

## Methodology

The PEJ News Coverage Index (NCI) 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 measures 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 2012 News Index dataset is the summation of five months of coding conducted by PEJ. The coding was done in close to real-time and weekly findings were reported on <http://www.journalism.org/>

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

The 2012 coding includes stories from January to May of that year.

The 2012 analysis for the NCI totals 20,447 stories for the first five months of the year. This consists of 1,977 newspaper stories, 3,242 online stories, 5,186 stories from network television, 6,472 stories on cable news, and 3,570 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 Index is designed to be broad. Therefore, our sample, based on the advice of our academic team, is designed to include a broad range of outlets, illustrative but not strictly representative of the media universe.

The sample is also a purposive one, 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 also weight the various sectors on the back end to account for differences in audience. The weighting process is discussed below 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 number of outlets and individual programs vary considerably, as do the number of 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 of the Index is 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 code the lead portion. In other words, we code 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 skew the overall universe toward more “serious” stories, but this is also the most likely time period to include coverage of the “main” 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.

**Note:** The statistics cited here are the statistics that were accurate at the time of the launch of the Index (January 2007). When available, updated data are included in the endnotes section.

## **Network News**

### **Sector Reach**

Each evening, the three broadcast network news programs on ABC, NBC and CBS reach about 27 million viewers. The morning news shows on those networks are seen by 14.1 million viewers.[1] In addition, the nightly newscast on PBS reaches 2.4 million viewers daily, according to its internal figures. Because the universe of national broadcast networks is limited to these four, it is practical to include all of the networks in our sample universe.

### **Sector Sample**

Each of the three major broadcast networks produces two daily national general interest news shows, one in the morning (such as Good Morning America) and one in the evening. It is practical, therefore, to include at least part of all these news programs on ABC, CBS and NBC in our sample. At the same time, because the PBS NewsHour is considered by many as an alternative nightly news broadcast to the three major networks, and it reaches a substantial audience, we also include that program.

### **Units of Study**

For the commercial evening newscasts, the study codes the entire program. For the morning programs, it codes the news segments that appear during the first 30 minutes of

the broadcast, including the national news inserts but not local inserts. 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 point. The stories that the networks feel are most important will appear during the first 30 minutes and are included in our study.

For PBS NewHour, where the second half of the program differs from the first half, we began rotating beginning March 31, 2008, between the first and second half-hours of the show in order to get a closer representation of the program's overall content.

The resulting network sample is as follows:

Commercial Evening News: Entire 30 minutes of 2 out of 3 programs each weekday (60 minutes).

Commercial Morning News: First 30 minutes of 1 or 2 out of 3 programs each weekday (30 or 60 minutes).

PBS NewsHour: Rotate between first and second 30 minutes each day, then skip a day.

The combination of morning and evening news broadcasts results in 1.5 to 2.5 hours of programming 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 network news shows. During prime-time hours, 2.7 million viewers watch cable news, while 1.6 million watch during daytime hours.[2] But ratings data arguably undercount the reach of cable news. Survey data now find that somewhat more people cite cable news as their primary or first source for national and international news than they do 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.)

## Units of Study

Since these channels provide programming round the clock, with individual programs sometimes reaching fairly small audiences, it is not practical for us to code all of the available shows. On the one hand, there is a great challenge in selecting several time periods out of the day to serve as a sample of cable news overall.

On the other hand, earlier studies have shown that for much of the day, most people find one cable news program on a channel to be indistinguishable 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, the questioner might get a confused look in response. For blocks of hours at a time, the channels will have programs with generic titles such as 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. [3] Some of these shows, however, are not news programs per se, but rather have content that 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, and 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 continuing or live events are being covered. The study includes two 30-minute segments of daytime programming each day, rotating among the three channels.
- 2) Early evening and prime time (5 p.m.-11 p.m.) together as a unit, rather than separating out talk and news or early prime and late prime. Within this six-hour period, we included all programming that focuses on general news events of the day. Basically, this removes programs such as Fox's Greta Van Susteren, which is more narrowly focused on crime; CNN's Piers Morgan Live, which as often as not is focused on entertainment or personal stories rather than news events.

Prior to January 1, 2009, three out of four evening cable programs were coded each evening for both Fox News and CNN. For MSNBC, two out of four evening programs were coded each evening. These shows were rotated each weekday. In the past, MSNBC's ratings were significantly lower than the ratings for Fox News and CNN, and that was the justification for including one fewer of their shows each evening.

At the start of 2009, PEJ made a change to the evening cable sample. MSNBC now beats CNN in ratings, and all three channels have seen significant changes in ratings over the years. Therefore, it becomes harder to justify having a different amount of programming for the three stations. Consequently, beginning on January 1, 2009, all three stations have

two of their four evening cable shows coded on a nightly basis. On May 30, 2011, we started to rotate and code 1 or 2 evening shows from CNN and MSNBC.

To include the most cable offerings possible each week, the study codes the first 30 minutes of selected programs and rotates 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 to 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**

Two 30-minute segments for Fox News (60 minutes)

One or two 30-minute segments for CNN (30 or 60 minutes)

One or two 30-minute segments for MSNBC (30 or 60 minutes)

The Index rotates among all programming from 5 p.m. to 10 p.m. that was focused on general news events of the day.

	CNN	Fox News	MSNBC
5 p.m.	Situation Room	-----	-----
6 p.m.	John King, USA	Special Report with Bret Baier	PoliticsNation
7 p.m.	Erin Burnett Outfront	Fox Report with Shephard Smith	Hardball with Chris Matthews
8 p.m.	Anderson Cooper 360	The O'Reilly Factor	The Ed Show
9 p.m.	-----	Hannity	The Rachel Maddow Show

This results in 3.5 or 4 hours of cable programming each day (including daytime).

# Newspapers

## Sector Reach

About 54 million people buy a newspaper each weekday. [4] 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, some newspapers, such as the New York Times and Washington Post, have an even greater influence on the national and international news agenda 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 has declined over recent years, newspapers still play a large and consequential role in setting the news agenda that cannot be strictly quantified or evaluated by circulation data alone. There is a growing body of data to suggest that the total audience of newspapers, combining their reach in print and online, may actually be growing slightly.

## Sector Sample

To create a representation of what national stories are being covered by the 1,450 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:

First, papers need to be available electronically on the day of publication. Three websites—[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 nondaily 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 original sample list included seven papers in Tier 1, 44 papers in Tier 2 and 22 papers in Tier 3.

Tier 1: We wanted to include a representation from the large nationally reputed or distributed papers, so we code each week day two of four of the largest papers—the New York Times, the Los Angeles Times, USA Today and the Wall Street Journal. We code 1 out of 4 papers on Sunday.

Tiers 2 and 3: Four and three newspapers were selected from Tier 2 and Tier 3 respectively. To ensure geographical diversity, each of the four newspapers within Tier 2 and three within Tier 3 was randomly selected from a different geographic region according to the parameters established by the U.S. Census Bureau—Northeast Region, Midwest Region, South Region and West Region. An effort was also made to ensure diversity by ownership. In 2012, we rotate two of the four newspapers in Tier 2 and one of the three in Tier 3 each week day, and we code 1 paper from each Tier on Sunday.

### **1st Tier**

New York Times  
Los Angeles Times  
USA Today  
Wall Street Journal

### **2nd Tier**

Washington Post  
The Denver Post  
Houston Chronicle  
Orlando Sentinel

### **3rd Tier**

Traverse City Record-Eagle (MI)  
The Daily Herald (WA)  
The Eagle-Tribune

### **Units of Study**

For each of the papers selected, we code only articles that begin on Page A1 (including jumps). The argument for this is that the papers have made the decision to feature those stories in that day's edition. That means we do not code the articles on the inside of the A section or in any other section.

The first argument for ignoring these stories is that they would 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 track each week. The weakness of this approach, arguably, is that it undercounts the full agenda of national and international news in that it neglects those stories that were not on Page A1 on certain days but were on others. While this is perhaps less pertinent in the weekly Index, at the end of the year, when trying to assess the full range of what the media covered, those stories that appeared on the inside of the paper but did not vanish may be 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 would be an enormous amount of work for relatively minimal gain.

The other argument for forgoing national and international stories that fail to make Page A1 is more conceptual. We are 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 do not code every story that might appear on a website, an even more daunting task, but instead code just the top stories.)

The other challenge with newspapers that we did not face with some other media is that we include only 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 pertain 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 results in a newspaper sample of approximately 20 stories a day.

## **Online News Sites**

### **Sector Reach**

About 30 million Internet users go online for news each day. [5] About 6.8 million people read a blog each day, some of the most popular of which are news oriented. Both online news sites and blogs are becoming more important in the overall news agenda. Any sample of the modern news culture must include a representation of some of the more popular online news sources.

### **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 that aggregated news from throughout the Web.

When the Index was originally launched in 2007, the sample included five prominent Web sites that were tracked each weekday. These sites were Yahoo News, MSNBC.com, CNN.com, AOL News, and Google News. Considering the increased usage of the Internet for news as shown in recent surveys conducted by the Pew Research Center for the People & the Press, PEJ decided to expand our Internet content.

The increase in the number of sites included in the NCI took effect on January 1, 2009.

To choose the sites to be included in our expanded online sample, we referred to the lists of the top news sites based on the averages of six months of data (May-October 2008) from two rating services, Nielsen and Hitwise. (Data providing the most-visited news sites ranked by specific Web addresses were not available from Comscore at the time of our sampling.)



First, we found the top general interest news sites ranked by their average unique audience data for six months based on Nielsen NetView monthly data on a subdomain level. Second, we found the top general interest news sites ranked by their average market share for six months based on monthly rankings for top news and media websites provided by Hitwise. We then averaged the ranks of the top sites on these two lists to determine the top 12 general interest news websites. At the start of 2010, we updated our sample by averaging seven months of data (May 2009-November 2009) provided by Nielsen Media Research.

The sites included in our 2012 sample are as follows:

Yahoo News  
MSNBC.com  
CNN.com  
NYTimes.com  
Google News  
LATimes.com  
FoxNews.com  
USAToday.com  
WashingtonPost.com  
ABCNews.com  
HuffingtonPost.com  
Wall Street Journal Online

## **Units of Study**

For the online news sites, the study captures each site once a day. We rotate the time of day that we capture the websites between 9 a.m. to 10 a.m. Eastern time and 4 p.m. to 5 p.m. Eastern time. For each site capture, we code the top five stories, since those have been determined to be the most prominent at that 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 results in a sample of 30 stories a day.

## **Radio**

### **Sector Reach**

Radio is a diverse medium that reaches the majority of the people, with 94 percent of Americans 12 years and older listening to traditional radio each week. About 16% of radio listeners tune in to news, talk and information radio in an average week, which ranks it as the most popular of all measured radio formats.[6] Many more Americans get news from 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 different areas of radio news programming.

1. Public Radio: The Index includes 30 minutes of National Public Radio's (NPR) morning program, Morning Edition, each day or its broadcast of NPR's 4-6 p.m. program, All Things Considered.

NPR produces two hours of Morning Edition, and two hours of All Things Considered each day, which also include 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. In order to account for this unique relationship, in addition to rotating coding for each show, PEJ also rotates between coding the first 30 minutes of the first hour and the first 30 minutes of the second hour of the member station that we record the show from, WFYI. This gives us a closer representation of the overall content of Morning Edition and All Things Considered for a week.

### Sample NPR Coding Week

Radio Show	Monday	Tuesday	Wednesday	Thursday	Friday
Morning Edition (5-5:30 a.m.)	X				X
Morning Edition (6-6:30 a.m.)		X			
All Things Considered (4-4:30 p.m.)			X		
All Things Considered (5-5:30 p.m.)				X	

PEW RESEARCH CENTER'S PROJECT FOR EXCELLENCE IN JOURNALISM

*In this case, we would begin the next week by coding the 6-6:30a.m.broadcast of Morning Edition.*

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 conservatives, we included more conservative hosts than liberals. We code the first 30 minutes of each selected show.

The most popular conservative radio talk shows were Rush Limbaugh and Sean Hannity. We coded one of these two shows daily.

Since the politically liberal audience for talk radio is much smaller, we only coded one liberal talk show every other day. In 2012, we coded Ed Schultz. This show was the top liberal radio hosts based on national audience numbers. The Arbitron ratings, according to Talker's Magazine online, for spring 2006 are as follows:

Minimum Weekly Cume (in millions, rounded to the nearest .25, based on Spring '06 Arbitron reports) [7]

Rush Limbaugh (13.5)  
Sean Hannity (12.5)  
Michael Savage (8.25)  
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 about five minutes at the top of each hour and are national in the sense 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 code two national feeds, one from ABC radio and the other from CBS radio. Each network airs two feeds a day (9 a.m. and 5 p.m. Eastern time). We do one 9 a.m. broadcast from one network and the 5 p.m. feed from the other network every weekday.

The stations used to capture each program are selected based on the availability of a solid feed through the station's website. We have also compared their shows to that of other stations to ensure that the same edition is aired on that station as on other stations carrying the same program.

This results in the following sample:

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

Talk: The first 30 minutes of one or two talk programs each day -- one conservative (Rush Limbaugh or Sean Hannity) every day and one liberal (Ed Schultz) every other day.

Headlines: Two headline segments each day (one from ABC Radio and one from CBS Radio), about 10 minutes total.

This results in a sample of roughly one to two hours of programming a day.

## **Universe of Outlets**

Each day, then, the NCI includes about 6.5-8 hours of broadcast (television and radio), 5 newspapers (approximately 20 stories), and 6 news websites (30 stories).

### **Newspapers** (Monday through Friday and Sunday)

Code two out of these four every weekday; one out of four every Sunday

New York Times  
Los Angeles Times  
USA Today  
Wall Street Journal

Code two out of these four every weekday; one out of four every Sunday

Washington Post  
The Denver Post  
Houston Chronicle  
Orlando Sentinel

Code one out of these three every weekday; one out of three every Sunday

Traverse City Record-Eagle (MI)  
The Daily Herald (WA)  
The Eagle-Tribune

### **Websites** (Code 6 of 12 each day, Monday through Friday)

Yahoo News  
MSNBC.com  
CNN.com  
NYTimes.com  
Google News  
LATimes.com  
FoxNews.com  
USAToday.com

WashingtonPost.com  
ABCNews.com  
HuffingtonPost.com  
Wall Street Journal Online

**Network TV** (Monday through Friday)

Morning shows – Code one or two out of three every day

ABC – Good Morning America

CBS – The Early Show

NBC – Today

Evening news – Code two out of three every day

ABC – World News Tonight

CBS – CBS Evening News

NBC – NBC Nightly News

PBS – NewsHour – Code two days consecutively, then skip one day

**Cable TV** (Monday through Friday)

Daytime (2 p.m. to 2:30 p.m.) – code two out of three every day

CNN

Fox News

MSNBC

Nighttime CNN – code one or two out of three or four every day

Situation Room (5 pm)

John King, USA

Erin Burnett OutFront

Anderson Cooper 360

Nighttime Fox News – code two out of four every day

Special Report With Bret Baier

Fox Report With Shepard Smith

O'Reilly Factor

Hannity

Nighttime MSNBC – code one or two out of four every day

PoliticsNation

Hardball with Chris Matthews

The Rachel Maddow Show

The Ed Show

## **Radio** (Monday through Friday)

### News and Headlines every day

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

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

NPR Morning Edition or All Things Considered every day

### Talk Radio

Rush Limbaugh or Sean Hannity every day

Ed Schultz every other day

That brings us to 25 to 29 outlets each weekday. Sundays are accounted for with three newspapers.

## **Universe Procurement and Story Inclusion**

### **Newspapers**

For each of the eleven newspapers included in our sample, we code all articles 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 has a picture, caption or teaser to text inside the edition, we do not include that story in our sample. We code all stories that appear on the front page with a national or international focus.

Because we are looking at the coverage of national and international news, if a story is about an event that is solely local to the paper's point of origin, we exclude such a story from our sample. The only exception to this rule is when an article with a local focus is tied to a story that we have 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 war in Iraq has a local angle but is related to a national issue and is important in the context of our study.

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

Whenever possible, we have subscribed to the hard copies of the selected newspapers and have them delivered to our Washington D.C. office. This is 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 use the hard copy edition to determine the placement on the front page of the edition and to get all the text we will code. We use the LexisNexis computer database to determine the word count for each of the stories.

For all of the other papers that we are not able to get hard copies of within the same day of publication, we take advantage of Internet resources that have digital copies of the hard copy editions. Pressdisplay.com and Newsstand.com have subscription services offering same-day access digital versions of the hard copy. From these digital versions, we obtain the text of the relevant articles and also determine the word counts. To get the word counts, we copy the text of the articles (not including captions, headlines, bylines or pull-out quotes) into the Microsoft Word software program and run the "word count" function to get the final number. When necessary, we go to the paper's website in order to find the text of articles that are not available on either of the two Web services. Through examination of each individual article, we are able to determine when the text of the article on the website is the same as it would be in the hard copy of the paper.

### **Network and Cable Television**

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

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

Any story that fits the above criteria and begins within the first 30 minutes is included in the study, even if the story finishes outside of the 30 minute time period. A three-minute story that begins 28 minutes into a program would be coded in its entirety, even though the final minute ran after our 30-minute cutoff mark. The exception to this rule is when a television station is showing a speech or press conference that runs 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 is the same. PEJ subscribes to DirectTV satellite television service. We use a recording device called Snapstream to capture the shows we code. The Snapstream service digitally records each broadcast on an in-house server and then we archive the programs onto DVDs. There is redundancy in our recording method so that each show is also recorded on TiVo recording boxes that are directly linked to DirectTV. We do this in order to avoid problems in our capture that might result from technical error.

Occasionally outlets deviate from the regularly scheduled news programs. When a show is pre-empted for a special *live* event, such as a presidential campaign debate or the State of the Union address, we do not include that period as part of our sample.

## Radio

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

For each of the radio shows selected, we have found national feeds of the show that are available on the Web. As with television, we have two computers capturing each show so as to avoid errors if one feed is 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 archive those programs onto DVDs.

## Online

For each of the websites we are including in our sample, we capture and code the top five stories that appear on the site at the time of capture. Our capture times rotate on a regular basis. They occur either between 9 a.m. and 10 a.m. Eastern time or between 4 p.m. and 5 p.m. Eastern time each weekday. The captures occur 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 appear in our browsers at the time of the capture. We rely on people rather than a software package to capture sites because some software packages have proved invasive to websites.

With the current rotation of websites along with the rotation of the times of day that we capture the sites, we wanted to make sure that we did not always capture the same sites at the same time (CNN.com always at 9 a.m., for example). We also wanted to ensure that for the websites where we coded another outlet from the same news organization, such as USA Today's newspaper and the usatoday.com website, we did not code both of those outlets on the same days. In order to avoid these two concerns, we created a method of rotation in which the capturing times for the website rotate every two days.

This means that the pattern the capture times follow is 9 a.m., 9 a.m., 4 p.m., 4 p.m., 9 a.m., etc.

Here is an example of how the online rotation works:

	<b>Week 1</b>					<b>Week 2</b>					<b>Week 3</b>					<b>Week 4</b>				
	4 pm	9 am	9	4	4	9	9	4	4	9	9	4	4	9	9	4	4	9	9	4
CNN.com		x		x		x		x		x		x		x		x		x		x
Yahoo News		x		x		x		x		x		x		x		x		x		x
MSNBC.com		x		x		x		x		x		x		x		x		x		x
WSJ.com		x		x		x		x		x		x		x		x		x		x
FoxNews.com		x		x		x		x		x		x		x		x		x		x
USAToday.com		x		x		x		x		x		x		x		x		x		x
NYTimes.com	x		x		x		x		x		x		x		x		x		x	



LA Times.com	x		x		x		x		x		x		x		x		x		x	
WashingtonPost.com	x		x		x		x		x		x		x		x		x		x	
ABCNews.com	x		x		x		x		x		x		x		x		x		x	
Huffington Post.com	x		x		x		x		x		x		x		x		x		x	
Google News	x		x		x		x		x		x		x		x		x		x	

Beginning on May 30, 2011, we changed the rotation so that CNN.com and MSNBC.com are not captured on the same days.

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

Because each website is formatted differently, we came up with a standard set of rules to determine which stories are 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. First, we ignore all advertisements and extra features on the sites that are not reported news stories. We are interested only in the main channels of the websites where the lead stories of the day are displayed. Second, we determine the top “lead” story. That is the story with the largest font size for its headline on the home page. The second-most prominent story is the story that has a picture associated with it, if that story is different than the story with the next largest headline. By considering many sites, we realized that a number of sites associate pictures with stories that they find particularly interesting but are clearly not intended to be the most important story of the day. We do want those stories to be in our sample, however, because the reader’s eye will be drawn to them.

Having figured out the most and second-most prominent stories, we then rely on two factors to determine the next three most prominent stories. We first consider the size of the headline text and then the height on the home page. Therefore, for determining the third most prominent story, we look for the story with the largest headline font after the top two most prominent stories. If there are several stories with identical font sizes in headlines, we determine that the story that is higher up on the page is more prominent. In cases where two articles have the same font size and the same height on the screen, we choose the article to the left to be the more prominent one.

For the first two years of the NCI, we did not include online news stories that were audio or video features. Starting in 2009, PEJ changed its method of measuring online stories to allow for the inclusion of audio and video stories. See the section below entitled “Inclusion of Online Audio and Video in Index Calculations” for details on how the changes to the Index statistics have been incorporated.

### **Coding Procedures and Intercoder Reliability**

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

The source variable takes in all the media outlets we code. Variable broadcast start time applies to radio and TV broadcast news and gives the starting time of the program in which the story appears. Story word count designates the word count of each individual print/online news story. Duration in seconds is the amount of time for broadcast only, calculated by subtracting story start time from story end time. The placement/prominence variable designates where stories are placed within a publication, on a website 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 time period under study. Sub-storyline applies to stories that fit into some of the long-running big stories, reflecting specific aspects, features or narrower elements of some big stories. The geographic focus variable refers to the geographic area to which the topic is relevant in relation to the location of the news source. The broad story topic variable identifies which of the broad topic categories is addressed by a story. Variable lead newsmaker determines the person whose actions or statements constitute the main subject matter of the story.

The coding team responsible for performing the content analysis in 2011 was made up of sixteen individuals. The daily operation is directed by a coding manager, a training coordinator, a methodologist, and a senior researcher. Several of the coders have been trained extensively since the summer of 2006, and most of the coders have more than a year's worth of coding experience.

Numerous tests of intercoder reliability have been conducted since the inception of the NCI in order to ensure accuracy among all coders.

### **2011 Intercoder Tests**

In Spring 2011, PEJ conducted one major intercoder test. This test focused on complex variables, including broad story topic, big story, substory, geographic focus, format, and lead newsmakers.

97 stories (representing roughly 10% of a week's sample) were randomly selected from across all 5 media sectors. Coders were asked to recode 10 stories from newspaper sector, 15 from online, 24 from network, 33 from cable and 15 from radio sector.

A total of 16 coders participated in this round of testing. The number of stories recoded by a coder depended on the amount of coding executed by the coders in the previous six weeks.

The percent of agreement was as follows:

Big story: 98%

Substory: 82%

Geographic focus: 86%

Format: 95%  
Lead newsmaker: 89%  
Broad story topic: 78%

### **2010 Intercoder Tests**

PEJ conducted one major intercoder test over the course of 2010. This test focused on the complex variables that require extensive training, expertise and effective communication among the coders. These are the Main Variables.

#### **Main Variables**

Over the course of Spring and Summer of 2010, we tested intercoder agreement for Main Variables. One hundred and five stories (representing 10% of a week's sample) were randomly selected from across all 5 media sectors. Coders were asked to recode 12 stories from the newspaper sector; 15 from online; 20 from network; 34 from cable and 24 from the radio sector.

In this round of testing multiple coders were asked to recode the same stories. A total of 17 coders participated in this test. The number of stories recoded by a coder depended on the amount of coding executed by the coder in the previous six weeks.

The percent of agreement was as follows:

Format: 83%  
Big Story: 85%  
Substory: 82%  
Geographic Focus: 91%  
Broad Topic: 78%  
Lead Newsmaker: 84%  
Lead Newsmaker 2: 95%

#### **Select Housekeeping Variables**

In addition to these Main Variables, we also tested certain Housekeeping variables in the same round of testing. Housekeeping variables are those that are necessary for each story, but require little or no inference from each other.

Placement: 97%  
**Broadcast Only Variable**  
Broadcast Start Time: 97%

### **2009 - Early 2010 Intercoder Tests**

In 2009, PEJ conducted two phases of major intercoder testing to ensure continuing accuracy among all coders.

The first phase tested for variables that require little to no subjectivity from the coder. We refer to these codes as Housekeeping Variables. The second phase of testing was conducted in the fall of 2009. In this phase we tested for variables that are more complex and require more training and expertise. We call these the Main Variables.

### **Housekeeping Variables**

In summer 2009, we tested intercoder agreement for Housekeeping variables. These are variables that are necessary for each story but involve little inference from each coder.

We used a random sample of 131 stories, representing all five media sectors that we code. This sample represented 10% of the number of the stories we code in an average week.

A total of 15 coders participated in the study. Each coder was asked to recode each of the 131 stories.

A total of 27 print (12 newspaper, 15 online) and 104 broadcast (44 network, 36 cable and 24 radio) stories were sampled.

The percent of agreement was as follows:

Story Date: 99%

Source: 97%

Placement: 94%

#### **Print Only Variable:**

Story Word Count (+/- 20 words): 84%

#### **Broadcast Only Variables:**

Broadcast Start Time: 98%

Story Start Time (+/- 6 seconds): 97%

Story End Time (+/- 6 seconds): 91%

### **Main Variables**

The second group of variables we tested was referred to as the main variables, and they involve more training and interpretation. Having already demonstrated that we had a high level of agreement for all of our housekeeping variables, we then had the coders participate in separate tests for these main variables.

In the fall of 2009, we conducted intercoder testing for main variables. One hundred and thirty stories coded were randomly selected from all five media sectors (20 newspaper articles, 10 online stories, 36 network stories, 41 cable stories and 23 radio stories). These stories were coded over the course of 10 weeks.

A total of 16 coders participated in this test.

For main variables, we achieved the following level of agreement:

Format: 86%

Big Story: 85%

Sub-storyline: 83%

Geographic Focus: 89%

Lead Newsmaker: 86%

Lead Newsmaker 2: 90%

For our most complicated variable, Broad Story Topic, we conducted multiple tests in mid to late 2009 and early 2010. The average agreement for Broad Story Topic was 81%.

### **Testing Details**

All the percentages of agreement for the above variables were calculated using a software program available online called PRAM.[8]

Since the inception of the News Coverage Index, as new coders were hired and included in the coding team, they were given extensive training by the training coordinator, the content supervisor and other experienced coders. New coders were not allowed to participate in the weekly coding for the project 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 D.C. office and was trained to work on all the print and broadcast media included in the sample. The schedule for each coder varies, 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 ensure statistical reliability, generally no one coder codes more than 50% of a particular media sector within one week. And each coder codes at least three media sectors each week. In the case of difficult coding decisions about a particular story, the final decision is made by either the coding administrator or a senior member of the PEJ staff.

The coding data are entered into a proprietary software program that has been written for this project by Phase II Technology. The software allows coders to enter the data 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 used to determine top stories in broadcast and cable is time, and in text-based media, it is words. Thus for cable news, for example, we refer to the

percentage of total seconds that a certain story received. In other words, of all the seconds analyzed in cable news in a 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. Should we run a simple average of all five? Should we average 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, circulation in newspapers, unique visitors in online), any system based on audience figures raises serious issues of discontinuity. 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 audience 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 for the launch of the Index in January 2007, we tested multiple models:

Model 1: Compile percentages for big stories for each of the five media sectors (newspapers, online sites, network TV, cable TV 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 TV, cable TV 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 the 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 (June 2005, November 2005 and August 2006). First, we took the average response for each media sector across the three time periods. Next, we rebalanced the average percentages to match the five media sectors in the Index — newspapers, online, network TV, cable TV and radio — to equal 100%. Thus, the weight

for newspapers was 0.28, online was 0.16, network TV was 0.18, cable TV was 0.26, and radio was 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 regular media usage survey data, collected by the Pew Research Center for the People & the Press in its Biennial Media Consumption Survey 2006. Thus, the weight for newspapers was 0.307, online was 0.218, network TV was 0.165, cable TV was 0.201 and radio was 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 had the virtue of tracking the media use for national and international news, which is what the Index studies. Also, the Pew Research Center for the People & the Press asks this question about once every six months so we can reflect changes in media use. We adopted this model and update the weights when appropriate.

The weights used for data in Model 3 have been updated since the inception of the News Coverage Index.

On January 1, 2012, we updated our weights for the year. The 2012 weights were calculated by averaging survey data from July and December 2011. These data were collected by the Pew Research Center for the People & the Press. The weights used for the Index in 2012 were:

### **2012 Weights**

Newspapers: 0.19  
Online: 0.30  
Network TV: 0.15  
Cable TV: 0.23  
Radio: 0.12

### **Inclusion of Online Audio and Video in Index Calculations**

The decision to include audio and video stories for online beginning in 2009 meant that PEJ needed to create a method to incorporate different ways of measuring length (time in seconds versus amount of words) within the same media sector.

Prior to this change, the calculations for the percentage of the newshole for the online sector had been the percentage of words (in text). However, including multimedia elements in our Web sample created a challenge for coming up with a percentage of newshole calculation that can incorporate both text and the length of multimedia stories in seconds. PEJ undertook several tests to come up with a simple, yet accurate, method of creating an equivalent measure.

The process PEJ uses for valuing multimedia stories is to take the length of the multimedia story (in seconds) and multiply by 4 to get an approximate equivalent value to a text story of that number of words. For example, an online video that is 30 seconds in length would be given an equivalent value of 120 in words (30 x 4). An online video that is 60 seconds in length would be given an equivalent value of 240 words (60 x 4).

PEJ arrived at this method by first timing how long it takes for people to read news stories out loud. After timing five people reading different types of news stories, we discovered that people read approximately 3 words per second. However, simply multiplying the length of a story by 3 would not accurately reflect the value of a multimedia story.

We then compared the distribution of the length of online text stories to the distribution of the length of online multimedia stories. To make this comparison, we took the distribution of 3,500 online text stories over the last six months in the NCI and compared that to the distribution of length for 280 video stories compiled from seven separate Web news sites. The median length of a text story was approximately 600 words, while the median length of a video story was approximately 150 seconds.

Drawing from these comparisons, we determined that multiplying the length by 4 gave a reasonable approximate value to use in comparison to the length of a text story in words. No single multiplier would match exactly since the distribution of the length of Web videos is not linear and because there is no simple way to quantify the value of visuals within multimedia stories along with the text. However, this simple method of multiplication gives us a straightforward way to make approximate equivalents between two measures (seconds and words) that are not otherwise easily compared.

---

#### **ENDNOTES:**

[1] Data from 2012 indicate that the three evening networks reach about 22.1 million viewers and the three morning newscasts average about 12.6 million people daily. Nielsen Media Research, used under license.

[2] Data from 2012 indicate that 3.3 million viewers watch cable news during prime time hours and 1.9 million watch during daytime hours. Pew Research Analysis of Nielsen Media Research, used under license.

[3] Larry King Live was no longer on the air in 2012, and was replaced by Piers Morgan Live.

[4] For 2012, circulation numbers indicate that roughly 44 million people buy a newspaper each weekday. Pew Research Center analysis of Newspaper Association of America data.

[5] According to the August 2011 survey, 28% of adults go online for news each day. Pew Internet, August 2011 survey.

[6] In 2012, 33% of adults said they listened to “news radio yesterday.” Pew Research Center media consumption survey.



[7] Current ratings data available at Talkers Magazine [online](#).

[8] Kimberly A. Neuendorf, “The Content Analysis Guidebook,” Sage Publications, 2002.