

## FINDING NEWS STORIES: A COMPARISON OF SEARCHES USING LEXISNEXIS AND GOOGLE NEWS

By David A. Weaver and Bruce Bimber

*This study compares Google News to LexisNexis for finding stories in the New York Times, eight large-circulation U.S. newspapers, and all English-language news outlets in each database. Inter-database agreement between Google News and LexisNexis ranged from 29% to 83%, with much of the discrepancy due to wire service exclusions: LexisNexis missed half or more of stories appearing in major papers and in broad searches of English-language news because it is blind to wire stories. Wire-service blind spots in news archives can be a substantial limitation.*



Researchers studying media commonly employ digital research archives for data about news. Services such as LexisNexis, NewsBank, ProQuest, and others provide full-content access to many news sources and have long been standard research tools. Recently, free Web-based news portals such as AOL News, Google News, Yahoo! News, and Slashdot have presented potential alternatives to traditional archives, but are designed for public rather than academic use. Research relying on these services is now being reported and shedding new light on some of the limitations of traditional archives.<sup>1</sup>

One well-known drawback of traditional news archives is that they typically offer text-only versions of original stories lacking graphics and information about headline size and story placement, which can be important to certain kinds of research. A potentially more serious problem with traditional archives is the fact that content originating with wire services is typically stripped out of newspapers before stories are archived.<sup>2</sup> This means that news databases do not necessarily constitute archives of the whole content of news appearing in a particular news outlet or in a specific news market. This limitation is of particular concern for research requiring accurate assessments of all news in distribution about a particular subject or in a particular venue, but it is not widely recognized by researchers. It is clear that a good deal of news, especially national news distributed at the local

*David A. Weaver is a Ph.D. student in the Department of Political Science at the University of California, Santa Barbara. Bruce Bimber is professor in the Departments of Political Science and Communication at the University of California, Santa Barbara. The authors thank the UCSB Center for Nanotechnology in Society; Lucia Snowhill and Harold Colson for consultation on LexisNexis; as well as Jerry Macala, Erica Lively, Gary Haddow, and Nicole Tyler for research assistance. This material is based upon work supported by the National Science Foundation under grant SES 0531184.*

**J&MC Quarterly**  
Vol. 85, No. 3  
Autumn 2008  
515-530  
©2008 AEJMC

---

level, is comprised of wire stories that are lost in the news archiving process.

To explore this methodological challenge to news research, this study compared a traditional news archive with a Web-based news aggregator: LexisNexis, the most widely used news archive in the social sciences,<sup>3</sup> and Google News, which reaches an audience of about nine million people per month and which has been the subject of a good deal of scholarly and professional interest.<sup>4</sup> The study used the two databases to track a news topic across many months, examining the entire news stream from elite newspapers to local outlets in order to assess the comparability of search results between LexisNexis and Google News, and to use Google News to estimate the number of stories missing from LexisNexis because of omission of wire service content.

As the news business struggles with fragmenting audiences, shrinking advertising revenue, and reductions in news staff, many observers expect wire services, syndicated content, and even blogs and public-contributed material to play an increasingly important role in the news.<sup>5</sup> This would mean that, absent changes in practice about what is archived, traditional news databases would grow less useful for identifying what news reaches the public.

For the analysis, a topic was chosen that is part of a larger study of news coverage of technological innovation, especially issues of regulation and public safety of new technologies. This study focused specifically on the class of technologies labeled "nanotechnology," which is emerging as a news topic. For example, in 2006, a product recall occurred in Germany that generated news in Europe and the United States. The same year, the Environmental Protection Agency (EPA) announced it would regulate products containing nanotechnology in cases where manufacturers make bacteria-killing claims, and environmentalists and consumer groups petitioned for regulatory action from the Food and Drug Administration (FDA), which declined to act.

These and other developments associated with nanotechnology have two useful characteristics. First, nanotechnology presents a discrete topic with a distinctive and identifiable name, facilitating reliable and valid searches. Second, because the topic is new to journalists and editors, a good deal of variation in attention to it over time and across outlets is expected, and this variation should provide a strong test of how the databases compare at finding news stories.

---

### **Methodological Challenges in Use of News Databases**

Research on media typically treats the use of archival news databases as unproblematic. This is true of work on agenda setting and other aspects of news and public opinion, of research on independence of news media from government and the relationship of local to national news, and studies of news consonance, among other topics.<sup>6</sup> However, the methodological literature on use of databases has identified several limitations. Research has shown problems arising from variation between original news content and the content in databases due to truncations of articles, mistakes in capturing headlines, unannounced changes in con-

---

tracts between news sources and news databases, and idiosyncratic exclusions from databases, such as corrections or retractions from newspapers that are not archived.<sup>7</sup> Variation of this kind can introduce error into studies of news content, but this error is not large and more importantly does not necessarily bias results systematically.

A more troublesome class of problem involves systematic database exclusions.<sup>8</sup> Some items are obvious in their absence from text-based databases, such as photographs and other graphics, headlines, and cues about article placement. To find such information, researchers may be able to consult the original news source in print or microfilm, or in many cases the news source's own online archive. A more troublesome exclusion involves wire service stories. Stories originating with services such as the Associated Press (AP) are virtually always removed when a newspaper's content is archived in research databases, because licensing agreements customarily permit one-time use only of wire content by a news outlet. The same stipulation typically applies to newspapers' own online archives, which also generally omit wire-originated stories.

This problem was noted by Snider and Janda in an unpublished manuscript in 1998, but has not been assessed systematically.<sup>9</sup> Research designs that require finding news stories that reached the public—as opposed to studying news production, for example—rarely address the problem. For instance, a recent study of agenda setting in health news used LexisNexis for finding articles in a set of major newspapers in order to analyze public exposure to positively and negatively valenced news.<sup>10</sup> The technique produced 715 stories. Our own search using the authors' search terms shows that wire services *produced* roughly a hundred articles on the topic during the period under study, some unknown number of which would have appeared in the papers examined, and some of which would have appeared in multiple outlets.<sup>11</sup> The Associated Press accounts for over 90% of those stories, meaning that omitted news was dominated by a single organization's perspective.

The fraction of news reaching the public that originates with wire services is not known. Niederdeppe, Frosch, and Hornik estimate that more than 85% of U.S. newspapers subscribe to the AP service, but do not estimate how much news in distribution is actually AP-based.<sup>12</sup> Within a single newsroom, an editor may accept anywhere from a few to half of available stories from the AP.<sup>13</sup> While large news businesses typically report more stories written by their own staff, local news outlets may rely entirely on wire stories and syndicated material for national news.<sup>14</sup> Some researchers have approached this problem by coding AP content itself as a proxy for national news, but this technique is not common and is limited by absence of information about the distribution of AP stories.<sup>15</sup>

The changing structure of the news business may be exacerbating this problem of finding what news the public actually sees. Shrinking audiences for national news broadcasts and falling subscriptions and advertising revenues have led to substantial reductions in news staff at

---

newspapers across the country; many news businesses appear to be relying more over time on wire stories and syndicated material, as well as Web-based content such as blogs, which are also typically excluded from news archives.<sup>16</sup> One study showed that after purchase by Gannett, more than half of stories in the *Louisville Courier-Journal* originated from wires rather than staff writers.<sup>17</sup>

---

**Background  
on Lexis-  
Nexis  
and Google  
News**

LexisNexis and Google News offer a useful comparison for this problem. LexisNexis is the most widely used news archive in the social sciences.<sup>18</sup> For the United States, it offers roughly 300 newspapers and about 500 general print publications, along with transcripts from about three-dozen broadcast outlets. It covers virtually all big-city newspapers in the United States; many papers from mid-size markets, such as the *Charleston Gazette* and *Anchorage Daily News*; and a smaller sample of local news outlets.<sup>19</sup> LexisNexis omits wire service stories in distribution, though it provides searches of wire content as it is produced.<sup>20</sup> Methodological analyses of LexisNexis are few in number. The most comprehensive report is that of Deacon, who compared LexisNexis and the Chadwyck Healy newspaper archives for the United Kingdom and reported a number of small to moderate-sized reliability problems due to idiosyncratic differences.<sup>21</sup>

Google News has been a Web-based “portal” to news since 2002, and some of its features are complementary to those of LexisNexis. It is unique among news aggregators in providing a fully automated sweep of global news, conducted roughly every fifteen minutes and without human editorial judgments.<sup>22</sup> Google claims to include 4,500 news sources. Its database can be searched in full text mode, like LexisNexis, and is also represented by a changing home page of some three dozen news topics currently covered by the largest number of news businesses globally.<sup>23</sup> A major news topic, such as an election outcome, might be linked to 5,000 or more stories, while less prominent topics are linked to a hundred or fewer. Though it produces no news of its own, Google News has been described as a paradigmatic example of the changing news business, because it provides “radically expanding access to diverse viewpoints on any issue,” in contrast to journalism’s traditional emphasis on creating filtered, ordered, and prioritized news.<sup>24</sup>

One of the most useful features of Google News for the present study is the fact that between its inception in 2002 and late summer of 2007, it included wire stories as they appeared in distribution throughout the news system. An AP story, for example, that ran in a hundred newspapers around the country would produce a hundred results, making possible analysis of news distribution. By contrast, that same article would appear once in the LexisNexis database, associated with its producer, the AP. In September of 2007, Google News stopped its practice of displaying “duplicate content,” instead agreeing to license stories directly from four major wire services and thereby reducing a source of complaint about copyright infringement.<sup>25</sup> For the five years that Google News provided a window into news-in-distribution, it was possi-

---

ble to compare results with LexisNexis's database of news-in-production.

## Research Questions

The first research question addresses the *New York Times*, which is a common focus of research on news because it is viewed as intrinsically important, or is treated as a proxy for the agenda of the entire news system on the assumption that it serves as a primary "intermedia agenda setter."<sup>26</sup> The first step in our comparison examines if searches for *Times* articles with Google News can replicate LexisNexis searches. Then the comparison is extended to other large-circulation newspapers in the United States.

**RQ1:** What is the level of agreement between LexisNexis and Google News when searching for news stories in the *New York Times* and other large-circulation U.S. papers?

Our general expectation is that the two databases will produce similar pictures of news coverage because both databases include the *Times* and roughly the same major papers, and because these papers rely heavily on their own reporting. Any differences will be chiefly the result of idiosyncratic variation in the construction of the two archives.

The second question turns to the contents of the two databases beyond the major newspapers, in order to identify how each portrays news content broadly. Expanding the analysis from major papers to include local news businesses and foreign outlets suggests divergence in results, since LexisNexis's news-production emphasis should result in fewer stories than Google News's news-distribution emphasis.

**RQ2:** How large is the difference in the total number of stories found by LexisNexis and Google News when searching all U.S. and world news?

Even when wire service stories are removed from Google results, some differences should remain in the portraits of the news system provided by LexisNexis and Google News, because they do not search the same body of news sources—Google News reports 4,500 sources on its home page, while LexisNexis has roughly half that number.

In addition to assessing the total volume of news stories that LexisNexis and Google News find, the study compares results when the attention of the media coalesces on a particular issue or development. The assumption is that a baseline level of coverage of the topic will be occasionally interrupted when some key development, such as an action by public officials, leads to a rise in news volume. Without any *a priori* judgment about what should be a newsworthy development, the study compares what LexisNexis and Google News suggest are big news events.

**RQ3:** What is the level of agreement between LexisNexis and Google News content in depicting major news events?

**News Topic: Regulation of Technology.** The analysis employs a longitudinal design focused on one topic in order to track coverage in LexisNexis and Google News long enough to observe the rise and fall of various developments in the news, and to see how the two databases handled these. The topic, the safety and regulation of nanotechnologies, offers several useful features for the present inquiry. "Nanotechnology" is a readily searchable topic that produces relatively unambiguous results. As a news topic, nanotechnology is new to most journalists, as it is to the public more generally, yet a number of potentially newsworthy developments occurred during 2006 and 2007, producing a variety of opportunities for reporting.<sup>27</sup> In the United States, environmental and other groups were active in advocating for new regulations; several high-level advisory bodies and think tanks issued expert reports about policy; the city of Berkeley, California, adopted an ordinance requiring firms to disclose the presence of nanotechnology within city limits; and the EPA and FDA took preliminary actions on public safety and nanotechnology.

**Data and Analysis.** To examine how the databases capture the response of news media to such developments, a data set was constructed containing search results for news coverage of nanotechnology in bi-weekly searches from January 29, 2006, to August 15, 2007. This period was chosen in order to begin the dataset prior to the first actions historically by any U.S. regulatory agencies on nanotechnology and also prior to the recall in Germany of a cleaning product called MagicNano. Three months of baseline news prior to these developments were included; data collection continued until about a half-dozen newsworthy events had occurred involving public officials, advocacy groups, or experts. This criterion was met in mid-2007, not long before Google's database change that excluded wire service stories in circulation. The result is an eighteen-month dataset that exceeds the power of traditional news-sample techniques such as the constructed-week sampling approach, and it captures any English-language news found by the two sources.<sup>28</sup> The dataset allows plotting of stories over time and calculation of level of agreement between the two.<sup>29</sup>

The design employed a set of search terms that provides a broad sweep through full text news. After construction of equivalent Boolean terms in LexisNexis and Google News, searches were performed for stories containing the term "nanotechnology" or several permutations, and at least one term from a list of twenty-one concepts relating to safety, risk, environmental issues, and regulation or public policy, which were developed from the literature on societal implications of nanotechnology and through discussion with topical experts at the National Science Foundation's Center for Nanotechnology in Society at the University of California, Santa Barbara.<sup>30</sup> In LexisNexis, the search involved: "General News," which consisted of 52 "Major Papers" and 502 "Magazines & Journals"; "United States News," which consisted of 267 sources from the Midwest, Northeast, Southeast, West, all 50 states, and the District of Columbia; "News Transcripts" which includes material from 35 sources; "World News," which includes 1,289 sources; and "News Wire" sources,

---

which numbered 357. Accounting for entries appearing in multiple categories, this sample included approximately 2,500 unique news sources. In Google News, parallel searches were conducted of the entire database of 4,500 sources every two to three weeks, in order to stay within the moving window formed by Google's policy of archiving its main news content for only thirty days.<sup>31</sup> Google's page rank, placement, and other subjective indicators of importance were ignored.

For purposes of examining the major papers including the *New York Times*, the ten largest circulation papers in the United States were identified, using figures from the Audit Bureau of Circulations as of September 30, 2006, which were *USA Today*, *Wall Street Journal*, *New York Times*, *Los Angeles Times*, *Denver Post*, *Chicago Tribune*, *Washington Post*, *New York Daily News*, *New York Post*, and *Houston Chronicle*.<sup>32</sup> The *Chicago Tribune* and the *Wall Street Journal* were dropped because LexisNexis does not provide full-text searching of these. The resulting list of eight newspapers constituted the population of news sources for the major-papers comparison.

Each story returned from the searches was coded as relevant or not to the topic, and then, for relevant stories, variables were generated directly from the databases, including *story*, a text string identifying each article by title; *date* as found by the database; *LexisNexis*, a dichotomous variable describing whether the article was found by LexisNexis; *Google News*, which similarly describes whether Google News found the article; *wire*, a dichotomous variable describing whether the article was produced by the Associated Press, United Press International, or another syndicated news service; and *publication*, a text variable giving the name of the news publication.

**Reliability.** Reliability for coding of raw search results as relevant or irrelevant was assessed, using a simple dichotomous choice about whether each article returned from Google News and LexisNexis searches was on the topic of interest, namely nanotechnology and possible health, environmental, safety, regulatory, or other societal concerns. The reliability procedure for this hand-coded step was as follows. Two coders were used, with about fifteen minutes of initial training. A pilot sample of roughly 50 stories was drawn at random from Google News and coded independently by the coders. The result was perfect agreement, so no further training was employed and the coders proceeded to the main reliability assessment.

The reliability sample has two components. For each database, a week was chosen at random, and all stories from that week were included, up to a target of approximately 150 cases from each database. This sample size permitted 95% confidence or better (two-tailed) in our reliability estimates for each database individually, following Riffe, Lacy, and Fico<sup>33</sup> and assuming actual agreement of 90%. This produced a reliability sample of 304 cases, with 147 from Google News and 157 from LexisNexis. Cohen's kappa for the combined sample was 0.76, and was 0.79 for the Google News sub-sample and 0.71 for the LexisNexis sub-sample. These figures are satisfactory for a project of this kind, though not high.

---

## Results

**New York Times and Major Papers.** The searches returned 2,565 stories from Google News and 558 from LexisNexis. For **RQ1**, just stories from the *New York Times* were examined. Google News and LexisNexis each returned six stories, five of which they have in common. The common stories covered retail products containing nanotechnology (March 12, 2006); background on societal response to risks from emerging technologies (May 17, 2006); a report by the National Research Council (September 26, 2006); potential regulatory issues associated with food additives (October 10, 2006); and the Berkeley ordinance (January 14, 2007). LexisNexis found an article in the *Times* about an agreement by DuPont and Environmental Defense regarding voluntary safety guidelines (June 21, 2007) that was missed by Google News, while Google found a "Bits" blog entry (August 14, 2007) about potential hazards in sunscreens not included by LexisNexis. Simple agreement was used to compare the two, because the study's focus is on consistency in the list of stories found by each, rather than in chance-corrected measures that would account for the total number of stories in the databases. For the *New York Times*, agreement was 71% between Google News and LexisNexis.

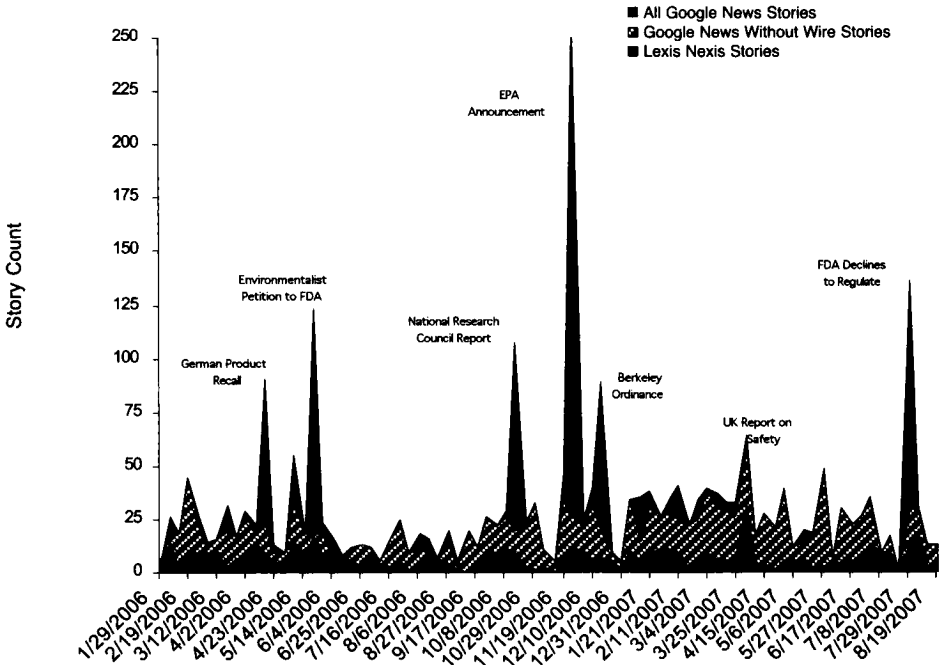
Next, the comparison was expanded to include the rest of the eight large-circulation newspapers. Among these, LexisNexis found 20 articles and Google News 47. These results include 32 articles found by Google News but not LexisNexis, and 5 found by LexisNexis but not Google News. Agreement for these results was 29%. Of the articles found uniquely by Google News, 29 of 32 originated with wire services. Google News thus revealed that LexisNexis's wire service blind spot caused it to miss more than half of the news in the eight major papers, including stories in the *Washington Post*, *Los Angeles Times*, and *Houston Chronicle*. Six of the eight major papers carried the same AP article, a July 2007 piece by Andrew Bridges about the FDA. With these wire stories removed from the analysis, agreement rose to 61%.

**All News Sources.** **RQ2** involved comparison of the 2,565 stories from all sources returned by Google News to the 558 returned by a search of LexisNexis. Figure 1 plots the number of stories per week found by each. Google News returned an average of 31 stories per week, compared to about 7 for LexisNexis. The difference is consistent with the general expectation, as Google News searches nearly twice as many sources as LexisNexis, in addition to including each occasion where a wire story runs. Google News found 893 instances of news wire stories on the topic, which exceeds the total number of stories found by LexisNexis. Without wire stories, the result was 1,675 Google News stories, for an average of about 20 per week—still more than three times the stories returned by LexisNexis. These are shown in Figure 1 by the line labeled "Google News Without Wire Stories." There are no obvious patterns in coverage that would explain these differences. For instance, Google News found 10 articles from the *Sydney Morning Herald* missed by LexisNexis, which in turn found 3 in the *Canberra Times* missed by Google News.



**FIGURE 1**

*Comparing Weekly News Volume Using LexisNexis and Google News,  
January 29, 2006, to August 15, 2007,  
Stories on Nanotechnology*



**News Events.** The differences between the two databases persist when one turns to questions of identifying discrete news events, the focus of RQ3. In the Google News data, there are a number of spikes in news volume, six of which happen to exceed one standard deviation (35.6) above the weekly mean, and four of which exceeded two standard deviations. Examining the content of stories in these six spikes shows that each reflects coverage of a discrete event, such as the German recall, the petition to the FDA, and an EPA announcement.<sup>34</sup> From the perspective of Google News, the biggest news day of the year for the topic was November 23, 2006, when 121 news outlets carried stories about the EPA.

In the LexisNexis data, meaningful news events are less obvious statistically. On nine occasions news volume rises one standard-deviation (4.9) or more above the mean, but inspection of these clusters of stories shows that four do not correspond to any single event or news development, instead reflecting the regular ebb and flow of mis-

---

cellaneous stories. Four of the six events that trigger one standard-deviation spikes in Google News findings also do so in LexisNexis.<sup>35</sup> By this simple approach to identifying spikes in news volume, the databases give quite divergent pictures of what developments were most newsworthy during the study period—that is, of when the attention of media turned more heavily to the topic. Agreement at identifying news events is 36%.<sup>36</sup>

---

## **Discussion and Conclusion**

For a topical search exclusively of the *New York Times*, agreement between Google News and LexisNexis is a modest 71%. One of the discrepancies appears idiosyncratic and one is attributable to the fact that Google News includes blog entries. Filtering the blog entry raises agreement to 83%, which approaches an acceptable level and was the highest agreement in the study. From the perspective of inclusiveness of the databases, there seems to be no particular advantage to Google News over LexisNexis for searching the *New York Times*, though some scholars might find Google advantageous because of its link to actual stories in context.

A search of eight major newspapers in the United States yielded poorer agreement (61%) if wire service stories found by Google News are set aside, and only 29% with wire stories in the picture. On this basis, LexisNexis and Google News do not provide comparable portraits of news in major news outlets. Google News revealed that LexisNexis missed more than half of the news stories in major papers. The importance of the wire services to news is not, at least in this case, confined to local coverage of national issues but extends to coverage in major newspapers.

As expected, differences between Google News and LexisNexis for broad searches of U.S. and English-language world news were substantial. Google News found four times as many stories as LexisNexis, and still about three times as many when wire service stories were excluded. For searches of local and foreign news, Google News and LexisNexis should be treated as representing different samples of news sources. So long as Google News keeps its source list proprietary, researchers will not be able to resolve these differences. Google News shows that LexisNexis is blind to a great many news stories because of the wire exclusion, and this problem extends to major news outlets. Caution should be exercised in making inferences from LexisNexis about what news actually reaches audiences. The prominent role of news wires in raising the volume of disseminated stories down the media system is consistent with classic research on news wire services, which finds that wire service news tends to be “uncritically” accepted by local media and by editors operating in national bureaus.<sup>37</sup>

Which database is better? This study was not designed to render such a judgment for any particular task. Though we regard LexisNexis as the best available news archive, these results are discouraging for researchers who need to identify all stories appearing in a set of news outlets. The problem will grow larger as restructuring of the news business continues and wire services grow more important. Ironically, the

---

fidelity of news archiving may have diminished since the days when researchers turned to microfilm. Since Google's change in practice in late 2007, both databases now suffer from the wire service blind-spot. It is clear that other techniques should be employed by researchers to account for missing stories when using LexisNexis, such as specifically analyzing AP-produced stories when conducting content analysis of newspaper content, or using selected daily searches of individual news Web sites or print editions in order to estimate the fraction of coverage accounted for by wire content.

**Limitations and Areas for Further Research.** This study focused on a national news topic low on the political agenda, using Google News and LexisNexis, neither of which is the sole source of data in its class. Our topic has some similarities and a number of differences from other topics of research on news. From the perspective of local news, it is similar to many national issues in that few local outlets cover developments themselves. We would expect the findings about the wire service blind spot to apply in cases of other news topics. From the perspective of national news, the situation may be different. The topic received fairly light and irregular attention from the major papers. The *New York Times* ran only a handful stories, all of which were produced by staff writers, while the rest of the major papers ran a mix of their own reporting and wire service stories. We are hesitant to make inferences from this study to high-salience news topics such as war, political campaigns, or economic news, which might feature stronger gatekeeping, more uniform news agendas, and comparatively less importance of wire service content in coverage by major news outlets.

Several lines of inquiry could improve understanding of these issues. The most important is characterizing the apparently growing reliance of news businesses on wire service and other syndicated content. This could be accomplished through analysis of print news and news Web sites in real time, before archiving processes remove this content.<sup>38</sup> Some researchers have employed commercial news clipping services that make daily sweeps through original news, in order to be more inclusive and create datasets with greater fidelity to actual news in distribution, a technique we endorse.<sup>39</sup> Comparisons of other news portals, such as Yahoo! News, with other online databases could also improve our understanding of archive biases. In the most general terms, our analysis provides a reminder that any news database provides nothing more than a sample of the actual news to which the public is exposed, and it is important to understand better the nature of sampling effects arising from this fact.

---

## NOTES

1. Clark C. Freifeld, Kenneth D. Mandl, Ben Y. Reis, and John S. Browstein, "Health Map: Global Infectious Disease Monitoring through Automated Classification and Visualization of Internet Media Reports,"

---

*Journal of the American Medical Informatics Association* 15 (March-April 2008): 150-57; Nathaniel D. Poor, "A Cross-National Study of Computer News Sites: Global News, Local Sites," *The Information Society* 23 (2007): 73-83; Tsuyoshi Murata, "Detection of Breaking News from Online Web Search Queries," *New Generation Computing* 26 (November 2007): 63-73; Also see Matt Carlson, "Order versus Access: News Search Engines and the Challenge to Traditional Journalistic Roles," *Media, Culture & Society* 29 (2007): 1014-30.

2. J. H. Snider and Kenneth Janda, "Newspapers in Bytes and Bits: Limitations of Electronic Databases for Content Analysis" (paper presented at the annual meeting of the American Political Science Association, Boston, MA, 1998).

3. David Deacon, "Yesterday's Papers and Today's Technology: Digital Newspaper Archives and 'Push-Button' Content Analysis," *European Journal of Communication* 22 (winter 2007): 5-25.

4. Project for Excellence in Journalism, "The State of the News Media 2007: An Annual Report on American Journalism," 2007, <http://www.stateofthenewsmedia.com/2007>; Carlson, "Order versus Access."

5. Project for Excellence in Journalism, "The State of the News Media 2004: An Annual Report on American Journalism," 2004, <http://www.stateofthenewsmedia.com/2004>.

6. Lutz Erbring, Edie N. Goldenberg, and Arthur H. Miller, "Front-Page News and Real-World Cues: A New Look at Agenda-Setting by the Media," *American Journal of Political Science* 24 (February 1980): 16-49; Maxwell E. McCombs and Donald L. Shaw, "The Agenda-Setting Function of the Mass-Media," *Public Opinion Quarterly* 36 (summer 1972): 176-87; Diana C. Mutz and Joe Soss, "Reading Public Opinion: The Influence of News Coverage on Perceptions of Public Sentiment," *Public Opinion Quarterly* 61 (autumn 1997): 431-51; David Weaver, Maxwell E. McCombs, and Donald L. Shaw, "Agenda-Setting Research: Issues, Attributes, and Influences," in *Handbook of Political Communication Research*, ed. Lynda Lee Kaid (Mahwah, NJ: Lawrence Erlbaum, 2004), 257-82.

7. Snider and Janda, "Newspapers in Bytes and Bits: Limitations of Electronic Databases for Content Analysis"; Philip A. Kaufman, Carol Reese Dykers, and Carole Caldwell, "Why Going Online for Content Analysis can Reduce Research Reliability," *Journalism Quarterly* 70 (winter 1993): 824-32; Katheleen A. Hansen, "Online Inaccuracies: The Use and Misuse of Electronic Information Sources" (paper presented at the annual meeting of the AEJMC, Washington, DC, 1995); Daniel Riffe and Alan Freitag, "A Content Analysis of Content Analyses: Twenty-Five Years of *Journalism Quarterly*," *Journalism & Mass Communication Quarterly* 74 (autumn 1997): 515-24.

8. Snider and Janda, "Newspapers in Bytes and Bits: Limitations of Electronic Databases for Content Analysis."

9. Snider and Janda, "Newspapers in Bytes and Bits: Limitations of Electronic Databases for Content Analysis."

10. Thomas Hugh Feeley and Donald Vincent III, "How Organ Donation is Represented in Newspaper Articles in the United States," *Health Communication* 21 (May 2007): 125-31.

---

11. We used LexisNexis to search wire services for articles containing [(organ or tissue) and (donation or donor or transplant)] for the period of the study.

12. Jeff Niederdeppe, Dominick L. Frosch, and Robert C. Hornik, "Cancer News Coverage and Information Seeking," *Journal of Health Communication* 13 (March 2008): 181-99.

13. Glen L. Bleske, "Ms. Gates Takes Over: An Updated Version of a 1949 Case Study," *Newspaper Research Journal* 12 (fall 1991): 88-97.

14. Snider and Janda, "Newspapers in Bytes and Bits: Limitations of Electronic Databases for Content Analysis." It has also been argued that the Associated Press has become *the* dominant wire service in the news business. See Richard A. Schwarzlose, "The Associated Press and United Press International," in *The Future of News*, ed. Philip S. Cook, Douglas Gomery, and Lawrence W. Lichty (Washington, DC: Woodrow Wilson Center Press, 1992), 145-66.

15. Niederdeppe, Frosch, and Hornik, "Cancer News Coverage and Information Seeking."

16. Project for Excellence in Journalism, "The State of the News Media 2004."

17. David C. Coulson and Anne Hansen, "The *Louisville Courier-Journal's* News Content After Purchase by Gannett," *Journalism & Mass Communication Quarterly* 72 (spring 1995): 205-15.

18. Deacon, "Yesterday's Newspapers and Today's Technology: Digital Newspaper Archives and 'Push-Button' Content Analysis."

19. As context for the size of the LexisNexis database, the 2006 newspaper market in the United States included 1,437 daily papers and more than 6,000 weeklies, and the television market comprised 1,943 licensed broadcast stations. Sources are Project for Excellence in Journalism, "The State of the News Media 2008," 2008, <http://www.stateofthenewsmedia.com/2008>; and Federal Communications Commission, "Broadcast Totals as of December 31, 2006," 2007 (Washington, DC: FCC), [http://hraunfoss.fcc.gov/edocs\\_public/attachmatch/DOC-269784A1.pdf](http://hraunfoss.fcc.gov/edocs_public/attachmatch/DOC-269784A1.pdf).

20. Snider and Janda, "Newspapers in Bytes and Bits: Limitations of Electronic Databases for Content Analysis."

21. Deacon, "Yesterday's Newspapers and Today's Technology: Digital Newspaper Archives and 'Push-Button' Content Analysis."

22. Carlson, "Order versus Access: News Search Engines and the Challenge to Traditional Journalistic Roles."

23. A 2007 analysis of the twenty-five news sources that appear most commonly among the top Google sources showed that the *New York Times* was the most common source of stories appearing on the Google News home page, followed by the *Washington Post*. Of the ten largest-circulation newspapers in the United States as measured by the Audit Bureau of Circulations, seven appear among the twenty-five most frequent sources of Google News home-page stories, along with several news wires, the major television networks, a number of foreign sources such as the BBC, the *Times Online*, the *Guardian*, and *Xinhua*, and some big-city papers below the ten largest, such as the *Boston Globe* and the

---

*Seattle Post*. See Greg Jarboe, "Revealing the Sources of Google News," *Search Engine Land*, May 31, 2007, <http://searchengineland.com/070531-133609.php>. For circulation data, see Audit Bureau of Circulations, "Top 200 Newspapers by Reported Circulation," October 18, 2006, <http://www.accessabc.com/reader/top150.htm>.

24. Carlson, "Order versus Access: News Search Engines and the Challenge to Traditional Journalistic Roles," 1015.

25. Scott Karp, "Google News Hosting Wire Service Stories Diminishes Value of Duplicate Content," *Publishing 2.0*, September 2, 2007, <http://publishing2.com/2007/09/02/google-news-hosting-wire-service-stories-diminishes-value-of-duplicate-content/>, accessed May 20, 2008; Reuters, "Google Shift on Handling of News," *New York Times*, September 1, <http://www.nytimes.com/2007/09/01/technology/01news.html>, accessed May 20, 2008.

26. Frank R. Baumgartner and Bryan D. Jones, *Agendas and Instability in American Politics* (Chicago: Chicago University Press, 1993). See also Yue Tan and David H. Weaver, "Agenda-Setting Effects among the Media, the Public, and Congress, 1946-2004," *Journalism & Mass Communication Quarterly* 89 (winter 2007): 729-44; Maxwell McCombs and Amy Reynolds, "News Influence on Our Pictures in the World," in *Media Effects: Advances in Theory and Research*, 2d ed., ed. Jennings Bryant and Dolf Zillmann (Mahwah, NJ: Lawrence Erlbaum Associates, 2002), 1-18; Maxwell McCombs, Edna Einsiedel, and David H. Weaver, *Contemporary Public Opinion: Issues and the News* (Hillsdale, NJ: Lawrence Erlbaum, 1991), 47-49. For review of this concept, see Weaver, McCombs, and Shaw, "Agenda-Setting Research: Issues, Attributes, and Influences," 272-73.

27. Dan Kahan, Paul Slovic, Donald Braman, John Gastil, and Geoffrey L. Cohen, "Affect, Values, and Nanotechnology Risk Perceptions: An Experimental Investigation," *Social Science Research Network*, March 7, 2007, [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=968652](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=968652); Chul-Joo Lee and Dietram A. Scheufele, "The Influence of Knowledge and Deference Toward Scientific Authority: A Media Effects Model for Public Attitudes Toward Nanotechnology," *Journalism & Mass Communication Quarterly* 83 (fall 2006): 819-34; Jane Macoubrie, "Informed Public Perceptions of Nanotechnology and Trust in Government" (Washington, DC: Woodrow Wilson International Center for Scholars, September 2005), <http://www.wilsoncenter.org/events/docs/macoubriereport.pdf>.

28. Joe Bob Hester and Elizabeth Dougall, "The Efficiency of Constructed Week Sampling for Content Analysis of Online News," *Journalism & Mass Communication Quarterly* 84 (fall 2007): 811-24.

29. Our data collection procedure missed eleven days of Google News stories during the eighteen-and-a-half-month period: December 29-31, 2006, and July 13-20, 2007. We exclude these from the analysis. LexisNexis returns no articles for these periods from the *New York Times* or the eight major papers, and produces no news events.

30. The search protocol included the following provisions: (a) To identify stories on nanotechnology, we searched in Google News using

"nanoscience"; or "nanoscale"; or "nanotechnology"; or "nano" but not "iPod" or "Apple" or "nanosecond." In LexisNexis, we used "nano!"; or "nano-technology"; and also excluded "Apple" and "iPod" and "nanosecond." All searches examined full text. In the case of Google News, we included stories originally labeled as "omitted" by Google due to potential duplication. (b) To identify stories about politically relevant subjects, we included in the search strings an "and" requirement to at least one occurrence of any from a list of concepts related to politics or political issues.

Because of the absence of wildcard options in Google, we constructed the searches slightly differently. The following is the list of twenty-one pairs of concepts; the first in each pair is the LexisNexis term or terms, and the second, italicized concept is the Google News term or terms. (1) risk!: *risk or risks*; (2) toxic!: *toxic or toxicity*; (3) priva!: *private or privacy*; (4) surveillance!: *surveillance*; (5) disast!: *disaster or disasters or disastrous*; (6) goo or self-replicating or swarm or sludge: *goo or self-replicating or swarm or sludge*; (7) ethic! or unethical: *ethics or ethical or unethical*; (8) regulat! or rules: *regulate or regulation or regulations or regulatory or regulators or rules or regulated*; (9) harm!: *harm or harmful*; (10) danger!: *danger or dangers or dangerous*; (11) health or illness or sick: *health, illness, sick*; (12) environment!: *environment, environmental*; (13) standards: *standards*; (14) law! or legal: *law, laws, legal*; (15) policy: *policy*; (16) govern or governance: *governance, govern*; (17) safe!: *safe, safety*; (18) hazard!: *hazard, hazards, hazardous*; (19) controvers!: *controversial, controversy, controversies*; (20) damage: *damage*; (21) concern\*: *concern, concerns*.

31. See "Google News: How it Works: Oldest articles in Google News," <http://www.google.com/support/news/bin/answer.py?answer=40215&ctx=sibling&topic=8851>. Beginning in June of 2006, Google added a news archive selectively containing news older than thirty days.

32. Audit Bureau of Circulations, "Top 200 Newspapers by Reported Circulation."

33. Daniel Riffe, Stephen Lacy, and Frederick G. Fico, *Analyzing Media Messages: Using Quantitative Content Analysis in Research* (Mahwah, NJ: Lawrence Erlbaum Associates, 1998).

34. For the Google News events, these are the German recall, labeled in Figure 1 as "Recall;" the petition to the Food and Drug Administration ("FDA Petition"); a report by the U.S. National Research Council ("NRC Report"); release of the FDA advisory report ("FDA Report"); the EPA announcement that it would regulate nanoparticles if germicidal claims are made ("EPA"); and the Berkeley regulation ("Berkeley Ordinance").

35. Exceptions are EPA and NRC.

36. We explored procedures for screening results, such as using a two-standard deviation threshold or removing news spikes not associated with specific events, but were not able to produce higher agreement.

37. D. Charles Whitney and Lee B. Becker, "'Keeping the Gates' for Gatekeepers: The Effects of Wire News," *Journalism Quarterly* 59 (spring

---

1982): 60-65.

38. For a variation of this approach, see Jeongsub Lim, "Microlevel Agenda-Setting Effects of News Sites" (paper presented at the annual meeting of the International Communication Association, San Francisco, CA, 2007).

39. Katherine Clegg Smith, Melanie Wakefield, Catherine Siebel, Glen Szczypka, Sandy Slater, Yvonne Terry-McElrath, and Sherry Emery, "Coding the News: The Development of a Methodological Framework for Coding and Analyzing Newspaper Coverage of Tobacco Issues," Research Paper Series, no. 21 (Chicago: ImpactTeen, 2002), [http://www.impactteen.org/generalarea\\_PDFs/Newsmethodspaper\\_smithMA Y2002.pdf](http://www.impactteen.org/generalarea_PDFs/Newsmethodspaper_smithMA Y2002.pdf), accessed May 22, 2008.