

# METHODOLOGY

## Overview

The PEJ News Coverage Index analyzes a wide swath of American news media to identify what is being covered and not covered — the media's broad news agenda. Each week, the Index will issue a report on the top stories across the mainstream news media, as well as a breakdown of how the news agenda that week differed among the media sectors— network TV, for instance, vs. cable or newspapers.

The 2007 News Coverage Index dataset is the summation of a year's worth of coding conducted by PEJ. The coding is continuous throughout the year, with weekly findings reported in the News Coverage Index reports ([Click to read the latest News Coverage Index](#)).

All coding was conducted in-house by PEJ's staff of researchers.

The 2007 analysis totals 70,737 stories. This consists of 6,559 newspaper stories, 6,520 online stories, 21,320 stories from network television, 22,823 stories on cable news, and 13,515 stories from radio programs.

## The Universe: What we Studied

Because the landscape is becoming more diverse — in platform, content, style and emphasis — and because media consumption habits are also changing, even varying day to day, the analysis is designed to be broad. Therefore, our sample, based on the advice of our academic team, was designed to include a broad range of outlets, illustrative but not strictly representative of the media universe.

The sample is also by design, selected to meet these criteria rather than to be strictly random. It is a multistage sampling process that cannot be entirely formulaic or numeric because of differences in measuring systems across media. It involves the balancing of several factors, including the number of media sectors that offer news, the number of news outlets in any given sector, the amount of news programming in each outlet and the audience reach. In addition to front-end selections, we have also weighted the various sectors on the back end to account for differences in audience. The weighting process is discussed further down in this document.

The mainstream, or establishment, daily news media in the United States can be broken down into five main sectors. These are:

- Network TV News
- Newspapers
- Online News Sites
- Cable News
- Radio News

Within each media sector, the outlets and individual programs vary considerably in number, as do the stories and size of the audience. We began by first identifying the various media sectors, then identifying the news media outlets within each, then the specific news programs and finally the stories within those.

The primary aim was to look at the main news stories of the week across the industry. With that in mind, for outlets and publications where time does not permit coding the entire news content offered each day (three hours of network morning programming, for instance), we coded the lead portion. In other words, we coded the first 30 minutes of the cable news programs, the first 30 minutes of the network morning news programs, the front page of newspapers, etc. This may have skewed the overall universe toward more “serious”

stories, but this is also the most likely time period to include coverage of the main, national news events of the day, those that would make up the top stories each week or each month.

Below we describe the selection process and resulting sample for each main sector.

## **Network News**

### Sector Reach

Each evening, the three broadcast news programs on the commercial networks -- ABC, NBC and CBS -- together reach about 26 million viewers. The morning news shows on those networks are seen by 13.6 million viewers.<sup>1</sup> In addition, the nightly public television newscast on PBS reaches 2.2 million viewers daily, according to its internal figures. Because the universe of national broadcast channels is limited to these four outlets, it is practical to count all of the networks as our sample universe.

### Sector Sample

Each of the three major commercial broadcast networks produces two daily national general interest news shows, one in the morning (such as Good Morning America) and one in evening. It is practical, therefore, to include at least part of all these news programs on ABC, CBS, and NBC in our sample. (The magazine genre of programs was not included in the universe both because in most cases they are not daily, except for Nightline, and because they are not devoted predictably to covering the news of the day). At the same time, the NewsHour With Jim Lehrer on PBS is considered by many as an alternative nightly news broadcast compared to the three major networks, and because it reaches a substantial audience, we included that program.

### Units of Study

For the evening newscasts, the study coded the entire program. For the morning programs, it coded the news segments that appear during the first 30 minutes of the broadcast, including the national news inserts but not the local ones. By selecting this sample of the morning shows, it is possible that we will be missing some news stories that appear later in the programs. Through prior PEJ research, however, we have learned that the morning shows generally move away from the news of the day after the first 30 minutes, save for the top-of-the-hour news insert, and present more human interest and lifestyle stories after that. The stories that the networks feel are most important will appear during the first 30 minutes and be included in our study.

The resulting network sample is:

Commercial Network Evening News: Entire 30 minutes of all three programs each day (90 minutes)

Commercial Network Morning News: First 30 minutes of all three programs each day (90 minutes)

PBS NewsHour: First 30 minutes each day

Total: 3.5 hours each day

## **Cable Television**

### Sector Reach

According to ratings data, the individual programs of the three main cable television news channels -- CNN, MSNBC and Fox News -- do not reach as many viewers as those of the broadcast commercial network news shows. During prime time hours, 2.5 million viewers watch cable news, while 1.5 million watch during

daytime hours.<sup>2</sup> But ratings data arguably undercount the reach of cable news. Survey data now find that somewhat more people cite cable news as their first source for national and international news than cite broadcast network news.

### Sector Sample

The most likely option was to study CNN, MSNBC and Fox News. These represent the dominant channel of programming from each news-producing cable company. (This means selecting MSNBC as opposed to CNBC, and CNN as opposed to CNN Headline News, and MSNBC over Headline News, which now sometimes beats MSNBC in ratings.)

### Units of Study

Since these channels provide programming round the clock, with individual programs sometimes reaching fairly small audiences, it was not practical for us to code all of the available shows. There is a great challenge in selecting several times out of the day to serve as a sample of cable news over all.

On the other hand, earlier studies have shown that for much of the day, one cable news program on a channel is indistinguishable to most people from another. If one were to ask a daytime viewer of cable news which program he or she preferred, the 10 a.m. or the noon, you might get a confused look in response. For blocks of hours at a time, the channels will have programs with generic titles CNN Newsroom, Your World Today or Fox News Live. Our studies have shown that there are four distinct programming sectors to cable - early morning, daytime, early evening and prime time.

Working with academic advisers we weighed various options. A selection based on the most-watched programs would result in the O'Reilly Factor (1.8 million viewers a night) for Fox and Larry King Live (500,000 viewers a night) for CNN. However, some of these shows are not news programs per se, but rather their content derives from the host's opinions and guests on any given day. Separating news and talk also proved problematic because it is often difficult to distinguish between the two categories, while several programs offer both news and talk in the same hour.

The best option, we concluded, was to draw from two time periods:

1. The daytime period, to demonstrate what live events are being covered. The study includes two 30-minute segments of daytime programming each day, rotating among the three networks.
2. Early evening and prime time (6 to 11 p.m.), together as a unit, rather than separating out talk and news or early prime and late prime. Within this five-hour period, we included all programming that focuses on general news events of the day. Basically, this removes three programs: Fox's Greta Van Susteren, which is focused on crime; CNN's Larry King, which is focused on entertainment or personal stories rather than news events, and MSNBC's documentary program. Because MSNBC's audience numbers are so much lower than those for Fox or CNN, we also decided to include slightly less of its programming. Even though CNN trails Fox in Nielsen ratings, its monthly cumulative, or "cume," audience figure is higher, so the two are sampled equally.

To include the most cable offerings possible each week, the study coded the 30 minutes of selected programs and rotated them daily. Morning shows were not included because those shows are run at the same time for every part of the country, meaning that a broadcast that starts at 7 a.m. on the East Coast will begin at 4 a.m. on the West Coast. Those programs appear far too early for much of the country to actually view. This is in contrast with the broadcast morning programs, which are shown on tape delay in different parts of the country, in the manner of other broadcast programs.

This process resulted in the following cable sample:

## Daytime

Rotate, coding two out of three 30-minute daytime slots each day (60 minutes a day)

## Prime Time

Three 30-minute segments for Fox (90 minutes)  
Three 30-minute segments for CNN (90 minutes)  
Two 30-minute segments for MSNBC (60 minutes)

The index rotates among all programming from 6 to 11 p.m. that was focused on general news events of the day, excluding CNN's Larry King and Fox's Greta Van Susteren.

Both CNN and MSNBC made some programming changes during 2007, and the replacement shows are included when appropriate.

	CNN	Fox	MSNBC
6 p.m.	Situation Room	Special Report w/Brit Hume	Tukcer Carlson
7 p.m.	Lou Dobbs Tonight (was previously at 6 p.m.)	Fox Report w/ Shephard Smith	Hardball
8 p.m.	Paula Zahn Now/Out in the Open	The O'Reilly Factor	Countdown w/Keith Olbermann
9 p.m.	-----	Hannity & Colmes	Scarborough Country/Live w/ Dan Abrams
10 p.m.	Anderson Cooper 360	-----	-----

= 5 hours each day (including daytime)

## Newspapers

### Sector Reach

About 54 million newspapers are sold each weekday.<sup>3</sup> This number does not include the “pass along” rate of newspapers, which some estimate, depending on the paper, to be approximately three times the circulation rate. In addition, specific newspapers, such as the New York Times and Washington Post, have an influence on the national and international news agenda even greater because they serve as sources of news that many other outlets look to in making their own programming and editorial decisions. So while the overall audience for newspapers is declining over recent years, newspapers still play a large and consequential role in setting the news agenda that cannot be strictly quantified or justified by circulation data. There is a growing body of data that the total audience of newspapers, combining their reach in print and online combined, may actually be growing slightly.

### Sector Sample

To create some representation of what national stories are being covered by the 1,450 daily newspapers around the country, we divided the country's daily papers into three tiers based on circulation: over 650,000, 100,000 to 650,000, and under 100,000. Within each tier, we selected papers along the following criteria:

Papers needed to be available electronically the day of publication. Three Web sites, including [www.nexis.com](http://www.nexis.com), [www.newsstand.com](http://www.newsstand.com) and [www.pressdisplay.com](http://www.pressdisplay.com), offer same-day full-text delivery service. Based on their general same-day availability (excluding non-daily papers, non-U.S. papers, non-English-language papers, college papers, and special niche papers) a list of U.S. general interest daily newspapers was constructed. The list included seven papers in Tier 1, 44 papers in Tier 2, and 22 papers in Tier 3.

Tier 1: Due to its national prominence and readership, and the desirability of having at least one newspaper that was coded every day without any interruption due to rotation (in the same way the network newscasts are coded), we decided to code the New York Times each day (Sunday through Friday). We then wanted to include a representation from the other large nationally known or distributed papers, so each day we coded two out of four of the largest papers, the Washington Post, the Los Angeles Times, USA Today, and the Wall Street Journal.

Tier 2 and 3: Four newspapers were selected each from Tier 2 and Tier 3. To ensure geographical diversity, each of the four newspapers within Tier 2 and Tier 3 was selected from a different geographic region according to the parameters established by the U.S. Census Bureau, i.e., Northeast Region, Midwest Region, South Region and West Region. An effort was also made to ensure the ownership diversity. One selected newspaper was found too difficult to capture during our testing, and it was replaced by another newspaper from the same region within the same circulation category. We rotated two of the four newspapers in Tier 2 and Tier 3 each day.

This process resulted in the following newspaper sample:

#### **First Tier**

The New York Times

The Washington Post

The Los Angeles Times

USA Today

The Wall Street Journal

#### **Second Tier**

The Boston Globe

The ( Minneapolis) Star Tribune

The Austin ( Texas) American-Statesman

The Albuquerque Journal

### **Third Tier**

The ( Attleboro, Mass.) Sun Chronicle

The ( Ashtabula, Ohio) Star Beacon

The Chattanooga Times Free Press

The Bakersfield Californian

### **Units of Study**

For each of the papers selected, we coded only articles that began on page A1 (including jumps). The argument for this is that the papers have made the decision to feature those stories on that day. That means we did not code the articles on the inside of the A section, or on any other section. The first argument for ignoring these stories is that they will be unnecessary for our index, which measures only the biggest stories each week. If a story appears on the inside of the paper, but does not make A1 at any point, it would almost certainly not be a big enough story to make the top list of stories we will track each week. The weakness of this approach, arguably, is that it undercounts the full news agenda of national and international news in that it neglects those stories that were not on Page 1 on certain days but were on others. While this is less pertinent in the weekly index, perhaps, at the end of the year, when trying to assess the full range of what the media covered, the stories that spent time on the inside of the paper but did not disappear were undercounted.

Part of the reasoning for excluding those national and international stories that begin inside the front section of the paper is practical. Coding the interior of the papers to round out the sample for year end purposes is an enormous amount of work for relatively minimal gain.

The other argument for forgoing national and international stories that fail to make Page 1 is more conceptual. We were measuring what newspapers emphasize, their top agenda. Given the cost versus the benefit, capturing the front page of more newspapers seemed the better alternative. In the same regard, we were not coding every story that might appear on a Web site, an even more daunting task, and recorded just the top stories.

The other challenge with newspapers that we did not face with some other media is that we will only include stories that are national or international. National is defined as a story being covered by newspapers from different locations, as opposed to a local story that is only covered in one paper. The only local stories included in the study are those that are pertinent to a larger national issue — how the war in Iraq is affecting the hometown, for instance, or new job cuts at the local industries because of the sliding economy.

This resulted in the newspaper sample of about 25 stories a day.

### **Online News Sites**

#### **Sector Reach**

About 30 million Internet users go online for news each day.<sup>4</sup> About 6.8 million people read some blog each day, some of the most popular of which are news-oriented.<sup>5</sup> Both online news sites and blogs are becoming more important in the overall news agenda. Any sample of the modern news culture must include representation of some of the more popular examples of these sectors.

### Sector Sample

The online news universe is even more expansive than radio and has seemingly countless sites that could serve as news sources. To get an understanding of online news sources, we chose to include several of the most popular news sites in our universe as a sample of the overall online news agenda. We also wanted balance in the type of online news sites, between those that produced their own content and those who aggregated news from throughout the Web.

To choose the sites we were to include in our sample, we referred to the list of the top 10 news sites with the most unique visitors according to the Nielsen/NetRatings from September 2006. Out of that list, we choose five sites that represented a mix of sites that either created their own material for their Web site ([MSNBC.com](http://MSNBC.com), [CNN.com](http://CNN.com)), popular sites that aggregated material from other Web sites ([Google News](http://Google News), [Yahoo News](http://Yahoo News)), and a site such as AOL News, which usually uses material from news wire services but also creates some unique material at times as well.

The sites that were coded were as follows:

Site	Unique Audience (000) <sup>6</sup>
Yahoo News	90,162
MSNBC.com	26,745
CNN.com	24,676
AOL News	18,646
Google News	9,425

### Units of Study

For the online news sites, the study captured the site once a day from 9 to 10 a.m. Eastern time. For each site capture, we coded the top five stories, as those are the most prominent as determined at that point in time by the particular news service. As is true with our decision about page A1 in newspapers, if a story is not big enough for the online sites to highlight it in their top five stories, it is likely not a story that will register on our tally of the top stories each week.

This resulted in a sample of 25 stories a day.

## **Radio**

### Sector Reach

Radio is a diverse medium that reaches the majority of Americans – 94 percent of Americans 12 years and older listen to traditional radio each week.<sup>7</sup> Approximately 16% of radio listeners tune into news, talk and information radio in an average week, which ranks it as the most popular of all measured radio formats.<sup>8</sup> Many more Americans get from news headlines while listening to other formats as well.

The challenge with coding national radio programs is that much of radio news content is localized, and the number of shows that reach a national audience is only a fraction of the overall programming. On the other hand, our content analysis of radio confirms that news on commercial radio in most cities has been reduced

to headlines from local wires and syndicated network feeds, plus talk, much of which is nationally syndicated itself. The exception is in a few major cities where a few all-news commercial radio stations still survive, such as Washington, D.C., where WTOP is a significant all-news operation.

### Sector Sample

The Index includes three areas of radio news programming.

1. Public radio: The index includes 30 minutes of a public radio's broadcast of National Public Radio's morning program, Morning Edition, each day. Note: NPR produces two hours of Morning Edition each day, which also includes multiple news roundups produced by a different unit of NPR. Member stations may pick any segments within those two hours and mix and match as fits their programming interests. Thus, what airs on a member station is considered a "co-production" of NPR and that member station rather than programming coming directly from NPR.

2. Talk radio: The index includes some of the most popular national talk shows that are public affairs or news oriented. Since the larger portion of the talk radio audience, and talk radio hosts, are politically conservative, we included more conservative than liberal hosts.

Each day we coded the first 30 minutes of Rush Limbaugh's radio show, which is heard by more listeners than any other talk show, according to Talkers Magazine (December 2006). We also rotated each day between the two next most popular conservative shows, Sean Hannity and Michael Savage. Since the liberal audience for talk radio is smaller, we only coded one liberal talk show a day, rotating daily between Ed Schultz and Randi Rhodes, two of the top liberal radio hosts based on national audiences. The Arbitron ratings, according to Talkers Magazine online, for spring 2006 are as follows:

Minimum Weekly Cume (in millions, rounded to the nearest 0.25, based on Spring '06 Arbitron reports)<sup>9</sup>

Rush Limbaugh 13.50

Sean Hannity 12.50

Michael Savage 8.50

Ed Schultz 2.25

Randi Rhodes 1.25

Alan Colmes 1.25

3. Headline feeds: Hourly news feeds from national radio organizations like CBS and CNN appear on local stations across the country. These feeds usually last five minutes at the top of each hour, and are national in that people all over the country get the same information. They frequently supplement local talk and news shows.

To get a representation of these feeds, we coded two national feeds, each twice a day (9 a.m. and 5 p.m. Eastern time). The networks that were included were CBS Radio and ABC Radio. The stations that were used to capture the CBS Radio headlines, primarily KKZN in Colorado and WCBS in New York City, each generally carried five minutes of CBS Radio headlines, while the stations used primarily to capture the ABC Radio headlines, WMAL in Washington, D.C., and WLS in Chicago, generally carried two minutes of syndicated news headlines.



The stations that were used to capture each program were selected based on the availability of a solid feed through the stations' Web sites. We also compared their shows to that of other stations to ensure that the same edition was aired on that station as with others carrying the same program.

This resulted in the following sample:

News: 30 minutes of NPR's Morning Edition each day, as broadcast on a selected member station.

Headlines: Four headline segments each day (two from ABC Radio and two from CBS Radio), about 14 minutes.

Talk: The first 30 minutes of three talk programs each day -- two conservative (Rush Limbaugh and either Sean Hannity or Michael Savage) and one liberal (Ed Schultz or Randi Rhodes). About 2.5 hours a day.

## **Universe of Outlets**

### **Newspapers (Sunday to Friday)**

The New York Times every day

Coded two out of these four every day

The Washington Post

The Los Angeles Times

USA Today

The Wall Street Journal

Coded two out of these four every day

The Boston Globe

The ( Minneapolis) Star Tribune

The Austin ( Texas) American-Statesman

The Albuquerque Journal

Coded two out of these four every day

The Sun ( Attleboro, Mass.) Chronicle

The ( Ashtabula, Ohio) Star Beacon

The Chattanooga Times Free Press

The Bakersfield Californian

**Web sites (Monday to Friday)**

CNN.com

Yahoo News

MSNBC.com

Google News

AOL News

**Network TV (Monday to Friday)***Morning shows*

ABC – Good Morning America

CBS – Early Show

NBC - Today

*Evening news*

ABC – World News Tonight

CBS – CBS Evening News

NBC – NBC Nightly News

PBS – NewsHour with Jim Lehrer

**Cable TV (Monday to Friday)***Daytime (2 to 2:30 p.m.)\* – coded two out of three every day*

CNN

Fox News

MSNBC

(\* From January 1 to March 16 2007, coded from 1-1:30pm EST; from March 19 2007 on, coded from 2-2:30pm EST)

Nighttime CNN – coded three out of the four every day

Lou Dobbs Tonight

Situation Room (6 p.m.)

Paula Zahn Now/Out in the Open

Anderson Cooper 360

Nighttime Fox News – coded three out of the four every day

Special Report With Brit Hume

Fox Report With Shepard Smith

O'Reilly Factor

Hannity & Colmes

Nighttime MSNBC – coded two out of the four every day

Tucker Carlson (6 p.m.)

Hardball (7 p.m.)

Countdown With Keith Olbermann

Scarborough Country/Live With Dan Abrams

**Radio (Monday to Friday)**

*News Radio*

Headlines - every day

ABC Radio headlines at 9 a.m. and 5 p.m.

CBS Radio headlines at 9 a.m. and 5 p.m.

NPR's Morning Edition – 5:00-5:30 a.m. as broadcast on an East Coast member station.

*Talk radio*

Rush Limbaugh every day

One out of two additional conservatives each day:

Sean Hannity

Michael Savage

One out of two liberals each day:

Ed Schultz

Randi Rhodes

That resulted in 35 outlets each weekday, and seven newspapers were included only on Sundays.

## **Universe Procurement and Story Inclusion**

### **Newspapers**

For each of the seven newspapers included in our sample, we coded all stories where the beginning of the text of the story appears on the front page of that day's hard copy edition. If a story only had a picture, caption or teaser to text inside the edition, we did not include that story in our sample. We coded all stories that appeared on the front page with a national or international focus. Because we were looking at the coverage of national and international news, if a story was about an event that was solely local to the paper's point of origination, we excluded such a local story from our sample. The sole exception to this rule was when a story with a local focus was tied to a story that we determined to be a "Big Story" – defined as one that has been covered in multiple national news outlets for more than one news cycle. For example, a story about a local soldier who has come back from the Iraq War has a local angle but is related to a national issue and was important in the context of our study.

We coded the entirety of the text of all the articles we include. If an article included a jump to an inside page in the hard copy edition, we coded all the text including that which makes up the jump.

When possible, we subscribed to the hard copies of the selected newspapers and had them delivered to our Washington office. This was possible for national papers that have same-day delivery (the New York Times, the Washington Post, the Wall Street Journal and USA Today). For these papers, we used the hard copy edition to determine the placement on the front page of the edition and to get all the text we coded. One element of the hard-copy stories we were not able to achieve easily with those editions was the word count of each article. For that, we used the LexisNexis computer database to determine the word count for each of the stories coded.

For all of the other papers that we are not able to get hard copies of within the same day of publication, we took advantage of Internet resources that have digital copies of the hard copy editions. Pressdisplay.com and Newsstand.com offer services where we could subscribe to digital versions of the hard copy and get them the same day. From these digital versions, we got the text of the relevant articles and also determined the word counts. To get the word counts, we copied the text of the articles (not including captions, headlines, bylines or pull-out quotes) into the Microsoft Word software program and ran the word count function to get the number. When necessary, we went to the paper's Web site in order to find the text of articles that were not available on either of the two Web services. Through prior experience and through examination of each individual article, we were able to determine when the text of the article on the Web site was the same as it would be on the hard copy of the paper.

### **Network and Cable Television**

For television programs, we coded the first 30 minutes of the broadcast, regardless of how long the program lasts. As with newspapers, we coded all stories that are news reports that relate to a national or international issue. Therefore, we did not code stories that are part of a local insert into a national show. For example, NBC's Today Show cuts each half-hour to a local affiliate that will report local stories and weather. We did not include those local insert stories.

We also excluded from our sample commercials, promos and teasers of upcoming stories. We were interested only in the actual reporting that takes place during the broadcasts.

Any story that fit the above criteria and began within the first 30 minutes was included in the study, even if the story finished outside of the 30-minute time period. A three-minute story that began 28 minutes into a program was coded in its entirety, even though the final minute ran after our 30-minute cutoff mark. The exception to this rule was when a television station showed a speech or press conference that ran longer than the 30-minute period (often much longer). In those cases, we cut off the coding at the 30-minute mark in order to prevent that event from unduly impacting our overall data.

The method of collection of all television programs was the same. PEJ is a subscriber to the DirectTV satellite television service and we have nine TiVo recording boxes hooked up to the DirectTV signals. Through these TiVo boxes, we digitally recorded each broadcast and then archived the programs onto DVDs. There is redundancy in our recording method so that each show is recorded on two machines to prevent an error from causing problems in our capture.

Occasionally outlets deviated from the regularly scheduled news programs. When the deviation was a special program *produced by that news outlet* (such as "A Special Look Into Iraq With Wolf Blitzer"), we coded the same 30-minute period that we would have coded otherwise in order to get a realistic account of what news would be available to the consumer who tuned in at that time. However, when a show was pre-empted for a special *live* event, such as a presidential campaign debate or the State of the Union address, we did not include that period as part of our sample.

There were 261 weeknights in our 2007 sample of the network nightly newscasts. On three nights (January 1, November 23 and November 25), ABC did not air its regular nightly newscast, and on two nights (November 22 and 23), CBS did not air its regular nightly newscast. On most occasions, the newscasts were pre-empted by sporting events.

## **Radio**

The rules for capturing and selecting stories to code for radio were very similar to television. We coded the first 30 minutes of each show regardless of how long the show lasts. We also excluded local inserts from local affiliates and continued coding any story that ran past the 30-minute mark.

For each of the radio shows selected, we found national feeds of the show that were available on the Web. As with television, we have two computers capturing each show so as to avoid errors if one feed was not working. The actual recording is done using a software program called Replay A/V, which captures the digital feeds and creates digital copies of the programs onto our computers. We then archived those programs onto DVDs.

## **Online**

For each of the Web sites we included in our sample, we captured and coded the top five stories that appeared on the site at the time of capture. Our captures took place from 9 to 10 a.m. Eastern time each weekday. The captures physically occurred with a coder going to each site using an Internet browser and saving the home page and appropriate article pages to our computers, exactly as they appeared in our

browsers at the time of the capture. We relied on people rather than a software package to capture sites because some software packages interfere with the operation of Web sites.

As with newspapers, some stories are longer than one Web page. In those cases, we included the entire text of the article for as many Web pages that the article lasted.

Because each Web site is formatted differently, we came up with a standard set of rules to determine which stories were the most prominent on a given home page. We spent a significant amount of time examining various popular news sites and discovered patterns that led us to the best possible rules. We ignored all advertisements, audio/visual features, or extra features on the sites that were not news reported stories. We were interested only in the main channels of the Web sites where the lead stories of the day were displayed. We determined the top "lead" story. That was the story with the largest font size for its headline on the home page. The second most prominent story was the story that had a picture associated with it, if that story was different than the story with the largest headline. By considering many sites, we realized that a number of sites put pictures with stories they find particularly interesting but are clearly not intended to be the most important story of the day. We wanted those stories to be in our sample, however, because the reader's eye is often drawn to them.

Having figured out the first and second most prominent stories, we then relied on two factors to determine the third. We considered the size of the headline text and then the height on the home page. Therefore, for determining the third most prominent story, we looked for the story with the largest headline font after the top two most prominent stories. If there are several stories with identical font sizes, we determined that the story that is higher up on the page is more prominent. In cases where two articles had the same font size *and* the same height on the screen, we chose the article to the left to be the more prominent.

## **Coding Procedures and Inter-coder Reliability**

A coding protocol was designed for this project based on PEJ's previous related studies. Key variables include story ID number, story date, source, broadcast start time, headline, story word count, duration in seconds, placement/prominence, story format, big story, sub-storyline, geographic focus, broad story topic, lead newsmaker, and campaign mention.

Variable source includes all the outlets we coded. Variable broadcast start time applies to radio and TV broadcast news and gives the starting time of the program in which the story appears. Variable headline determines whether the story is part of a regular news roundup segment. Variable story word count designates the word count of each individual print/online news story. Duration in seconds is the amount of time for a broadcast story only, calculated by subtracting story start time from story end time. Variable placement/prominence designates where stories are placed within a publication, on a Web site, or within a broadcast. The location reflects the prominence given the stories by the journalists creating and editing the content. Story format measures the type and origin of the text-based and broadcast stories, which designates, at a basic level, whether the news story is a product of original reporting or drawn from another news source. Big stories are particular topics that occurred often in news media during the period under study. Sub-storyline applies to stories that fit into some of the long-running big stories or other common storyline, reflecting aspects, features or development of some big stories. Variable geographic focus concerns the geographic area to which the topic is relevant in relation to the location of the news source. Variable broad story topic determines the type of broad topic categories addressed by a story. The variable for lead newsmaker names the person or group who is the central focus of the story (This variable was added to the coding scheme on July 1, 2007). Variable campaign mention determines whether the story has any mention at all of a U.S. campaign or election.

The team responsible for performing the content analysis was made up of 13 coders, a coding administrator and a senior research methodologist. Five of coders have been trained extensively since the summer of 2006.

Two major tests of intercoder reliability were conducted over the past year in order to ensure accuracy among all the coders. The first test was conducted just prior to the launch of the weekly News Coverage Index reports beginning in January 2007.

For the first test, two random datasets were combined for running intercoder reliability statistics. In total, 126 stories from 16 outlets within various media categories (newspapers, online, television, radio) were randomly selected to assess the reliability, resulting in roughly 9% of one week's total story count (one week's total story count ranges from 1,200 to 1,400). Over all, 22 stories were coded by seven coders (17%), 66 stories were coded by six coders (52%), and 38 stories were coded by five coders (30%).

For the more difficult or subjective variables, including story format and broad story topic, we conducted furthering testing. Ninety stories from 10 outlets within various media sectors were randomly selected for reliability assessment for these variables as well as for a newly added variable campaign mention. Most of the coders coded all of these stories. Data shows that the percent of agreement for all the variables in the index were above 80.

A second test of intercoder reliability was conducted in April and May of 2007. For this test of intercoder reliability, we divided the variables into two groups and tested those variables separately as to ensure the accuracy for each variable.

#### Housekeeping Variables

The first group of variables we tested for was called housekeeping variables. These are variables that are necessary for each story, but often involve little to no subjectivity from the coder. For this test, we selected a random sample of 151 stories from each of the five media sectors we covered. Each story was coded by two coders. This represented more than 10% of the number of stories we code in a given week.

Of those stories, 32 were print stories (newspaper and online) and 119 were broadcast stories (television and radio). For our housekeeping variables, we achieved the following levels of agreement:

##### Print (32 cases)

Story Date: 100%

Source: 100%

Story word count (+/- 20 words): 90%

Placement: 94%

##### Broadcast (119 cases)

Story Date: 100%

Source: 100%

Broadcast start time: 100%

Headline: 100%

Story start time (+/- 6 seconds): 91%

Placement: 93%

Story end time (+/- 6 seconds): 92%

#### Main Variables

The second group of variables we tested are referred to as the main variables, and are variables that take more training to utilize and comprehend. Having already demonstrated that we had a high level of agreement for all of our housekeeping variables, we then had the coders participate in an additional test to determine the level of agreement for these main variables.

We randomly selected 116 stories from both print and broadcast mediums, which represented about 8% of the stories we code in a typical week. The level of agreement for each of our key variables was as follows:

Format: 89%  
Big Story: 91%  
Sub-storyline: 87%  
Geographic Focus: 91%  
Topic: 85%

Lead newsmaker: 83%

All the percentages of agreement for the above variables were calculated using a software program available online called PRAM (See "The Content Analysis Guidebook," by Kimberly A. Neuendorf, Sage Publications, 2002.).

Throughout 2007, as new coders were hired and included in the coding process, they were given extensive training from both the coding administrator and from other experienced coders. New coders were not allowed to participate in the weekly coding for the News Coverage Index until they had demonstrated a level of agreement with experienced coders for all variables at an 80% level or higher.

Each coder worked between 20 and 37.5 hours a week in our Washington and was trained to work on all the mediums included in the sample. The schedule for each coder varied, but since all of the material included in the index is archived, the actual coding can be performed at any point during the week.

To achieve diversity in the coding and statistics, generally no one coder coded more than 50% of a particular sector within one week. And each coder coded at least three mediums each week. In the case of difficult coding decisions about a particular story, the final decision was made by either the coding administrator or a senior member of the PEJ staff.

The physical coding data was entered by the coders into a proprietary software program that has been written for this project by PEJ Phase II Technologies. The software allows coders to enter the data using the appropriate conditions for each variable, and also allows coders to review their work and correct mistakes when needed. The same software package compiles all of the coding data each week and allows us to perform the necessary statistical tests.

### **Total Media Combined: Creation and Weighting**

The basis of measurement for top stories is time in broadcast and cable and words in text-based media. Thus for cable news, for example, we refer to the percent of total seconds that a certain story received. In other words, of all the seconds analyzed in cable news this week, ground events in Iraq accounted for xx% (or xx seconds out of a total of xxx). The industry term for this is "newshole" — the space given to news content.

The main index considers broadcast and print together, identifying the top stories across all media. To do this, words and seconds are merged together to become total newshole. After considering the various options for merging the two, the most straightforward and sensible method was to first generate the percent



of newshole for each specific medium. This way all media are represented in the same measurement — percent.

Next, we needed to create a method for merging the various percentages. There were several options. We could have run a simple average of all five. We could have averaged all print and all broadcast and then average those two? Or should we apply some kind of weight based on apparent audience?

Because each medium measures its audience differently (ratings per month in television, daily circulation in newspapers, unique visitors in online), it is nearly impossible to create a reliable system based on audience figures. Nonetheless, several of our advisers thought some kind of weight should be applied. Various options were considered, including a combination of different metrics, such as actual data alongside supplemental survey data. One consistent measure is that of public opinion surveys. The same question is posed about multiple media. Two such questions are asked regularly by the Pew Research Center for the People & the Press. One asks about “regular usage” and the other asks where people go for “national and international news.”

Before arriving at a method, we tested multiple models:

Model 1: Compile percentages for big stories for each of the five media sectors (newspapers, online sites, network television, cable television and radio), and then average those five lists into one final list.

Model 2: Divide the media sectors into two groups, text-based media (newspapers, online sites) and broadcast (network television, cable television, and radio). Average the lists of percentages between the two groups to get one final list.

Model 3: Compile percentages for big stories for each of the five media sectors, and then add the weighted five lists together into one final list. The weights given to each media sector were calculated by averaging three most recent survey data in terms of where people get news about national and international issues, collected by the Pew Research Center for the People & the Press (August 2006, November 2005 and June 2005). First, we take the average response for each media across the three time periods. Next, we rebalance the average percents to match the five media sectors in the index — newspapers, Internet, network television, cable television and radio — to equal 100%. In this calculation, the weight for newspapers is 0.28, for Internet 0.16, for network television 0.18, for cable television 0.26, and for radio 0.12.

Model 4: Compile percentages for big stories for each of the five media sectors, and then add the weighted five lists together into one final list. The weights assigned to each media sector were generated based on the regularly media usage survey data, collected by the Pew Research Center for the People & the Press in its Biennial Media Consumption Survey 2006. In this, the weight for newspaper is 0.307, for Internet 0.218, for network television 0.165, cable television is 0.201 and radio 0.109.

By testing two trial weeks’ data, we found that the lists of top five stories were exactly the same (top stories’ names and their ranks) using all four of these models, although some percentages varied. In the end, the academic and survey analysts on our team felt the best option was Model 3. It has the virtue of tracking the media use for national and international news, which is what the index studies. Thus, we adopted Model 3 and plan to update the weights when appropriate.

## Footnotes

1. Nielsen Media Research, used under license.
2. PEJ Analysis of Nielsen Media Research data, used under license
3. 2004 Editor and Publisher Yearbook.
4. Pew Internet, December 2005 survey
5. Ibid.
6. Nielsen/NetRatings available at <http://www.cyberjournalist.net/news/003791.php>
7. Arbitron ratings, December 2004
8. Ibid.
9. Talkers Magazine [online](#).