Matthew Hindman
Assistant Professor
School of Media & Public Affairs
The George Washington University

June 10, 2011

Response to Peer Review

I was pleased to receive Prof. Chyi's peer review, and to read her generally positive comments regarding the overall study. Given the substantial work that the report required, it is gratifying to read her assessment. In response to her suggestions, I have made several clarifications and additions that I believe improve the final report.

I will reply to Prof. Chyi's specific points in detail below, in the order in which they appear in her written review.

Prof. Chyi argues that some aspects of the sample construction might underestimate the online news population, particularly the relatively small sample of workplace users. This is likely correct (as I noted on page 10). The quote she highlights was intended, in context, to apply to just one aspect of the comScore's sampling procedure: its likely overrepresentation of heavy Internet users. I stand by this assessment regarding the likely over-prevalence of heavy users, and have revised the text in the interest of clarity. But Prof. Chyi is right that comScore's much smaller work sample raises important unanswered questions about its ability to capture work usage, and I have added slightly to this discussion (including a reference to Pablo Boczkoski's recent book on this issue). My stance is somewhat more skeptical than hers appears to be. Without greater disclosure from comScore, it is difficult to assess whether the construction of the comScore work sample is likely to undercount local news usage overall—but the possibility certainly exists. Unfortunately, similar biases are possible in other Web measurement firms' data as well.

My biggest divergence from Prof. Chyi's analysis concerns her discussion of the audience reach threshold that sites need to achieve in order to be included in the overall analysis. The report looks just at sites that reach one percent of comScore's panelists in at least one of the three months. Prof. Chyi argues for using a lower threshold. In fact, *it is simply not possible to use a lower threshold in the full 100-market study*, at least using the comScore data provided.

Why not 0.5% or 0.3%, as Prof. Chyi suggests? The answer, as the report noted, is that the smallest markets in the sample (such as Madison WI or Burlington VT) average just 600-some panelists across the three months. 0.5% reach for 600 panelists is 3 unique

visitors. But as the report also stated, for ALL local markets, the comScore data do not include any sites that receive fewer than 6 visitors. The threshold cannot be set below 1% unless we are willing to have a different audience share threshold in different-sized markets. The worry is that, as I wrote, "A site that got five panelist visits in Burlington would be omitted from the analysis, while a site that got eight visits in New York would be included—even though the market reach of the Burlington site is 18 times higher."

Prof. Chyi suggests that using 1% as a threshold "may not be the best decision." But it is incorrect to characterize this as a "decision" at all. I did not "decide" that 6 is 1% of 600, of course—and if we do not have data on sites that get fewer than 6 visitors, we cannot go lower than 1% and have consistent standards across all markets, as the study's protocol calls for.

Even lowering the threshold slightly would raise the potential for censoring in many or most of the smallest markets in the study. Using a 0.5% threshold would require a panel size of at least 1200 users (i.e. 6 / 1200) to avoid censoring issues, larger than the panel size of 33 of our 100 markets. A 0.3% threshold requires 1800 panelists, which would raise censoring issues in 54 of the 100 markets. Additions to the text highlight these limitations and further explain the issue.

Prof. Chyi suggests as well that is "room for speculation" that a large number of hyperlocal sites, too small to be included, might potentially add up to substantial audience share. The data make such a claim extremely implausible. In an added paragraph, I examine local news sites with between 1 and 1.2 percent audience reach—just above the threshold for inclusion. These sites average just 0.008 percent of monthly page views each in the local market; by contrast, the top site in the market averages .22 percent of local page views. Omitted sites must necessarily have less traffic on average than sites that *were* included. In the added paragraph, I discuss the case of the median market, where 9 online news outlets collectively add up to .43 percent of local page views. Even if there were a dozen omitted outlets with an average of .006 percent of local page views each—an implausible assumption—the total local news market would still only account for 0.5 percent of local page views.

I suspect that Prof. Chyi would not have objected if, instead of looking at audience reach, I had included only sites that account for at least one-hundredth of one percent of local page views. In practice, however, this would be an *easier* standard than one used in the report.

Claims about the "long tail" distribution of online media are widespread, particularly the suggestion (as here) that numerous small media outlets might together rival the largest outlets in audience. I critique these claims at length in my recent book, *The Myth of Digital Democracy* (Princeton University Press, 2009). The notion of a "long tail" necessarily means that the number and size of small sites are distributed in a predictable fashion. Simply put, the math does not work: over the entire Web, or within

categories of news or political Websites, the observed power law (or extreme lognormal distribution) is too steep for small sites in practice to add up to substantial collective market share. That lesson certainly seems to hold here too, though these markets have too few data points to calculate anything like a power law (or lognormal) curve.

Moreover, the report does look, in great detail, for dozens of specific hyperlocal news outlets that might have fallen below the one-percent threshold, drawing on a number of diverse sources. It finds very few in the comScore data, presumably because they fail to receive the requisite six visitors among the comScore panelists. The closer examination of five markets also addresses these potential concerns; it similarly finds few local news outlets left out.

In short, we can be confident that any local news sites omitted because they failed to reach a larger portion of the public do not, in fact, add up to more than a fraction of online local news consumption.

I also include two additional figures, which graph the distribution of audience reach and monthly minutes per user for the top-ranked sites in our data. While not directly responsive to the peer review, the added figures help clarify the scale of the disparity between the largest and smallest local news outlets.

The discussion of the Pew report in the initial paragraph has been clarified. Pew has asked multiple questions on this topic for decades, but the main survey question which this claim references asks respondents to identify up to two primary news sources.

I have made a few other minor changes, such as fixing the identified typos, and clarifying the significance levels in the regression analysis. "Per person" metrics are (with noted exceptions) per Internet user, using comScore's estimates of the total online population in each broadcast market.