and the additional number of changes in the effective executive when the executive is a directly elected president.

Surprisingly, however, scholars working with similar—or even identical—data coding rules but different record sources report very different frequencies for important events in the same set of countries for the same time periods. Some illustrative comparisons are presented in Table 1. Using the same data coding rules, Doran, Pendley, and Antunes (1973) report from one source a number of coups in the Caribbean region for 1948–64 that is 25 percent larger than another source. Banks (1997) and Taylor and Jodice (1983) disagree on one-third of the possible cases of successful coups in the same eleven Caribbean countries for the same period. They agree on eighteen coups; Banks identified two not coded by Taylor and Jodice, and Taylor and Jodice reported seven not coded by Banks. Taylor and Jodice (1983, 186) note other research showing a cross-country correlation of only 0.65 with their data on assassinations.

Entries in Table 1 show significant divergences between coders in the numbers of high visibility, significant events reported for the same sample countries and time period. These divergences undoubtedly reflect both record-coding biases and data-coding differences. Three of the sources report very similar frequencies for elections, events that may be relatively impervious to scholarly coder effects. Unfortunately, the data summarized in Table 1 do not support the conclusion that irregular, infrequent, and important events are reliably observed by most media.

A few studies involve important events for which a credible baseline of "real" events can be constructed.⁵ Gaddy and Tanjong (1986) show that the *New York Times* reports only about 20 percent of the serious earthquakes outside of the United States and with no apparent bias regarding developing countries. Adams (1986), Singer, Endreny, and Glassman (1991), and Keshishian (1997) report various forms of record bias in earthquake reporting.

Weimann and Winn (1994) show that the *New York Times* reported only 33 percent of world-wide terrorist events occurring between 1968 and 1980 and only 35 percent of events occurring in North America.⁶ More likely to be reported were larger, more violent events and events involving certain terrorist groups and certain targets (e.g., Palestinians and Israelis).⁷

In short, the best and most complete media sources involve biases and shortcomings even when relied upon to provide hard facts. This does not mean that bias is

⁵ In the case of price inflation, where verification is easy, I found no correlation between the annual inflation rate and the annual number of *Wall Street Journal* articles identified in an online newspaper index (Melvyl) search for the keyword "inflation" between 1982 and 1996.

⁶ Weimann and Winn use an authoritative compilation of terrorist events prepared by RAND.

⁷ Terrorism is media-oriented; that is, it relies on media coverage for its effects, as in case of black insurgency (McAdam 1982). Thus, these gaps in coverage are especially relevant.

fatal or cannot be detected. On the contrary, these reports reinforce the importance of thinking in terms of assembling a "media portfolio" as the basis of events coding.

More Evidence on Record Coding Bias and Cross-Source Agreement

The issue of record-coding bias arises also in studies of media focus. Many problems can be traced to the desire to measure a vague concept—"the media." Even if "media focus" is a meaningful true event, which measure is best or most relevant? In longitudinal studies, how can we accommodate dramatic changes in both the available media and citizens' use of media?

Scholars have made two generic claims to simplify research. (1) Different media sources tell the same agenda-setting story, and therefore the choice of a record coding is not terribly important. (2) Some media outlet (usually the *New York Times*) is a critical agenda setter for other media so we can safely, and perhaps correctly, focus on that outlet. If the second claim were true, then, logically, the first claim would also be true. The second claim cannot be true if the first is false. These claims are not sufficiently well validated to be generally accepted.

In considering these issues, we should distinguish the volume of media attention (or number of events) from the content of the media or characterization of the events. Agenda-setting effects probably depend on the consistency of media content. Further, if the agenda-setting effects of media attention are a nonlinear function of volume, or are subject to threshold effects, then large differences in volume between media may be important even if a high correlation suggests a similar kind of bias (Zhu 1992).

Baumgartner and Jones (1993, 258–259) argue that media focus measures based on the *New York Times Index* and the *Reader's Guide* are highly correlated. Therefore either index yields the same substantive conclusion about trends in media attention and timing of peaks in media attention (see also Patterson and Caldeira 1990; Terkildsen, Schnell, and Ling 1998).

A strong correlation over time in media attention is consistent with significant differences in content or volume. For example, in the case of tobacco smoke, the content of the *New York Times* was twice as positive overall (30 percent) as the *Reader's Guide* (15 percent; Baumgartner and Jones, 1993, 255). Hufker and Cavender's (1990) data on coverage of the "Mariel flotilla" in the *Washington Post* and *Los Angeles Times*, show that despite a strong correlation over time in the number of stories ($r = .98$), the *Post* had 45 percent more articles. There was zero correlation

between the two papers in the proportion of stories coded as "negative" in tone. Hertog, Finnegan, and Kahn (1994) report strong inter-media correlations in the volume of AIDS coverage, but large variation among newspapers in content (Table 1, 297). Gamson and Modigliani (1987) find significant differences between television and newsmagazines in the content of coverage of affirmative action over time. Gilens (1996) reports significant differences among newsmagazines in their depiction of poor people in the US.

Baumgartner and Jones's media volume correlations in the cases of child abuse and smoking are strongly influenced by single extreme years.⁸ Omitting that single extreme year substantially reduces the correlation between the volume measures. Aside from those peak years, it is not obvious that general trends are almost identical between *New York Times* and the *Reader's Guide* coverage volume. The first differences (or year to year changes) are not correlated at all. Thus, a study on a sample that happened *not* to include one of those critical years would yield results that vary according to the media indicator.

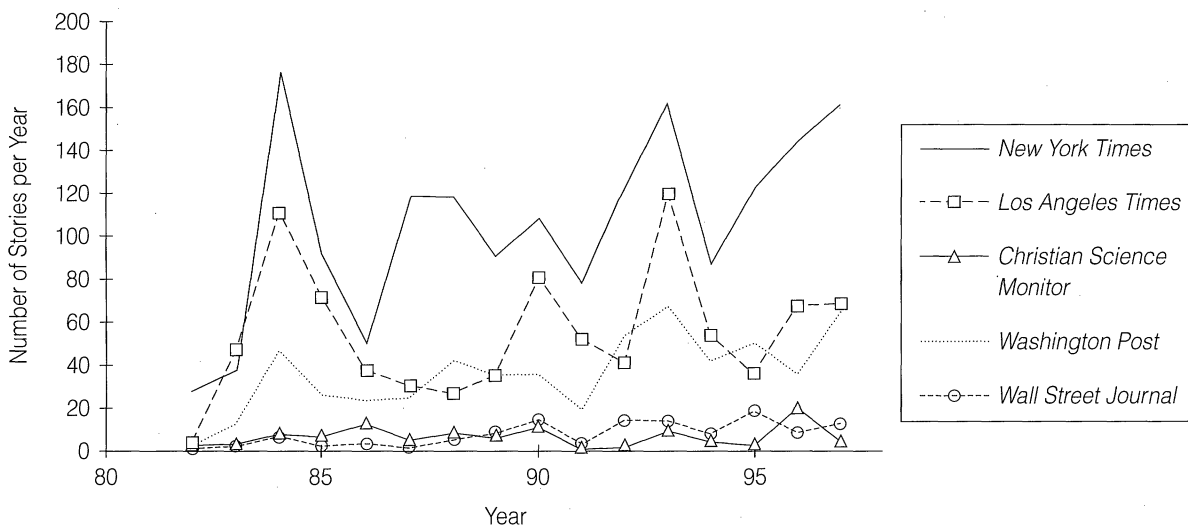
Evidence of discrepancies between and among media is apparent throughout the literature. Mazur shows significant differences between the *New York Times* and the *Reader's Guide* coverage in the 1980s (see Mazur 1990 Figure 1, 314). Some earlier works cited by scholars to support the notion of inter-media similarity are, upon reexamination, somewhat ambiguous. Research comparing congressional coverage in ten newspapers reported differences among papers in the volume of coverage of 2.5 times even excluding the *Washington Post* (Tidmarch and Pitney 1985).⁹ The frequency of original reporting on Congress differed by a factor of 4.9 between the paper with the least and the most coverage. There were "remarkable differences among newspapers" in the "focus" of coverage (Tidmarch and Pitney 1985, 472). Only in terms of the positive or negative tone of content was there convergence among the newspapers (474).

Robinson states that national news media, both television and newspapers, have become increasingly hostile in coverage of Congress. However, he also writes that the local and national presses have become more separate and distinct. Indeed, they "are two separate worlds" (1981, 85, 88).

⁸ 1984 for child abuse and 1964 for smoking; see Figure 8.4 and Figure A.1, respectively. Visual inspection of plots in Schrodt and Gerner (1994) suggests that some of their correlations may also be affected by influential observations.

⁹ The *Washington Post* covers Congress more than any other paper in the country.

FIGURE 1 Media Attention to Child Abuse in Five National Newspapers, 1982–1997; Divergent Trends, Peaks, Variability



Source: Information Access Company, Melvyl "News" Database, University of California; Number of items returned in Keyword Search for "Child Abuse."

To provide an additional test of these claims with an electronic database, I ran a keyword search to generate citation frequencies for articles on "child abuse" from the University of California's Melvyl catalog "News" database, an index covering five national newspapers for the period 1982–97.¹⁰ The results appear in Figure 1. This analysis assumes equivalent index-coding biases across these publications.

These data show that we would *not* draw virtually identical conclusions from different national newspapers. The bivariate correlations between these series vary between 0.12 and 0.84 with a mean correlation of 0.47. With five different newspapers, there are four different conclusions about the year in which attention to child abuse "peaked." The relative size of the peak, compared to the mean, varies between 1.6 and 3.0. Our conclusions about the shape of the trend and about variability around the trend depend on the newspaper as well.

In summary, scholars have advanced two general claims to justify simplified sampling of media sources. The more general claim is that different media measures yield the same agenda-setting story; the other is that one media outlet is the critical agenda-setter for the other. The two are logically linked; and the more general claim appears not to be true in many cases, implying that the narrower claim is not true either. It is likely that the no-

tion of "the media" is too vaguely defined to be useful in research, especially given the degree of change over time in the composition of mass media.

Long-Term Trends in Media Selection Bias

Baumgartner and Jones (1993) coded the "tone" of media stories to reveal reporting favorable or hostile to the private interests involved. Their work reveals a dramatic change in media content (to emphasize health issues) starting in the 1950s. Moreover, there appears to be an association between negative tone and the volume of media attention.

Why this change occurred is an important research question. The relative frequency of negative reporting may itself reflect a changing media selection bias, one that Starobin (1995) has called the "new cynicism." Such a development is apparently not peculiar to the U.S. (see Westerståhl and Johansson 1986 on Sweden).

Several studies document a recent tendency of newspapers to over-report certain kinds of scientific results. Koren and Klein (1991) studied newspaper coverage of two contrasting articles published together in the *Journal of the American Medical Association*.¹¹ One showed that nuclear workers had an increased rate of cancer ("positive" results); the other showed that residents near

¹⁰New York Times, The Washington Post, The Los Angeles Times, The Wall Street Journal, and The Christian Science Monitor.

¹¹Also see letters concerning this article in *JAMA* February 19, 1992, 930.

nuclear power plants had no increase in cancer rates (“negative” results). In an analysis of nineteen newspaper articles reporting on the studies, Koren and Klein found that the positive results alone were reported in nine cases; and the other ten reviewed both studies. No newspaper reported the negative results alone. Moreover, within the latter ten, the positive results received nearly twice as much attention (column inches); and coders judged the descriptions of the positive results to be significantly clearer and more accurate.

Similar selection effects seem evident in studies on the long-term effects of spanking on child discipline (Rosellini 1998, Gilbert 1997) and on the link between alcohol use and breast cancer (Houn et al. 1995). These findings are consistent with Entwistle’s (1995) description of biases of medical journalists and with Rothman and Lichter’s (1987) research on biases of journalists reporting on nuclear power.

An implication of these selection effects is that changing media volumes reflect journalistic practices but may not be an accurate reflection of underlying expert judgment or the tone of elite understanding. These selection effects may distort and amplify some events in politically significant ways. If so, we need to understand better why journalistic practices changed in the post-war period and with what effect.

Index Record Coding and Sources of Index Biases

For research of any historical depth, scholars usually rely on periodical indexes in constructing media-based measures. While we know that these indexes reflect current ideas of relevance and topicality, we know little about the coding rules used in their construction.¹²

Taylor and Jodice (1983, 185), using the *New York Times Index*, identified only 29 percent of the events found using the entire *New York Times* daily edition. The drop-off varied across event types and countries; it was particularly severe for strikes, riots, demonstrations, and government sanctions. In a recent paper, Althaus, Edy, and Phalen (1998) find that the *New York Times Index* served as a good proxy for the underlying print edition concerning the 1986 Libya Crisis but *only* for work at a high level of aggregation.

Index coding rules may be sources of bias; and, probably more worrisome, they are unstable over time.

¹² *Reader’s Guide* coding rules are not publicly available; articles are indexed only if “one column in length or greater; B. Chen, H. W. Wilson Co., personal communication July 30, 1998. Also, Lawler 1950.

This helps focus attention on a critical conceptual issue: what is the policy “problem” or “issue”? When the volume of articles changes in the periodical index categories—categories we have selected *because* they have come to be politically important—what have we learned?¹³ Were there equally large changes in areas that did not become politically important? Was the change in volume due to changing indexing practices, changes in problem definition, or merely the rise or fall of some dramatic event?¹⁴ When a topic heading disappears, what is the significance of that change, and where did the articles on this topic go?

The potential for error is larger for studies of media focus than for event counts. The longer the time frame, the more certain that indexing practices have changed during the period studied. Categories come and go; subdivisions emerge and fall away. Even over the short run, Gilens (1996) found inconsistencies in *Reader’s Guide* cross-referencing practices. As issues arise, the indexes create specialized topic headings that may affect overall attention measures.

One approach to using indexes is to try to include all possible index entries related to an issue, a reasonable strategy. This approach depends heavily on whether appropriate indexing categories exist. It involves the risk of potentially misconstruing the volume of attention to a policy issue (e.g., smoking) by including many articles that may be peripheral to the concept of a “policy agenda” (e.g., crop reports, corporate news) or a “public” issue.

A better, but more costly, approach would be to define independent coding rules for identifying relevant media content. Published index categories would be only a starting point. The case of child abuse illustrates the importance of this issue in the Baumgartner and Jones data. The *Reader’s Guide* included no separate heading for child abuse prior to 1959, so relying on index headings alone would have led to misleading conclusions about the agenda status of the problem of child abuse. Fortunately, Baumgartner and Jones were also using a second index—the *New York Times*—which showed that there was substantial attention to this problem in the 1920s. Scholars must be aggressive in searching for relevant articles (not just topic headings), and must develop

¹³ Changing indexing practices may be evidence about changing problem definition. “Water pollution” appears as a *New York Times Index* category in 1924. From 1921 to 1924, the relevant topic is “Harbor and River Pollution.” Before 1921, the topic is “Oil on Beaches” or “Oil from Steamships.”

¹⁴ Lawler (1950, 102) writes that the *Guide* selects index headings to reflect “common usage” (1950, 102). See also his discussion of changing headings and heading categories, pages 104–106.

coding rules for identifying relevant articles independent of the category judgments of the indexing source.

Ironically, McAdam, who was creating an event count, adopted a rigid strategy of examining only certain headings in the *New York Times Index*. The logic of constructing an event count calls for consulting *any* possibly relevant index subject heading and then coding identifiably separate events. While McAdam had elaborate coding rules to identify relevant events, his event count was inherently vulnerable to shifting indexing practices—likely when events suddenly assume greater prominence. For example, in 1968, McAdam's procedure appears to have ignored relevant events indexed under the headings of specific "movement" organizations and the heading "urban riots."¹⁵ In 1972, his procedure ignored all of these topics plus a very large number of stories indexed under "Black Panther Party."

Second Order Record Coding and Composition Instability

One might assume that composite indexes such as the *Reader's Guide* are more representative of concerns on the public agenda than are measures based on narrower media samples. However, such aggregate indexes involve instability due to composition effects.

Unlike the *New York Times Index* or any data set based on a single periodical (such as Havrilesky's index), the composition of the journals included in the *Reader's Guide* has changed dramatically through time. Because of these changes the *Guide* is not a clear reflection of discourse in society or among elites.¹⁶ Some apparent change in media focus revealed by periodical indexes is due to composition effects, unrelated to any real shifts in media focus.

Often the magnitude of change is not immediately apparent. For example, the 1953 *Reader's Guide* indexed a few more journals in total than the previous issue. However, starting in April 1953, the *Reader's Guide* shifted fifteen scholarly journals from the *Reader's Guide* to the *International Index*, later renamed the *Social Sciences and Humanities Index*. Fifteen other periodicals were dropped at the same time; each had a fairly specialized or quasi-

academic audience.¹⁷ Together, these journals represented 25 percent of publications in the *Reader's Guide*. The items dropped were replaced with nonscholarly, generalist magazines.¹⁸ The thirty-two "new" journals were not new; the *median* initial date of publication ("volume 1") was 1910.

In 1961 and again in 1968, there was a large net expansion in the number of journals indexed by the *Reader's Guide*. As before, the titles added were not new magazines that had just entered popular or elite discourse. Of the magazines added to the index in 1961, the *median* initial publication date was 1938. In short, in historical work, the changed composition and expanded volume of agenda discourse suggested by *Reader's Guide* entries is partly an artifact of indexing practices. From 1951 to 1961, many index entries representing specialist and scholarly articles were replaced by a larger set of entries for more popularly oriented, market-driven content.

Two possible solutions may address the composition problems of the *Reader's Guide*. First, construct a relevant sample of journals or journal types to provide more controlled continuity in journal titles or content over time. This is feasible for several major newsweeklies since the mid-1930s (see, e.g., Terkildsen 1998, Zaller 1992, and Gilens 1996). Even for a longer time period this may not be far-fetched.¹⁹

Second, seek relative proportionality between circulation of the journals studied and total journal circulation.²⁰ This could involve the "top twenty," or periodicals accounting for a constant share of total circulation. This requires characterizing magazine circulation in the United States (or elsewhere). UNESCO is recognized as a source for comparative statistics on periodical circulation at a very broad definition of periodical, but the data are not consistent across time or countries. The Audit Bureau of Circulation is the authoritative source on US periodical circulation since the early 1900s (Taft 1982), but their data are not readily available in university libraries (although subscriptions are not very expensive).

¹⁵ Omitted Groups: NAACP, SCLC, and SNCC. McAdam also excluded the topic heading "housing discrimination." McAdam traces the 1968 riots to black insurgency in chapter 8.

¹⁶ Indexed periodicals "are selected by a poll of the subscribers" based on their judgment of the "reference value" of the periodical. "... [C]ertain highly popular periodicals with low reference value have tried for years to be elected but without success" (Lawler 1950, 118).

¹⁷ The list of journals dropped is available upon request to the author.

¹⁸ The list of journals added is available upon request to the author.

¹⁹ Possible examples with initial publication dates: *Harper's*, 1850; *The Atlantic* 1857; *Scientific American*, 1845; *The Nation*, 1865; *The Saturday Evening Post* 1839–1971.

²⁰ A related aspiration would be a "media indicator" weighted by the reported reliance of people on different sources for news. Reported rates of magazine use depend on the question asked. Gilens (1996) claims 20 percent "regular" magazine readers; Mayer (1993) reports only 4–6 percent get "most of their news" from magazines.

Online Data Base Searches as an Alternative to Print Indexes

Because of the unknown quality and instability of the secondary coding process, significant benefits may flow from the control over coding afforded the analyst by searching an electronic database. However, scholars are well advised to approach such projects with caution.

What is the relationship between the online database and the original published record of interest? As news is increasingly disseminated online, how might the "record" be defined? How does the output from an online search compare to the published contents of periodical indexes?

Recent research suggests that searches of "encyclopedic" databases like LEXIS/NEXIS are likely to produce different counts than would be produced "by hand" (Kaufman, Dykers, and Caldwell 1993; Snider and Janda 1998). This happens in part because many online databases differ from the published record, often in ways not well documented. Some newspapers send to the databases only stories produced by their own staff or freelancers; some send all stories published. Some do not include in the databases wire service stories carried without revision by the newspaper. Some do not send syndicated columns. Newspapers do not always consistently store the same edition with online databases. Wagers (1992) reports that same-subject searches on related data bases (DowQuest and Dow Jones News) retrieved the same sources but different articles.

Further, arriving at a well crafted and productive set of automated search instructions is by no means easy (Schrodt 1994). Substantial variation in online search results can be attributed to different search strategies (Saracevic and Kantor 1991).

Consider this illustrative search. The issue is child abuse coverage in the *Washington Post* in 1995. The search compares the print edition of the *Washington Post Index* (hereafter WPI); an online keyword index available through the University of California Library System (the "Melvyl" "News" database); the online archives of the *Washington Post* accessed through the *Post's* web site; and the publication archives of Dow Jones Interactive Publications Library, accessible to online subscribers of the *Wall Street Journal*.

First round results for the WPI produced 111 items indexed under the subject words "child abuse, child neglect." The initial online search, using the keyword "child abuse" (Melvyl archive) or the phrase "child abuse" (in headline or lead paragraph; text archive), produced forty-eight articles in Melvyl, forty-nine in Dow Jones, and twenty-three in the Washington Post Archives. All

items in the Washington Post Archives were also in the Dow Jones results—and I consequently focused on Dow Jones. Why the Dow Jones results were broader is not clear, but evidence suggests that different editions were archived at the two sites.

In my judgment, none of the WPI articles were misclassified—they all dealt substantially with child abuse—but some misclassification (about 10 percent) was apparent in Melvyl and even more (nearly 20 percent) in Dow Jones. More striking was the low rate of overlap between the WPI and the electronic searches. Only thirty-four identical items were identified by both the WPI and the Melvyl searches, and only eleven by both the Dow Jones and Melvyl searches. A large proportion of the Dow Jones items were quite short—fewer than 200 words. It appears that a coding rule for the WPI is to exclude most short items (or roughly 5 column inches). This kind of implicit rule of "importance" or visibility may be a virtue from the point of view of agenda-setting studies, but is a defect for scholars constructing event counts. A large number of relevant items were missed in both strategies, but the published print index provided the best combination of identified relevant articles and fewest omissions.

Even with a much more complicated Boolean search command, deliberately constructed with knowledge of the WPI results, the outcome was very mixed. The final Dow Jones search produced 148 articles, approximately triple the initial yield, of which thirty-six were short articles. The overlap with the print index quadrupled to forty-four articles (still only 40 percent of the WPI total). The Dow Jones search identified thirty-four (often relatively short) articles not included in the print index. Together, two alternative electronic searches, Melvyl and Dow Jones, identified fifty-two relevant articles longer than 200 words that were not in the WPI. That is, the print index *understated* the relevant media volume by at least 30 percent. Overall, even with a more precise set of search commands, the Dow Jones search produced less than half the number of long articles reported in the print index, and 29 percent of the total hits were errors.

This experience suggests thinking of online searching as a two-stage process. The first stage involves significant effort in the crafting of search terms. Having as a criterion a sample "by hand" content coding, such as a published index, is essential. The second stage involves careful evaluation of the search results using content coding rules previously specified.

In short, using automated searches to measure media focus does not assure us of more precise or more complete results than would be achieved relying on commercially produced print indexes. For many purposes, a print index, itself based on a direct, and obviously imper-

fect, reading of the underlying record, may be more efficient than online searching.

For scholars constructing an event count, automated searching may, with effort, bring improvements over print indexes. As Schrodtt stresses, this is not about saving time; rather it is about gaining precision and direct control over the data coding process. With a very large record coding, such as a complete wire service (Reuters or the AP), and diligence to assure against double counting of events due to marginal rewrites of stories (Schrodtt, Davis, and Weddle 1994), the process of events coding may well benefit from automation. Ideally, scholars would have several entire databases available electronically and would be able to search for any possibly relevant event that would subsequently be coded according to some clear coding rule.

Adjusting Measures of Media Focus across Time

Over time (see Figure 2), the number of annual pages in the *Reader's Guide* and in the *New York Times Index* has varied dramatically. Indeed, the number of *Reader's Guide* index pages *per magazine per year* has more than doubled. If we assume no increase in the rate of double counting in the indexing phase, this implies many more entries—articles—per magazine publication. Over time, the *Reader's Guide* has indexed many more magazines (sixty-seven in 1900, 205 in 1989). The number of pages in the *New York Times* index has ranged from a low of 164 to a high of 3356, so that a media count of twenty-five articles in one year could have a profoundly different implications from an equivalent number of articles in another year.²¹ Accordingly, counts based on indexes should be deflated, reported as a ratio to the total number of articles indexed or total number of index pages. This adjustment would not address any effects that may arise due to changing numbers of entries per index page, type faces or page size.

The effects of deflating citation data can be seen in data on media attention to nuclear power in Baumgartner and Jones (1983, 266) and Weart (1988). The deflated and unadjusted data do show peaks of media attention in the same years. However, the deflated data show that media attention to civilian nuclear power in 1955 was roughly three times as great as 1905; the unadjusted data show that

attention in 1955 was over nine times as great as in 1905. The two ways of presenting the data suggest different issue dynamics and highlight the relevance of possible threshold effects associated with changes in media focus.

The effect of deflating the data is further illustrated in Figure 3 for the case of auto safety. The unadjusted data show wild swings over time around a flat trend, while the deflated data indicate declining volatility around a declining trend. Similarly, the data in Hertog, Finnegan, and Kahn (1994) show that *Medline* counts for “cancer” research increase 80 percent from 1980 to 1990 without volume deflation, but only 16 percent with volume deflation. One measurement suggests a large change in agenda status, the other a modest one.

Another measurement approach would be to examine media volume relative to some critical peak, or relative to the recent “normal” levels such as a moving average. A possible measure of a critical peak is the attention given to the “top story” (or stories) of the year as identified in the annual Associated Press poll of editors, dating from 1935. Then we could compare the volume of media attention to any issue to the volume allocated to its primary competitor(s).²²

Validation of an Event Data Series

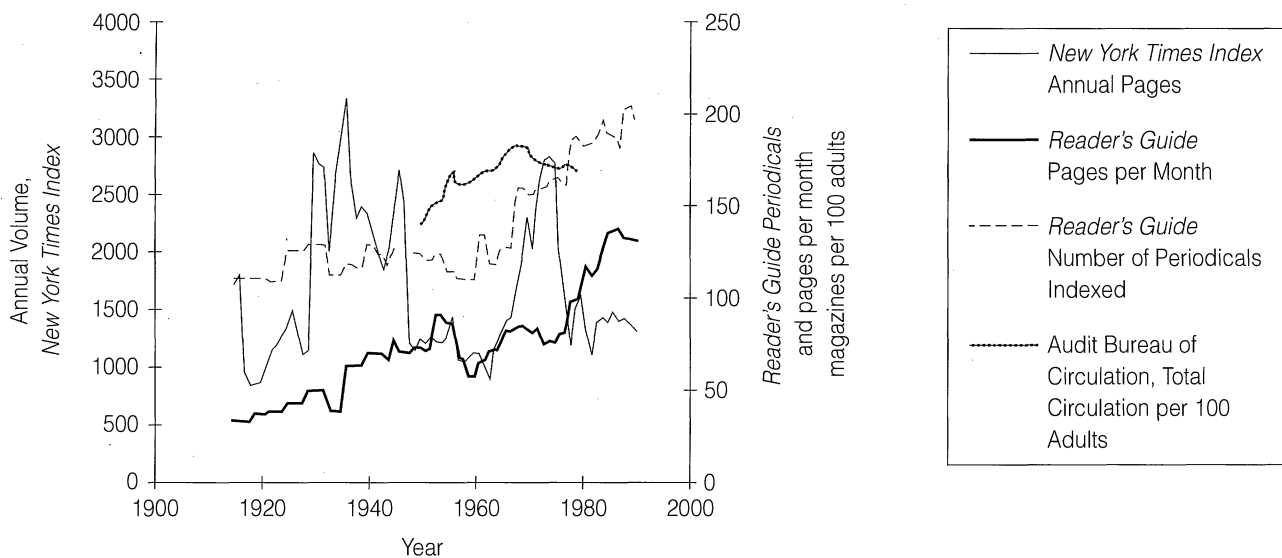
It would be good to have a means of validating an event data series other than by comparison to another media-based event series. The problem is to devise plausible benchmarks. The consistency of the data themselves with scholarly hypotheses is not necessarily a satisfactory test. Reporters' story lines may echo scholarly hypotheses.

In the instance of Havrilesky's Federal Reserve data, a plausible scholarly hypothesis might be that Democratic administrations are more likely to press for easier monetary policy in their first year in office and that Republicans are more likely to press for tighter policy. Reporters writing on economics probably share this expectation. Thus, reporters may, consciously or not, tend to look for evidence consistent with these two hypotheses and to write stories consistent with them. Out of the confusion of immediate events, it is likely that evidence can be discovered to support such stories. These stories have elements that make them more likely to be published than many alternative stories (especially the “frame” of conflict), and their consistency with received wisdom may enhance their plausibility. Is there objectively conflict of a

²¹ Other events affect publication volumes, too. Due to a strike, the *New York Times* was not published August 10, 1978 to November 5, 1978.

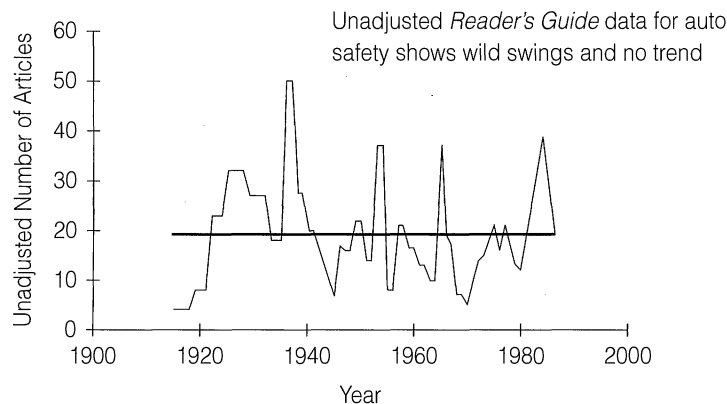
²² For example, in 1991 the top story was the Persian Gulf War with over 1800 *Washington Post* stories, 300 of them on the front page.

FIGURE 2 Variation Over Time in Total Volume, *New York Times* Index, *Reader's Guide to Periodical Literature*

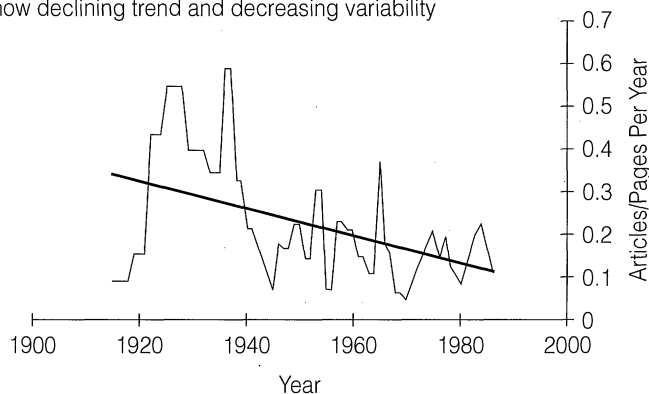


Source: Author's calculations, Baumgartner and Jones (1993); Taft (1982)

FIGURE 3 Deflating Article Counts Relative to Total Volumes Can Substantially Affect the Interpretation of Trends and Variability



Deflated *Reader's Guide* auto safety data show declining trend and decreasing variability



Source: Author's calculations, Baumgartner and Jones (1993)

particular sort with the Fed when a new Democratic or Republican administration takes office? As we shall see, perhaps so, perhaps not.

We must try to find other ways to evaluate such media-based event series. I illustrate three such ways here. The first, the internal logic of the event stream, notes that many event processes suggest related hypotheses that we can use to test the plausibility of media-based event data. The second test involves searching for evidence that characteristics of the news-production process may have influenced the reporting of events. The final test involves finding some nonmedia-based, preferably qualitatively superior, record of the same events that can be used to evaluate at least parts of the media-based series.

Internal Logic of the Event Stream

We can use information or theory about the logic of the event stream to evaluate the plausibility of the media-based count. Naturally, this logic will vary according to event stream.

Signaling is a process of persuasive communications, but the development of the Havrilesky data appear not to draw on any communications theory. Drawing from scholarship about persuasive communications (e.g., Reardon 1991; Stiff 1994), we can predict some patterns and effects of signaling for evaluating the validity of the Havrilesky data.

For example, the effect on monetary policy of clear signals should be different from that of mixed or conflicting signals. When an Administration gives conflicting signals, the effect on policy should be smaller than an equivalent volume of signals consistently indicating a preference about the direction of policy.

Further, the response to signals should attenuate as a function of the time elapsed since the signal. Success of an attempt at persuasion depends on the ability of recipients to respond. Thus, administration signals should coincide with key meetings—e.g., Federal Open Market Committee (FOMC) meetings, or meetings of the Fed Chairman with the President.

Finally, we should observe declining marginal efficacy of signals. For a given time period, each additional signal should produce a smaller effect than the first. Otherwise, there would be every incentive for the Administration to signal in huge volumes whenever it desired a particular outcome.

Havrilesky's series is largely inconsistent with these expectations. Consider the distinction between consistent and inconsistent signals. One can partition the Havrilesky data into two different weekly series—weeks with no contradictory signals (NOCONSIG) and weeks

with contradictory signals (CONSIG).²³ Then, regress the change in the weekly average federal funds rate on the two series—with and without conflicting signals. The results for January 1966 through December 1991²⁴ are as follows:

$$\text{DFFR} = 0.0060 - 0.0615 \text{ CONSIG} - 0.059 \text{ NOCONSIG}$$

(3.090) (2.823)

$$R^2: 0.0145; \text{df: } 1345 \text{ (t-test in parenthesis)}$$

The reaction of policy to the two different kinds of signals is statistically different from zero and undifferentiated one from the other. The results are very close to Havrilesky's in coefficient magnitude and overall goodness of fit. Contrary to theory, the Federal Reserve apparently responds in the same way to signals (presumably even those it disputes) whether or not the Administration appears to be in disarray. A related hypothesis would be that "tightness" signals (with which the recipient agrees) should produce a stronger Fed response than "ease" signals (with which the recipient disagrees). However, as Havrilesky (1993, 127–128) himself demonstrated, one cannot reject the hypothesis that the policy effects of signals for ease are equivalent to those for signals for tightness. These results are not what we would expect to find if these signals reflected persuasive communications.

If Havrilesky's data really measure signals, then presumably signals should cluster immediately prior to a FOMC meeting. The stimulus value of a signal would decrease the longer the time between the signal and the opportunity to act in response to it.²⁵ We can examine this by taking the date for each signal and relating it to the closest FOMC meetings. How many days before the next meeting did the signal occur, and how many days after the prior meeting?

Consider the following relatively weak test: The test is whether the rate of signaling in the week before FOMC meetings was greater than would have been expected purely by chance. The period of study is January 1964 through November 1991—1452 weeks and 257 signals. Thus, under the assumption of randomness, the probability of a signal in any given week is 0.177. During the same

²³ In this data series, signals for tightening are coded as negative and signals for easing are coded as positive. Thus, two tightens plus an ease equal a net signal of -1 . These are moving three-week sums of weekly net signals—Havrilesky calls this SAFER3.

²⁴ Havrilesky's sample starts in 1964. Data from the Federal Reserve Bank of St. Louis begin only in 1966. Overall results are very close, so this truncation is substantively unimportant.

²⁵ If the critical decisions occur between meetings, then this test would be mistaken. Critical decisions have been taken between FOMC meetings, but it is unusual.

period, there were 305 FOMC meetings. Purely by chance alone, we would expect to find fifty-four signals in the weeks before those FOMC meetings ($305 * 0.177$)—and the 0.95 confidence interval implies between sixty-seven and forty-one signals.²⁶ The actual count is sixty-two signals in the seven days *before* the FOMC meeting (including the meeting day itself). The count for the week immediately after the meeting is lower, at only forty-four signals. However, while the differences are in the “correct” direction, we cannot conclude that the reported signals were nonrandom.

Even the finding of more signals just before meetings may be spurious. The approaching date of an FOMC meeting may lead journalists to write (or editors to request) stories about the upcoming meeting. An obvious dramatic story line is of Administration dissatisfaction with policy. Further, the hypothesis consistent with persuasive communications is not “more signals than would happen by chance” but “many more than would happen by chance.” Persuasion attempts would be especially and prominently juxtaposed to critical decision dates. By a more stringent test, there would be many fewer signals than expected just before FOMC meetings.

Perhaps “signals” are not about FOMC meetings. An Administration may signal to reinforce messages delivered privately to the Fed Chairman. This can also be tested. During the Carter Administration, the President and the Fed Chairman met thirty-five times—we know this from the “Daily Diary,” at the Carter Presidential Library. Taking the dates of those meetings, we can relate Havrilesky’s signals to those meetings. Of the fifty Carter Administration signals, eighteen were in weeks before or after White House meetings—seventeen were expected from pure chance.²⁷

In a final test, we fail to identify a declining marginal efficiency of signaling. To test this, I partitioned the data between weeks with three or more signals (BIGSAF) and weeks with two or fewer signals (SMSAF). A simple regression of the change in the Federal Funds rate on these two series shows no difference in the Federal Reserve’s sensitivity to signaling volume.

$$\text{DFFR} = 0.0040 - 0.0566(\text{BIGSAF}) - 0.048(\text{SMSAF})$$

(3.180) (3.077)

R^2 : 0.014; df 1345 (t-test in parentheses)

²⁶ The 0.95 confidence interval around .177 with a sample of 305 is ± 0.0428 .

²⁷ Fifty signals in 208 weeks give a weekly signal probability of .24. The thirty-five meetings yield seventy sample weeks (before and after), and an expectation of $.24 * 70 = 16.8$ signals $\pm \approx 7.0$.

Indeed, these results suggest that a large volume of signals is more efficient in moving the federal funds rate than a small volume. Given these results, the main puzzle is why the Administration does not signal the Federal Reserve a great deal more. At six basis points per signal, it would be a simple matter for the Administration to move the Fed funds rate a half point in a month or 4 percent in a year. But this does not happen. In large part, I believe, that is because these data are not really a record of signals, but of the occasions on which journalists decided to write about signals.

In summary, by several tests, Havrilesky’s data fail to behave as if they were the result of persuasive communication to the Federal Reserve from the Administration. Indeed, the data look like largely random behavior. Scholars who use data drawn from media counts can develop hypotheses about the behavior of the underlying event series in order to evaluate their data. Of course, the less we know about the underlying process, the more challenging this task will be.

Media Processes

Scholars can also try to investigate directly how media processes might affect the generation of the event count data. This requires thinking about how and why the media might produce “news” accounts at rates different from the underlying event rate.

This can also be illustrated with the Havrilesky data. As a news organization, the *Wall Street Journal* has certain features that may affect the data, and the *Journal* has changed its coverage of the Fed over time in ways that also probably affect the data. An interesting fact about the Havrilesky data is that a disproportionate number of the reports of signaling have been published on Fridays or Mondays. By chance, 20 percent of signals would occur on any given day of the week; in fact, 26 percent of signals were reported on Monday and 25 percent on Friday.²⁸ Tuesdays only have 12 percent. This is not, as is evident from the data presented above, because those reports occurred just prior to FOMC meetings. A divergence as great as this is highly unlikely by chance.²⁹

By contrast, the Monday/Friday effect might reflect the *Wall Street Journal*’s practice of publishing certain

²⁸ Assuming signal stories are equally probable each day (the *Wall Street Journal* publishes Monday through Friday). Given the paper’s rigid and stylized layout, this seems like a fair assumption, but it might not hold for other newspapers (Riffe, Aust, and Lacy 1993).

²⁹ Data are for January 1964 through November 1991. Null hypothesis: signaling is equally likely each day of the week. Test: $\chi^2 = 17.3$ with 3 df; (χ^2 critical point is 11.1); thus, reject the null.

front-page opinion columns on Mondays and the “Washington Wire” column on Fridays. When interviewed for this research, one *Journal* correspondent reported that he was a regular contributor to a Monday column where his analysis of relations between the Fed and the Administration were “very welcome.”³⁰

The stories that comprise the Havrilesky data have particular characteristics that make for considerable flexibility as to when they appear. Tuchman (1978) distinguished “hard” from “soft” news. Hard news is news that will be obsolete if it is not distributed quickly. Soft news does not become obsolete so quickly—timeliness is not one of its characteristics. Gans (1979) also writes of “timeless features” not pegged to a specific event that can be run at any time. The Fed’s former public relations chief, Joseph Coyne, himself an experienced economics reporter before joining the Fed, advised a Fed chairman in a memo that stories about conflict at the Fed were likely to appear on Mondays because “reporters are frequently under pressure to produce a story for Monday’s papers which are usually ‘light’ so far as hard news is concerned.”³¹

Important changes took place at the *Wall Street Journal* and in monetary policy during the period covered by the Havrilesky data. In the 1980s, the rate of reported signaling increased—indeed, Havrilesky modeled this using a trend term. This increase coincided with an editorial decision by the *Journal* to increase the personnel devoted to covering the Fed. In the early 1980s, the relevant beat was narrowed to emphasize the Fed more exclusively, and two journalists often covered it. One may suspect that this is a typical media reaction to many news situations that seem to acquire increased importance or drama. The resulting increase in the number of news stories may simply reflect editorial decisions rather than significant changes in the underlying processes. Further, the *Wall Street Journal* correspondents starting in the early 1980s took special pride in their ability to “penetrate” the Fed.³² While their success may have been facilitated by objectively greater clarity within the Fed about its work, by greater conflict within the Board of Governors, and by changed attitudes about the desirability for secrecy central banking, it also seems likely to have increased the probability that signals would be reported.³³

³⁰Richard J. Levine interview, June 25, 1996.

³¹Coyne to Burns, May 22, 1972, Box D4, “Federal Reserve Bank of New York, 1971–72 (13),” Burns Collection, Gerald R. Ford Library.

³²Interviews with Alan Murray, June 27, 1996 and Kenneth Bacon, June 24, 1996.

³³Interviews with Bacon, June 24, 1996; Murray, June 27, 1996; Gannon, June 27, 1996; and Levine, June 25, 1996.

The Havrilesky data allow us to see how external events may structure the story selection by the media. In this case the peculiarity is an unusually high frequency of signals in February. The proportion of signal reports in February (15 percent) is nearly twice the expected proportion if signal reports are random (monthwise) through the year (8 percent). Why should there be such a high rate of signaling in February?

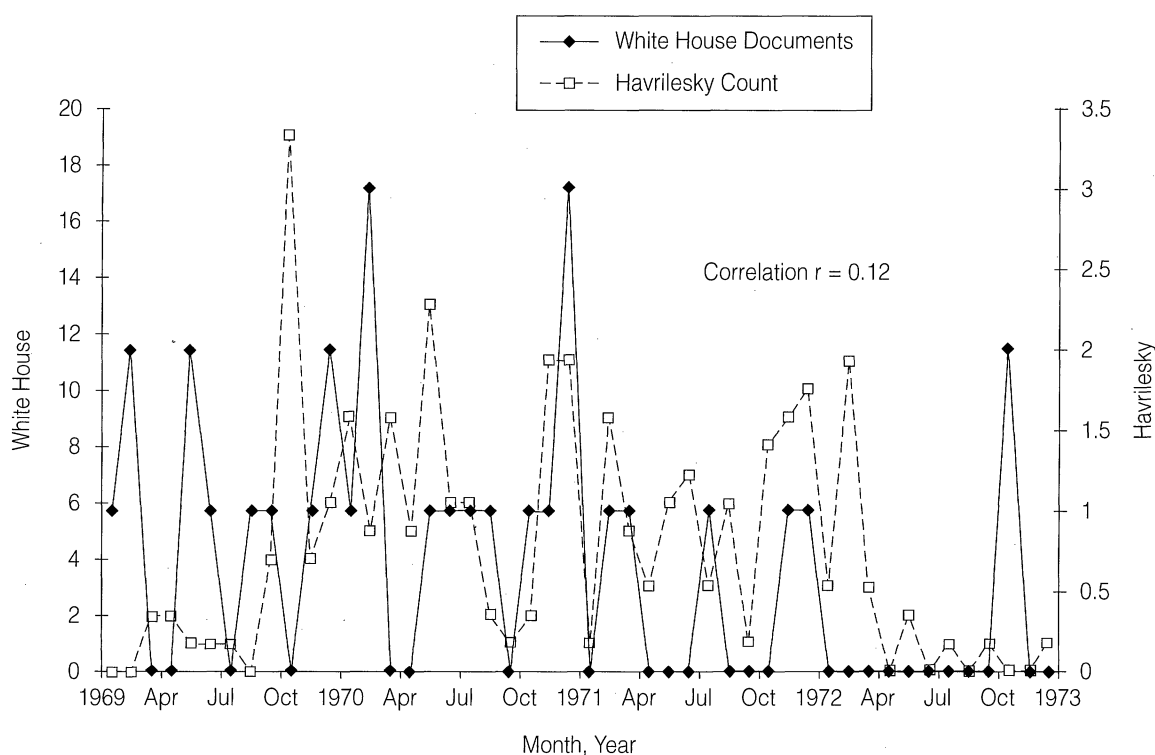
I doubt that there should be more signaling in February. However, there is a reason why more stories about signaling might be written in February. The President’s State of the Union address occurs in late January, and it is followed shortly thereafter by the submission of the President’s budget. The Joint Economic Committee holds hearings early in the year on the state of the economy and the stance of macroeconomic policy. Hearings on the economy and the budget are held by the Budget Committees and, often, by the taxing committees (House Ways and Means and Senate Finance). The Fed Chairman often testifies in all of these settings. All of these could provoke an increase in reports on monetary policy and its consistency with the Budget. All of these would suggest writing stories about conflict in February when in fact conflict would be no more likely than in January or December (when plans and projections are actually being made by the administration and the Fed).

In summary, scholars relying on media-based event counts should seek opportunities to examine the series closely for anomalies and patterns arising from media practices and in the structure of events that shape media story-selection. This may be especially feasible when counts are based on a particular publication. In the case of the Havrilesky data, we see several ways in which the series appears to have been affected by the rhythms of publication at the *Wall Street Journal*, by editorial decisions about the importance of monetary policy, by changes in access to the target organization, and by recurring external events that appear likely to suggest the relevance of “signal” stories independent of the actual level of signaling.

Validation from Independent Data

Event data can be validated in some cases when alternative, higher-quality documentary sources (record codings) become available. During the sample period, 65 percent of Havrilesky’s reported signals were for “monetary ease.” The predominance of ease signals is consistent with conventional wisdom that Presidents prefer easier policy than the Fed does. However, it well exceeds expectations that would be based on Kettl’s (1986) document-based tabulations of Council of Economic Advisor advice. The

FIGURE 4 *Wall Street Journal* Signal Count vs. Count of White House Documents Indicating Concern with Monetary Policy, 1969–72



CEA reported that policy was “too tight” only 52 percent of the time. This suggests that reporters tend to write according to a “script” in which the administration demands ease from a recalcitrant Fed much more often than is apparent in the contemporaneous documents.

Another illustration involves documentary evidence, this time from the first Nixon administration. For this period, in addition to the usual memoranda from advisors (in this case from the papers of Council of Economic Advisors chair Paul McCracken), available sources include the notes and diaries of Robert Haldeman and John Ehrlichman and the notes and correspondence available in the Burns Papers at the Gerald R. Ford Library.

Drawing on these papers, I constructed an independent event count of Administration concern about monetary policy.³⁴ In this case, an “event” is a memorandum to the President or a report of a meeting (only one count per meeting) in which explicit concerns were expressed about the Federal Reserve and monetary policy. Typically, in these records a high level of concern about mon-

etary policy was accompanied by an expressed intent to contact top Fed officials to convey Administration concerns. Although this event series is neither complete nor unbiased, it is a far better and more accurate reflection of true Administration signaling than could possibly have been available to even the most assiduous contemporaneous reporter.

The test of interest is whether there is, even in general terms, anything like a significant correlation between the document-based series and the media-based series. For this test, I have aggregated my counts and Havrilesky's counts up to the level of the calendar month. The results, displayed in Figure 4, are hardly encouraging about the media-based data. The overall correlation between the series is $r = 0.12$. The size of the Havrilesky count is generally a very poor predictor of the size of the document-based count. The Havrilesky data completely missed significant Administration concern in October 1969, March and April 1970, October 1971, and February 1972, and indicated substantial concern when there is almost none evident in the documents in February 1969, May 1969, and October 1972.

In summary, then, intensive documentary evidence does not provide ex-post validation for the basic assump-

³⁴Several scholars who have independently searched the archives collected the documents involved. These include Robert Hetzel, Wyatt Wells, and Donald Kettl in addition to me.

tion of the Havrilesky data—that it is a reasonably accurate “mirror” of the underlying processes.³⁵ I am not certain that the poor performance for the Nixon-era is typical of the whole series, but none of the other evidence I have seen makes me very confident.

Conclusion

Scholars interested in quantitative analysis have repeatedly turned to counts of stories in print media as a data source. Often, scholars use media counts to measure some process which is itself either unobservable or for which good data have not been available in relevant time periods. These data can be divided usefully into measures of media focus, usually associated with agenda-setting studies, and event counts measures that count the number of relevant events as reported in the media. Media-based data are attractive because they are relatively easy to acquire and to replicate. With the growing availability of online text databases, scholars will be increasingly inclined to turn to media sources for data.

The review and analysis in this paper indicate that media counts need to be used with great care. Like all record-codings, media sources involve bias and error that must be discovered and which are typically unstable over time. Even the most careful and complete journalistic accounts appear to present a partial and often biased representation of large, irregular, and important events. These biases can often be discovered, and scholars may be able to compensate for them by careful matching of media sources. Where data are based on some intermediary source, such as an annual periodical index, scholars must be sure that indexing practices do not constitute a significant source of variance in the data. Unwanted variance can arise from composition effects and from changing index-coding practices. Despite these flaws, published indexes may be as accurate a data source as a keyword-based online search.

All such data need to be tested for consistency with the internal logic of the processes involved. They should be examined for evidence that variation reflects the logic of the news media rather than the underlying social process. They should be validated wherever possible with other data that provide a more close and direct link to

the underlying process (see another example in Hill, Hanna, and Shafqat 1997).

Examination of the Havrilesky media-count series illustrates how an apparently valid indicator can be misleading. In this case, we have little basis for concluding that variation in the data truly arises from Administration behavior as opposed to the behavior of *Wall Street Journal* editors and reporters. The Havrilesky data are still interesting to study, but they are data not about policy behavior so much as about journalistic behavior.

Scholars relying on newspapers for data on policy-making processes need to remind themselves that what we read partly reflects the processes that direct the attention of journalists to certain stories or story lines. Trivial or transient events may receive a degree of attention equivalent to very important ones. Important events may not receive attention at all. “Positive” findings may be emphasized even though they float in a sea of “negative” findings.

There is every reason to seek to use media data. As always, measurement using these data demands careful attention to the underlying concept and diligent attention to the quality of available sources. Media sources may not provide good data even though the data are sometimes easily obtained. The ease of obtaining data through online searches is seductive, but good work continues to demand “tedious comparison of available sources” (Burrowes 1974).

Manuscript submitted May 11, 1998.

Final manuscript received July 19, 1999.

References

-
- Adams, William C. 1986. “Whose Lives Count? TV Coverage of Natural Disasters.” *Journal of Communication* 36:113–122.
 - Althaus, Scott L., Jill A. Edy, and Patricia Phalen. 1998. “Which Text is Best: Using Newspaper Indexes for Political Communication Research.” Presented at the Meeting of the American Political Science Association, Boston, MA.
 - Banks, Arthur S. 1997. *Cross-National Time Series Data Archive*. Binghamton, N.Y.: Computer Solutions Unlimited.
 - Baumgartner, Frank R., and Bryan D. Jones. 1993. *Agendas and Instability in American Politics*. Chicago: University of Chicago Press.
 - Baumgartner, Frank R., Bryan D. Jones, and Michael C. MacLeod. 1998. “Lessons from the Trenches: Ensuring Quality, Reliability, and Usability in the Creation of a new Data Source.” *The Political Methodologist* 8:1–10.
 - Bond, Doug, J. Craig Jenkins, Charles L. Taylor, and Kurt Schock. 1997. “Mapping Mass Political Conflict and Civil Society.” *Journal of Conflict Resolution*. 41:553–579.

³⁵Covering the Fed was like “trying to describe the movements of men in black suits in a dark room” (Gannon interview June 27, 1996). All correspondents emphasized the barriers to learning what was happening inside the Fed and described piecing together stories from many separate fragments of information.

- Brody, Richard A. 1991. *Assessing the President: The Media, Elite Opinion and Public Support*. Stanford: Stanford University Press.
- Buckman, Robert T. 1993. "How Eight Weekly Newsmagazines Covered Elections in Six Countries." *Journalism Quarterly* 70:780-792.
- Burrowes, Robert. 1974. "Mirror, Mirror on the Wall. . . : A Comparison of Event Data Sources." In *Comparing Foreign Policies: Theories, Findings, and Methods*. New York: John Wiley & Sons.
- Chappell, Henry, Thomas Havrilesky, and Rob McGregor. 1993. "Partisan Monetary Policies: Presidential Impact through the Power of Appointment." *Quarterly Journal of Economics* 108:185-218.
- Cohen, Bernard C. 1963. *The Press and Foreign Policy*. Princeton: Princeton University Press.
- Danzger, M. Herbert. 1975. "Validating Conflict Data." *American Sociological Review* 40:570-584.
- Doran, Charles F., Robert E. Pendley, and George F. Antunes. 1973. "A Test of Cross-National Event Reliability." *International Studies Quarterly* 17:175-203.
- Edwards, George C., III, William Mitchell, and Reed Welch. 1995. "Explaining Presidential Approval: The Significance of Issue Salience." *American Journal of Political Science* 39:108-134.
- Entwistle, Vikki. 1995. "Reporting Research in Medical Journals and Newspapers." *British Medical Journal* 310:920-923.
- Fan, David P. 1994. "Predicting Public Opinion from Press Coverage." *The Public Perspective* (August):21-23.
- Flemming, Roy B., John Bohte, and B. Dan Wood. 1997. "One Voice Among Many: The Supreme Court's Influence on Attentiveness to Issues in the United States, 1947-92." *American Journal of Political Science* 41:1224-1250.
- Froyen, Richard, Thomas Havrilesky, and Roger Waud. 1997. "The Asymmetric Effects of Political Pressures on US Monetary Policy." *Journal of Macroeconomics* 19:471-493.
- Franzosi, Roberto. 1987. "The Press as a Source of Socio-Historical Data: Issues in the Methodology of Data Collection from Newspapers." *Historical Methods* 20:5-16.
- Gaddy, Gary D., and Enoch Tanjong. 1986. "Earthquake Coverage by the Western Press." *Journal of Communications* 36:105-112.
- Gamson, William A., and Andre Modigliani. 1987. "The Changing Culture of Affirmative Action." *Research in Political Sociology* 3:137-177.
- Gans, Herbert J. 1979. *Deciding What's News*. New York: Pantheon.
- Gilbert, Susan. 1997. "Two Spanking Studies Indicate Parents Should Be Cautious." *New York Times*. August 20, B10.
- Gilens, Martin. 1996. "Race and Poverty in America: Public Misperceptions and the American News Media." *Public Opinion Quarterly* 60:515-541.
- Havrilesky, Thomas. 1993. *The Pressures on American Monetary Policy*. Boston: Kluwer Academic.
- Hertog, James K., John R. Finnegan, Jr., and Emily Kahn. 1994. "Media Coverage of AIDS, Cancer, and Sexually Transmitted Diseases: A Test of the Public Arenas Model." *Journalism Quarterly* 71:291-304.
- Hill, Kim Quaile, Stephen Hanna, and Sahar Shafqat. 1997. "The Liberal-Conservative Ideology of U.S. Senators: A New Measure." *American Journal of Political Science* 41:1395-1413.
- Houn, Florence, Mary Bober, Elmer E. Huerta, Stephen D. Hursting, Stephen Lemon, and Douglas Weed. 1995. "The Association between Alcohol and Breast Cancer: Popular Press Coverage of Research." *American Journal of Public Health* 85:1082-1086.
- Howell, Llewellyn D., and Gillian Barnes. 1993. "Event Data for Region-Specific Interactions: A Research Note on Source Coverage." In *International Event-Data Developments: DDIR Phase II*, ed. Richard L. Merritt, Robert G. Muncaster, and Dina A. Zinnes. Ann Arbor: University of Michigan Press.
- Hufker, Brian, and Gray Cavender. 1990. "From Freedom Flotilla to America's Burden: The Social Construction of the Mariel Immigrants." *The Sociological Quarterly* 31:321-335.
- Kaufman, Philip A., Carol Reese Dykers, and Carole Caldwell. 1993. "Why Going Online for Content Analysis Can Reduce Research Reliability." *Journalism Quarterly* 70:824-832.
- Keshishian, Flora. 1997. "Political Bias and Nonpolitical News: A Content Analysis of an Armenian and Iranian Earthquake in the *New York Times* and the *Washington Post*." *Critical Studies in Mass Communications* 14:332-343.
- Kettl, Donald F. 1986. *Leadership at the Fed*. New Haven: Yale University Press.
- Koren, Gideon, and Naomi Klein. 1991. "Bias Against Negative Studies in Newspaper Reports of Medical Research." *JAMA* 226:1824-1826.
- Kosicki, Gerald M. 1993. "Problems and Opportunities in Agenda-Setting Research." *Journal of Communication* 43:100-127.
- Lawler, John. 1950. *The H. W. Wilson Company: Half A Century of Bibliographic Publishing*. Minneapolis: University of Minnesota Press.
- Lichbach, Mark. 1984. "The International News about Governability: A Comparison of the *New York Times* and Six News Wires." *International Interactions* 10:311-340.
- Mayer, William G. 1993. "Trends in Media Usage." *Public Opinion Quarterly* 57:593-611.
- Mazur, Allan. 1990. "Nuclear Power, Chemical Hazards, and the Quantity of Reporting." *Minerva* 28:294-323.
- McAdam, Doug. 1982. *Political Process and the Development and the Development of Black Insurgency 1930-1970*. Chicago: University of Chicago Press.
- Merritt, Richard L. 1994. "Measuring Events for International Political Analysis." *International Interactions* 20:3-33.
- Patterson, Samuel C., and Gregory A. Caldeira. 1990. "Standing Up for Congress: Variations in Public Esteem Since the 1960s." *Legislative Studies Quarterly* 15:25-47.
- Reardon, Kathleen Kelly. 1991. *Persuasion in Practice*. Newbury Park: Sage.
- Riffe, Daniel, Charles F. Aust, and Stephen R. Lacy. 1993. "The Effectiveness of Random, Consecutive Day and Constructed Week Sampling in Newspaper Content Analysis." *Journalism Quarterly* 70:133-139.
- Robinson, Michael J. 1981. "Three Faces of Congressional Media." In *The New Congress*, ed. Thomas Mann and Norman J. Ornstein. Washington, D.C.: American Enterprise Institute.
- Rosellini, Lynn. 1998. "When to Spank." *U.S. News & World Report*. April 13, 124:52-53, 55, 57, 58.
- Rosenblum, Mort. 1981. *Coups and Earthquakes: Reporting the World to America*. New York: Harper Colophon.

- Rosenstone, Steven J., and John Mark Hansen. 1993. *Mobilization, Participation and Democracy in America*. New York: Macmillan.
- Rothman, Stanley, and S. Robert Lichter. 1987. "Elite Ideology and Risk Perception in Nuclear Energy Policy." *American Political Science Review* 81:383-404.
- Saracevic, Tefko, and Paul Kantor. 1991. "Online Searching: Still an Imprecise Art." *Library Journal* 116:47-51.
- Schrodt, Philip A. 1994. "The Statistical Characteristics of Event Data." *International Interactions* 20:35-53.
- Schrodt, Philip A., and Deborah J. Gerner. 1994. "Validity Assessment of a Machine-Coded Event Data Set for the Middle East, 1982-92." *American Journal of Political Science* 38:825-854.
- Schrodt, Philip A., Shannon G. Davis, and Judith L. Weddle. 1994. "Political Science: KEDS—A Program for the Machine Coding of Event Data." *Social Science Computer Review* 12:561-587.
- Sigler, John H. 1972. "Reliability Problems in the Measurement of International Events in the Elite Press." *Applications of Events Data Analysis: Cases, Issues, and Programs in International Interaction*, ed. John H. Sigler, John O. Field, and Murray L. Adelman. Beverly Hills: Sage.
- Singer, Eleanor, Phyllis Endreny, and Marc B. Glassman. 1991. "Media Coverage of Disasters: Effect of Geographic Location." *Journalism Quarterly* 68:48-58.
- Snider, J. H., and Kenneth Janda. 1998. "Newspapers in Bytes and Bits: Limitations of Electronic Databases for Content Analysis." Presented at the Annual Meeting of the American Political Science Association, Boston, Mass.
- Snyder, David, and William R. Kelly. 1977. "Conflict Intensity, Media Sensitivity, and the Validity of Newspaper Data." *American Sociological Review* 42:105-123.
- Starobin, Paul. 1995. "A Generation of Vipers: Journalists and the New Cynicism." *Columbia Journalism Review* March/April:25-32.
- Stiff, James B. 1994. *Persuasive Communication*. New York: Guilford Press.
- Taft, William H. 1982. *American Magazines for the 1980s*. New York: Communications Arts Books.
- Taylor, Charles Lewis, and David A. Jodice. 1983. *World Handbook of Political and Social Indicators Volume 2: Political Protest and Government Change*. 3rd ed. New Haven: Yale University Press.
- Terkildsen, Nayda, Frauke I. Schnell, and Christina Ling. 1998. "Interest Groups, the Media, and Policy Debate Formation: An Analysis of Message Structure, Rhetoric and Source Cues." *Political Communication* 15:45-61.
- Tidmarch, Charles M., and John J. Pitney. 1985. "Covering Congress." *Polity* 17:463-483.
- Tuchman, Gaye, 1978. *Making News: A Study in the Construction of Reality*. New York: Free Press.
- Wagers, Robert. 1992. "Dowquest and Dow Jones Text-Search: Which Works Best and When?" *Online* 16 (November):35-36, 38, 40-41.
- Walker, Jack L., Jr. 1977. "Setting the Agenda in the U.S. Senate." *British Journal of Political Science* 7:423-445.
- Weart, Spencer. 1988. *Nuclear Fear: A History of Images*. Cambridge: Harvard University Press.
- Weimann, Gabriel, and Conrad Winn. 1994. *The Theater of Terror: Mass Media and International Terrorism*. New York: Longman.
- Westerståhl, Jörgen, and Folke Johansson. 1986. "News Ideologies as Moulders of Domestic News." *European Journal of Communication* 1:133-150.
- Wood, B. Dan, and James E. Anderson. 1993. "The Politics of U.S. Antitrust Regulation." *American Journal of Political Science* 37:1-39.
- Wood, B. Dan, and Richard W. Waterman. 1993. "The Dynamics of Political-Bureaucratic Adaptation." *American Journal of Political Science* 3:497-528.
- Zaller, John. 1992. *The Nature and Origins of Mass Opinion*. New York: Cambridge University Press.
- Zhu, Jian-Hua. 1992. "Issue Competition and Attention Distraction: A Zero-Sum Theory of Agenda-Setting." *Journalism Quarterly* 69:825-836.