Local News Online: Aggregators, Geo-Targeting and the Market for Local News*

Lisa M. George†
Hunter College and the Graduate Center, CUNY
Christiaan Hogendorn‡
Wesleyan University

November 1, 2013

Abstract

We examine the effect of aggregating local news links on visits to news outlets. Using a sample of news visits by 43,087 US households before and after a *Google News* design change, we find that adding geo-targeted links increases both the level and share of local news consumed online. The magnitude is small: local news visits increase by less than 1% and the likelihood of a local news visit increases by 4-6% from a low baseline. The redesign also increases the local share of visits and referrals to outlets, indicating that aggregation can alter both the size and composition of the local media audience.

Keywords: Communications, Media, Internet, News, Advertising
JEL Classification.: D22, L13, L82

*This paper was written with financial support from the Time Warner Cable Research Program on Digital Communications. The views expressed are those of the authors and not necessarily those of Time Warner Cable or the Time Warner Cable Research Program on Digital Communications. We are grateful to helpful discussions by Kate Fang, Lapo Filistrucchi and Jeff Prince.

†Lisa M. George, Department of Economics, Hunter College, 695 Park Ave., New York, NY 10065. Email: `lisa.george@hunter.cuny.edu`.

‡Christiaan Hogendorn, Department of Economics, Wesleyan University, 238 Church Street, Middletown, CT 06459. Email: `chogendorn@wesleyan.edu`.

1 Introduction

The role of aggregators in the market for news has been fiercely debated for over a decade. Vocal newspaper moguls challenge the right of outsiders to collect, link and repost news in ways that divert readers from original sources, while digital media giants and bloggers alike defend intermediaries that reduce consumer search costs, improve matches between readers and content, and raise consumption.

Economic research has made limited progress in resolving this question. From the outlet perspective, Chiou and Tucker (2011) show that a contract dispute between the Associated Press and Google which removed Associated Press content from *Google News* for a short period in early 2010 reduced demand. The result suggests complementarity between aggregated links and original content, yet because the Associated Press had little standing as a news outlet on its own, the result does not readily generalize to the larger news market. On the consumer side, Athey and Mobius (2012) show that users who adopt a localization feature of *Google News* in France increase visits to local news outlets in the short run, yet over time most additional local news consumption derives from increased use of *Google News*. The question remains unresolved.

A major redesign of *Google News* on June 30, 2010 offers a rare opportunity to study the effect of aggregation on the demand for news. The redesign placed a permanent strip of geo-targeted local news headlines and links onto the *Google News* front page. Using a sample of news visits by 43,087 US households before and after the design change, we find that adding geo-targeted links increases both the level and share of local news consumed online. The magnitude of the effect is small: local news visits increase by less than 1% and the likelihood of a local news visit increases by 4-6% from a low baseline for heavy *Google News* users. But the results show no evidence of substitution away from direct outlet visits.

The redesign allows us to examine changes to the market for online news beyond straightforward complementarity. Adding geo-targeted links increases the number of different local outlets visited per day, but not the number of unique sites visited per month, suggesting that increases in local news consumption arise from more frequent visits to familiar news outlets rather than visits to additional news providers. The result implies that aggregators play less of a role in product discovery than is sometimes argued, at least in the narrow context of local news. Our household panel also allows some inference regarding the effect of aggregation on news outlets. We find that the

redesign raises the number of Google referrals from within the home market and reduces referrals from elsewhere, indicating that aggregation can alter not only the size but the composition of the audience for local media.

Our results highlight the importance of studying how intermediaries systematically *shift* consumption across media outlets rather than directly substitute for or complement original news. If aggregators reduce search costs uniformly, allowing consumers ready access to content previously too difficult to find, then observed shifts in readership are likely efficient. However if intermediaries reduce the costs of consuming some types of news relative to others, readers may switch to less-preferred but more readily available material rather than search. While this outcome might be cost-minimizing overall, the re-allocation of attention changes relative demand for different media types and can have important effects on competition.

The substantial effect of technology on relative demand has been documented in traditional media markets. George (2008) showed how the spread of the internet altered the composition of the audience for traditional local newspapers, pulling younger and more educated readers out of local newspaper markets. George and Waldfogel (2005) showed how national expansion of the *New York Times* made possible by satellite printing attracted highly educated readers away from local newspapers, with consequences for local media markets and also for local voting (George and Waldfogel 2008).

This work contributes to a small theoretical literature that is building a more nuanced picture of how aggregators shape the market for news. George and Hogendorn (2012) highlight the importance of transaction costs in consumption variety. Jeon and Esfahani (2012) show how consumer preferences for quality can lead aggregators to increase or decrease demand for original content. Alaoui and Germano (2013) study how time constraints impact news consumption in digital markets. Rutt (2012) studies competition in the presence of news aggregators.

Our results are relevant to policy at several levels. Strong provisions have long been in place in the US to promote local media, and localism is one of the three principles (along with diversity and competition) guiding Federal Communication Commission policy. Technologies that reduces barriers to local news consumption, even with small effects, are of great interest in reforming policy to reflect modern markets. In the context of digital news, a particular concern in the US has been that intermediaries have reduced the costs of locating and consuming national information relative to local content, facilitating readership shifts to national media that harm local outlets.

Evidence from aggregate data shows online news markets to be highly concentrated, with a handful of national news sites such as *CNN*, *Fox News*, and the *New York Times* attracting the vast majority of news visits. Until very recently, few aggregators offered tools to systematically and accurately identify local content. Geo-targeting is one of the few options available to offset this trend.

This study is also related to debate on privacy standards on the internet. While older personalization features of *Google News* and other aggregators allowed consumers with an interest in local media to "opt in" by providing geographic information, the redesigned *Google News* site automatically identifies local content based on IP addresses and other geo-targeting technologies. The low baseline of local news consumption among *Google News* users during the "opt in" period and the measurable increase with automatic targeting suggests that these technologies can have modest positive social effects that can offset privacy concerns that much of the debate.

The paper proceeds as follows. Section 2 describes the research design. Section 3 outlines construction of the data. Section 4 describe the empirical strategy. Sections 5 and 6 present results. Section **??** concludes.

2 Research Design

On June 30, 2010, Google introduced the first and only comprehensive redesign of the *Google News* page since the start of the service. The redesign altered the presentation of content in several ways and allowed users to prioritize subject areas. Most important for this study, the redesign added a strip of content on the right side of the page or "sidebar" with a set of local headlines and local news links, with the location of the user identified automatically through the IP address. While some customization of local content had been possible on the site since early 2008, the earlier features were available on an opt-in basis, requiring registration, log-in and user input. After June 30, 2010, local content was reported automatically through geo-targeting technology and could not be removed through customization.

Figure 1 shows a screen shot of a *Google News* page on June 28, 2010 before the design change and Figure 2 shows a screen shot on July 2, 2010 just after the change. The June 28 page reveals an area where viewers can enter preferences for topics or local content. This screenshot, scraped from the Internet Archives "Wayback Machine" shows a non-customized page. The screenshot in Figure 2, taken four days later, reflects

the new format. The page still includes personalization options for users who log into the site, but local headlines and links are now fixed on the right side of the page, shown with a large arrow (added for clarity). The location of the scraping server is identified as San Francisco, and the local content includes two local news headlines from the *San Francisco Chronicle* and one from the *San Jose Mercury News.*

Our basic empirical strategy is to measure the effect of adding local news links to *Google News* on news visits to outlets that are local to a sample of internet users. In one set of tests, we rely on a constructed treatment and control group that compares behavior of heavy *Google News* users to active Yahoo users before and after the redesign. In a second set of tests, local news consumption patterns before and after the redesign are compared based on intensity of *Google News* use prior to the change. A third set of tests examines the effect of the *Google News* redesign on the composition of the audience at local news outlets.

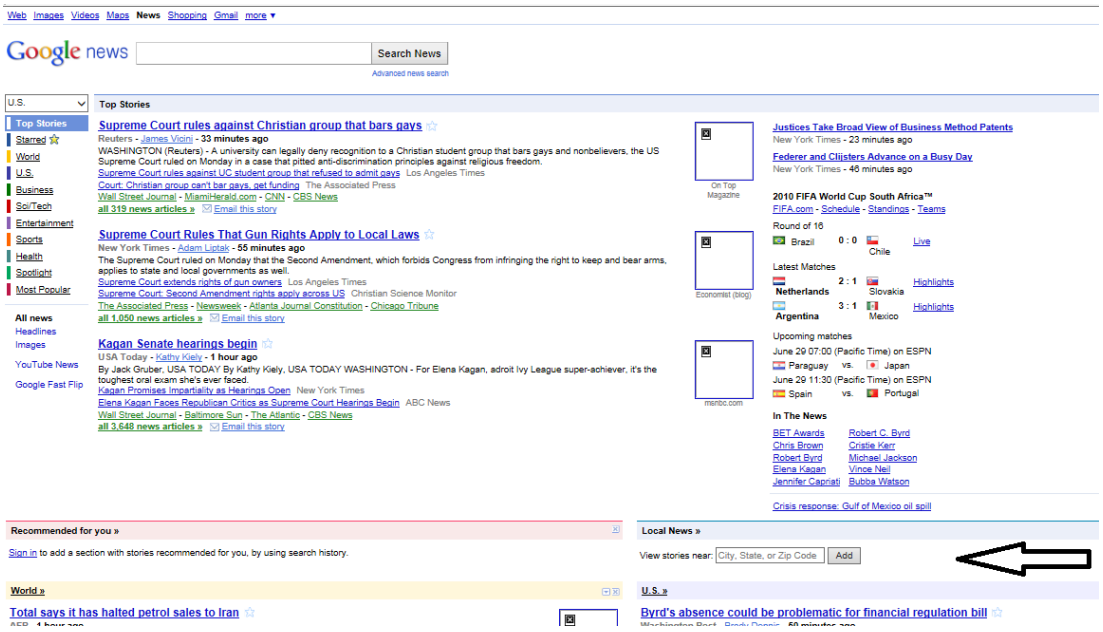
From a theoretical perspective, lower transaction costs of accessing local news should increase local news consumption relative to non-local news. But if time savings are substantial, users might also increase non-local news consumption through a time-based "income effect" leading to an ambiguous prediction regarding the local consumption share.[1]

More practically, one limitation of the experimental design is that the July 2010 changes to the *Google News* site expanded the customization features available to users signed into Google at the same time geo-targeted local news was added to the site. These "News for You" features allowed users to more easily customize and prioritize content categories on the main page. The dual change raises the possibility that Google referrals to non-local as well as local news outlets might increase after the re-design, again with ambiguous effects on the local news share. We return to this question in the empirical analyses.

The empirical strategy has features in common with Athey and Mobius (2012), who study the effect of a *Google News* design change on local news consumption in France. Other than application in the larger and more complex US news market, a key difference is that the US redesign in July 2010 placed geo-targeted news links in front of all users simultaneously, rather than as an "opt-in" feature. Studying this comprehensive change allows for a simpler estimation strategy than with the introduction of opt-in local content. In addition, results from the comprehensive redesign speak more

[1] See George and Hogendorn (2012) for a theoretical treatment of this topic.

Figure 1: *Google News* Front Page Before Redesign, June 28, 2010

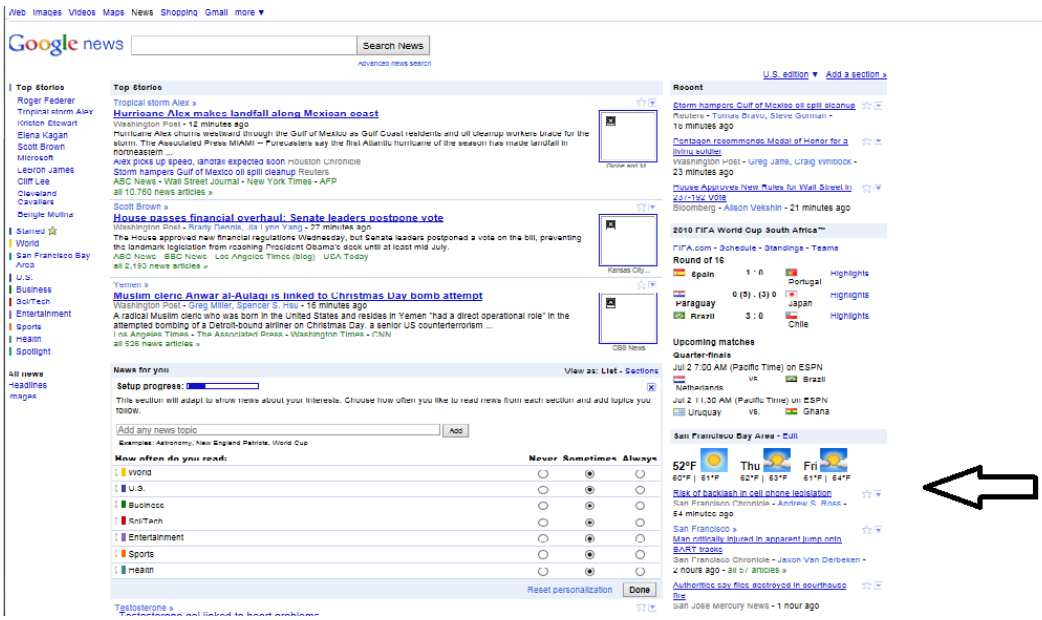


clearly to the overall effect of lower transaction costs on local news consumption and a closer link to privacy policies associated with geo-targeting. Perhaps more important, examining the comprehensive redesign also allows study of the effect of aggregation on local news outlets. Experimental conditions for testing whether aggregated links substitute for direct visits are uncommon, and this broad-based change to *Google News* allows estimates of how aggregation affects the composition of the audience for local news outlets.

3 Data

The basic working data is a panel of 43,087 household news visits each day fourteen weeks before and after the June 30, 2010 redesign of the *Google News* site. For each household each day, the data record the total number of news visits and the number of news visits to outlets local to the household. The data also record the number of news visits and local news visits referred by Google and Yahoo (news and search). For each household, the data include the Metropolitan Statistical Areas (MSA) of residence and basic demographic indicators for race, Hispanic origin, and income.

Figure 2: *Google News* Front Page After Redesign, July 2, 2010



The working data set is constructed from several underlying sources, discussed in turn.

3.1 Site Visit Data

Site visits logs come from the ComScore Web Behavior Database. The data record the complete browsing history for a nationally-representative sample of approximately 50,000 households that have opted to allow tracking of internet use. Data are collected at the machine level, so may reflect behavior of more than one user in each household and do not capture computer use on mobile devices or computers accessed outside the home. The opt-in nature of the collection program means that the sample may not be fully representative of the population in terms of computer use. These limitations are common to most sources of computer use microdata and largely affect the generalization of results rather than estimation.

The raw data include basic user demographics and household zip code. Zipcodes are merged with US Census geography to identify the Metropolitan Statistical Area for each household. Households living outside of MSA's are not included in the study.

3.2 News Outlets

News sites are identified from several proprietary databases and public sources. Newspaper, radio, magazine and television outlets are identified from *Burrelle's Media Directory* (2000 & 2005 edition), Bulldog Reporter's *MediaPro Directory* (2008) and the Newspaper Association of America web site (2010). In addition, all sources that appear on *Google News* in 2009 and 2010 are classified as news outlets, with the exception that we exclude the upper tail of high traffic sites for which the vast majority of visits are not news visits (about.com, aol.com, apple.com, blogspot.com, comcast.net, microsoft.com, msn.com, youtube.com). We also do not study visits to Associated Press content hosted by Google because we cannot fully identify these visits in the data for reasons outlined below. The final working data include visits to 6,407 domains.

The raw site visit data available for this study include only top-level domains. This limitation means that media outlets co-hosted on the same domain cannot be distinguished. (For example, local television and radio stations sometimes share a web site, as do some broadsheet and tabloid daily newspapers.) Since co-hosted sites are local to the same market and household visits are aggregated over domains each day, this limitation has minimal impact on the analyses. A more difficult problem is distinguishing *Google News* referrals from Google search referrals. The procedure for indirectly measuring *Google News* use is described in the subsection on intermediation, below.

3.3 Local Visits

In US markets, there is no consensus measure of what constitutes "local" media, especially on the internet. For this research, local news visits are identified based on demand. For each domain, the number of visits from each MSA over the entire year are counted. The MSA with the highest number of visits is defined as the home MSA for each media outlet. The share of visits to each domain from the home MSA is recorded as the home share for each domain.[2]

With this revealed preference designation of the home market and home market share for each outlet, we adopt two approaches for measuring local news consumption.

[2]To reduce measurement error, outlets with fewer than 10 visits over the year are excluded from the sample.

Our primary approach is to record a household news visit as local if the household MSA matches the home MSA for the domain and the local visit share for the domain exceeds 15%, which is the 25^{th} percentile in the local share. Foreign news outlets defined based on URL extensions are also classified as non-local sites. This categorization excludes from the local definition most national news sites, topical news sources, and international media. The *New York Times* (11% New York share) and the *Wall Street Journal* (7.1% New York) are not considered local to New York by this definition.

A second approach avoids the need for an arbitrary cut-off for national outlets by using a weighted continuous measure of local news consumption. Each household visit to a news site in its local MSA is characterized by the average home share of that outlet. So a household in New York that visits the *New York Times* (11% New York), the *New York Daily News* (25% New York) and *CNN* (4% Atlanta) would have a weighted local news consumption of $0.11 + 0.25 + 0 = 0.36$. An Atlanta resident with the same visit profile would have local news consumption for the day of .04.

For purposes of this study, the revealed preference measure of local interest has the feature of capturing implicit localism in different types of outlets which would not be identified with a media list based on the place of publication. For future research, the measure offers a useful way of characterizing competition between broadcast and print media, and also for understanding the extent to which demand for different types of information by different types of users is satisfied locally. For example, minority-targeted news outlets vary considerably in the share of viewing each receives in different markets. Relating demand for non-local media to individual and population demographics offers a systematic way of studying how groups with distinct tastes satisfy demand for information. For this study the relevant geography is set at the MSA level, but the data allow for finer definitions of local media for future research on community information needs.[3]

The share of visits from households in the home MSA provides a measure of localism for each news outlet. The 25 domains with the largest number of local visits in the sample are shown in the top portion of table 1 along with visit counts and the local share. The bottom portion of the table shows visit counts for the ten largest sites in

[3]There is considerable policy interest in this topic. For example, in 2012 the FCC commissioned a literature review examining how communities meet "critical information needs." Though the authors concluded that digital media were not likely to satisfy those needs, they highlighted the subjective nature of current research and need for generalizable measures of local supply and demand.

the sample, where the local share is again the share of all visits from the MSA with the largest number of visits. In most markets, the media outlet with the most local visits is the major urban daily newspaper, but in some markets the top outlet is a local radio or television station website. For example, *WPXI* News in Pittsburgh receives more visits from inside the MSA than the *Pittsburgh Post-Gazette*, although the newspaper site receives slightly more visits overall (4,881).[4] In general, the local visit share for radio and television stations is considerably higher for broadcast than for print media, with many broadcast sites reaching shares of over 90%. In table 1, the only sites with a local share exceeding 90% are broadcast sites.

3.4 Intermediation

News visits referred by intermediaries are identified from a field in the raw session data that lists the referring domain. Most relevant for this study are referrals from Google and Yahoo. The raw session data identify only top-level domains for referrals as well as visits. Because of this, referrals by *Google News* cannot be directly distinguished in the raw data from Google search referrals. The basic identification strategy, which relies on changes to the *Google News* page that do not affect search, does not require distinguishing referrals in the data. However, a measure of *Google News* use intensity is needed to construct a treatment and control group and also to identify households most affected by the redesign.

We develop an indirect measure of *Google News* usage by linking referrals in the visit data to outlets appearing on *Google News*. Specifically, we scrape *Google News* headlines from the archival site *Archive.org* that operates a well-known program called the "Wayback Machine." The scraped data identify for each domain-day whether or not an outlet appeared on the front page of *Google News*. In the working data, news visits referred by Google on days the domain was listed on the *Google News* page are classified as *Google News* referrals. For example, a visit to the *Atlanta Journal Constitution* referred by Google on a day the newspaper appeared on *Google News* would be coded as a *Google News* referral. A visit to the *Atlanta Journal Constitution* referred by Google on a day the newspaper did not appear on *Google News* would not be recorded as a *Google News* referral. Also, a visit to the *Atlanta Journal Constitution* on a day it appeared on *Google News* but was not referred by Google would not be

[4]The local visit totals and share suggest interesting implications for outlet competition for advertisers.

Table 1: Local Visit Share for Top Local News Sites

Outlet	Market	Total Visits	Local Visits	Local Share
Top 25 Outlets by Local Visits				
NY Daily News	New York, NY	23,474	5,315	23%
LA Times	Los Angeles, CA	23,471	5,179	22%
Washington Post	Washington, DC	21,963	5,549	25%
NJ.com	Newark, NJ	15,510	3,408	22%
Boston Globe	Boston, MA	14,448	5,131	36%
Atlanta Journal Constitution	Atlanta, GA	12,088	8,228	68%
KSL	Salt Lake City, UT	10,845	7,739	71%
Arizona Central	Phoenix, AZ	9,052	5,577	62%
Chicago Tribune	Chicago, IL	8,680	3,895	45%
Houston Chronicle	Houston, TX	8,247	5,041	61%
Cleveland Plain Dealer	Cleveland, OH	7,389	3,705	50%
WRAL	Raleigh/Durham, NC	7,275	5,223	72%
Philly.com	Philadelphia, PA	7,033	4,077	58%
WSYR	New York, NY	6,920	6,751	98%
New Orleans Times Picayune	New Orleans, LA	6,758	3,183	47%
Syracuse Post Standard	Syracuse, NY	5,937	4,493	76%
Orlando Sentinel	Orlando, FL	5,401	3,074	57%
Orange County Register	Orange County, CA	5,283	3,121	59%
Cincinati Post	Cincinatti, OH	4,696	4,021	86%
Minneapolis Star Tribune	Minneapolis, MN	4,642	3,362	72%
Newsday	Nassau-Suffolk, NY	4,330	2,970	69%
WSBTV	Atlanta, GA	4,031	2,974	74%
WPXI	Pittsburgh, PA	3,834	3,511	92%
WCPO	Cincinatti, OH	3,354	3,067	91%
WYFF	Greenville, SC	3,340	3,199	96%
Top 10 Outlets by Total Visits				
CNN	Atlanta, Georgia	205,591	8,160	4%
AOL News	New York, NY	110,056	6,333	6%
Major League Baseball	New York, NY	79,834	5,287	7%
NFL	Washington, DC	79,754	2,633	3%
Nickelodeon	New York, NY	79,076	5,106	6%
New York Times	New York, NY	72,419	8,206	11%
NBA	New York, NY	69,792	5,951	9%
Fox News	San Diego, CA	68,375	3,731	5%
Fox Sports	New York, NY	67,009	2,344	3%
Huffington Post Celebrity	New York, NY	66,714	3,126	5%

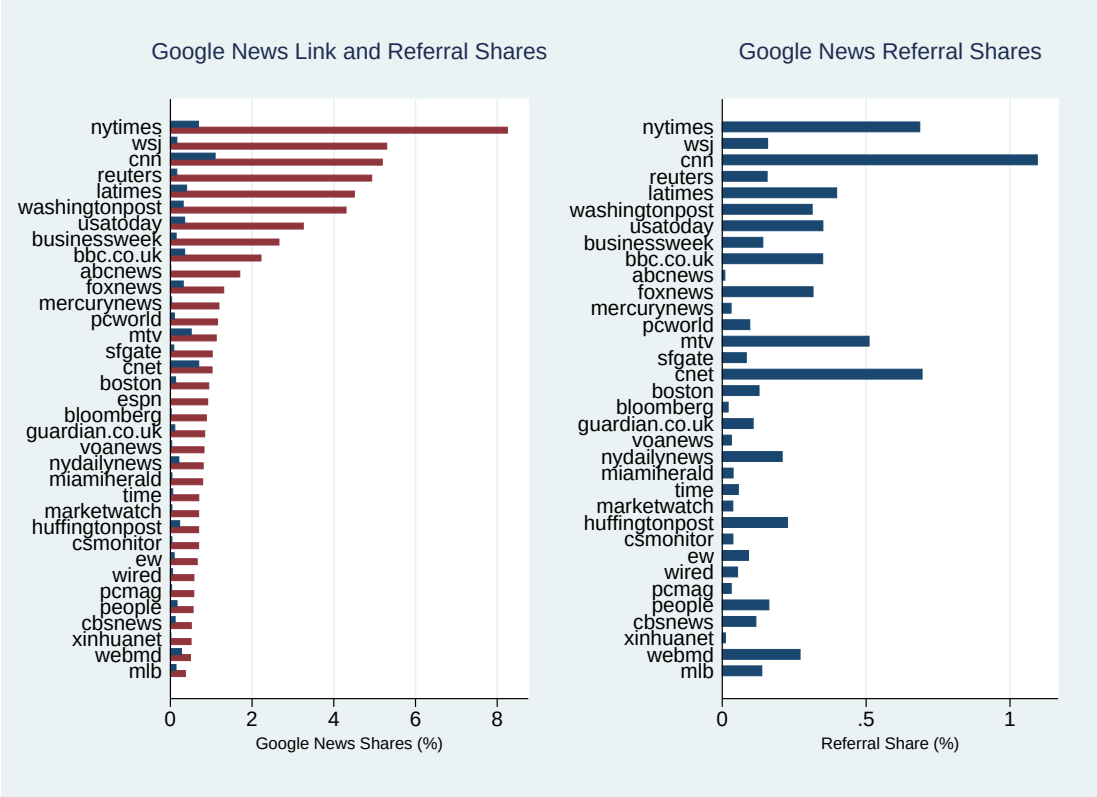
identified as a *Google News* referral. With this strategy, the imputed share of all news visits referred by *Google News* in the first six months of 2010 offers a measure of *Google News* use intensity that distinguishes households most likely to be affected by the redesign. Of the 6,407 news domains visited by households in the sample, 673 of them appeared during the sample period on *Google News*.

Our measure of *Google News* use intensity is limited in several ways. Most important, because the measure is based on referrals, it only captures behavior of users who "click through" to media outlets. This is a strong measure of *Google News* use, so will underestimate the effect of the redesign on local news consumption if users read more headlines but do not follow links. In some sense the measure will overstate the effect of the redesign on local news outlets if users switch from direct visits to local outlets to *Google News* viewing. But outlet visits are important for the structure of the market and at the center of the debate over whether aggregated links raise or lower direct visits, so in some sense referred visits is the correct measure. We return to this point in interpreting results.

Figure 3 summarizes *Google News* referrals from the perspective of both outlets and households for the most prominent outlets on *Google News*. The lower (red) shaded bar on the left image shows the share of all links scraped from the *Google News* page in the first half of 2010 (before the redesign) from each listed outlet. This measure can be viewed as a market share for each outlet on the *Google News* front page. National outlets such as the *New York Times*, *Wall Street Journal* and *CNN* dominate, with each outlet appearing on average in more than one place each day. Note that the link share on *Google News* for *YouTube* (10.7%) and the *Associated Press* (6.2%) do not appear on the charts because referrals cannot be inferred, but both outlets also appear daily on the page. The data show *Google News* links to be highly concentrated, with the top 25 outlets shown in the figure comprising 79% of links on the *Google News* page.

The upper (blue) shaded bars on the left image represent the share of all *Google News* referrals to each outlet, where referrals are measured using the procedure described above. In other words, we sum all of the outlet visits referred from *Google News* and calculate each outlet's share of these referrals. The right panel reproduces the statistic at a more legible scale. The referral share is a proxy for the click-through-rate for each outlet and thus offers one measure of the effectiveness of each *Google News* link. For example *CNN* and the *Wall Street Journal* appear with about the

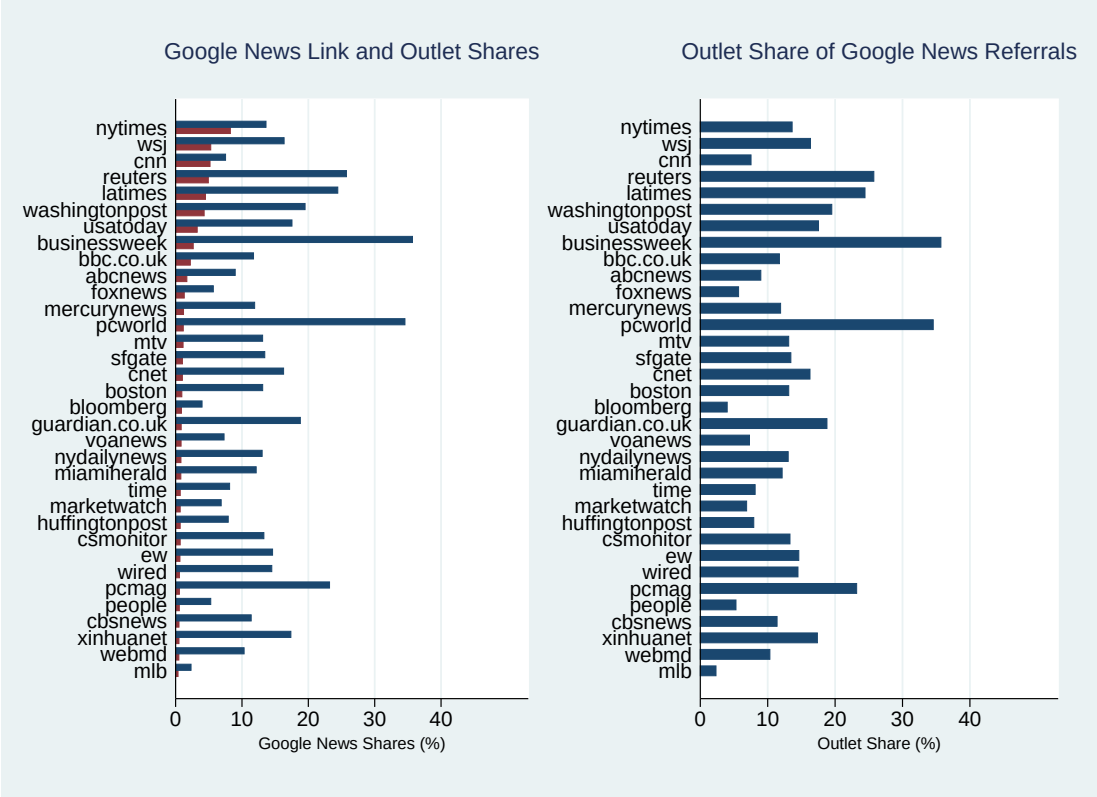
Figure 3: Link and Referral Shares on *Google News*



same frequency on the *Google News* page, but *CNN* receives a greater share of referrals. Similarly, *Fox News* attracts the same share of referrals as the *Washington Post* or *USA Today*, though its presence on *Google News* is smaller. Business outlets such as *Business Week*, *Bloomberg* and *MarketWatch* receive somewhat fewer referrals than indicated by their link share, but other specialty outlets such as entertainment (*Entertainment Weekly (EW)*, *People*), technology (*PC Magazine, Wired*) and health (*Web MD*) constitute a smaller fraction of the *Google News* page but attract more clicks per link than general news outlets. The pattern suggests that aggregators might shift attention to specialized content and expertise, which is an interesting avenue for further study.

The referral shares above offer a general measure of substitutability between aggregated links and source content. Another way of gauging the role of aggregators in the news market is to evaluate the share of visits to an outlet that are referred from *Google*

Figure 4: Link Shares and Clicks per User



News, shown in figure 4. As above, the lower (red) bar reflects the share on the *Google News* page. The upper (blue) bar in this case is the imputed share of all visits to the outlet referred from *Google News*. The figure emphasizes that aggregators are more important for some outlets than others. Business and technology periodicals have the highest imputed *Google News* referral share, with more than one third of outlet visits originating from the aggregator. General news sites such as *CNN*, while attracting a large share of *Google News* referrals, receive many more independent visits, hence rely less on the aggregator to feed demand.

It should be noted at this point that there are several sources of noise in measuring referrals. During the sample period, 1-3 snapshots are captured each day. If the *Google News* page is updated more often, we will miscount *Google News* referrals in the visit data. Second, the most popular news sites appear on the *Google News* page virtually every day, so there is little variation for these sites. Finally, our procedure attributes all

Google referrals to domains listed on *Google News* each day as *Google News* referrals, though it might be the case that users visited different news stories than those posted on *Google News*. In general, since the primary purpose of counting *Google News* referrals is to identify more active and less active *Google News* users, the noise introduced by the indirect measure does not undermine the basic empirical approach.

3.5 Summary Statistics

Table 2 and table 3 report summary statistics for domains and households, respectively. The top portion of 2 presents yearly totals for 2010, while the lower half summarizes daily data during the sample period. The average number of visits to each news outlet for the year is 842, with a range from the 5^{th} to 95^{th} percentile of 11 to 2,708. (Recall that outlets with fewer than 10 visits for the year day are dropped from the sample.) Total local visits average 117, with an average local share of 0.39. The probability a domain appears on *Google News* is 0.13, which is also the share of outlets in the sample with a *Google News* link. The average (imputed) share of *Google News* referrals is 0.002 over all domains. The probability of a *Google News* referral for the sub-sample of 673 outlets that appear on *Google News* is ten times higher at 0.02 of which 0.16 are local.

The daily data show similar patterns, with 2.4 visits to each domain each day and 0.36 local visits, giving an average daily local share of 0.37. Sites that appear on the *Google News* page are more popular, with an average of 10.7 links per day in the sample. Outlets appearing on *Google News* are also less local, with an average local share of 0.31 versus 0.37 for the whole sample.

Table 3 reports analogous measures for the sample of 43,087 households. Each household makes an average of 125.2 news visits over the entire year, with a local share of 0.11. The average share of local visits by households is much lower than the average local share of visits received by outlets (0.39), reflecting that many households frequently visit a small number of national outlets such as *CNN*, the *New York Times*, etc. The total share of visits referred by Google (news and search) is .29, while the imputed share linked from Google News is .03.

Daily data reflect similar patters. The sample of 38,648 users active during the 28 week study period make an average of 0.37 news visits and 0.06 local news visits per day. The probability of any news visit is 0.15 and the probability of a local news visit is 0.03 during the sample period. The daily local share during the sample period is

Table 2: Summary Statistics (Media Outlets)

	N	Mean	SD	5th%	95th%
2010 Totals: All Outlets (6,407 Domains)					
News Visits	6,407	842	4,912	11	2,708
Local News Visits	6,407	117	121	10	363
Local Visit Share	6,407	0.39	0.32	0.00	0.94
Probability of Google News Link	6,407	0.13	0.34	0.00	1.00
Google News Referral Share*	6,383	0.002	0.019	0.000	0.004
2010 Totals: Google News Outlets (673 Domains)					
Google News Referral Share*	673	0.022	0.055	0.000	0.122
Local Share of Google News Referrals*	474	0.163	0.317	0.000	1.000
Daily Sample: All Outlets (6,407 Domains)					
News Visits	627,886	2.40	19.22	0.00	8.00
Local News Visits	627,886	0.36	1.31	0.00	2.00
Local Visit Share	203,768	0.37	0.44	0.00	1.00
Daily Sample: Google News Outlets (673 Domains)					
News Visits	81,536	10.73	35.86	0.00	41.00
Local News Visits	81,536	1.15	2.57	0.00	6.00
Local Visit Share	61,187	0.31	0.39	0.00	1.00

* Calculated before re-design

Table 3: Summary Statistics (Households)

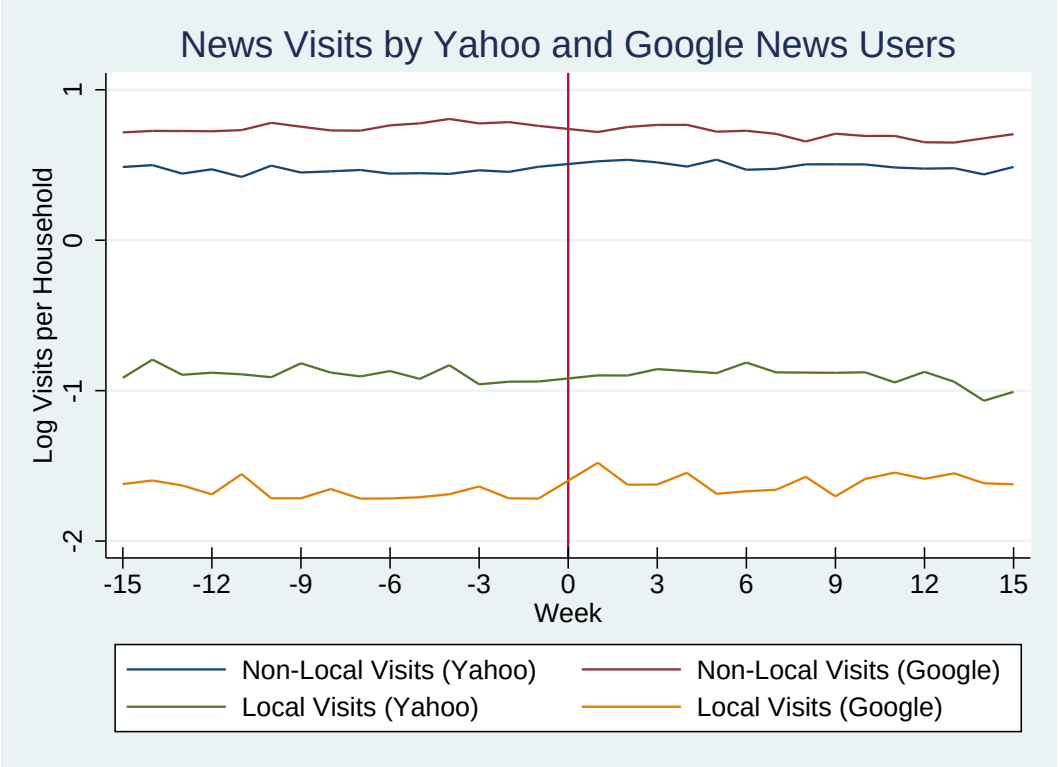
	N	Mean	SD	5th%	95th%
2010 Totals (43,087 Households)					
News Visits	43,087	125.22	275.98	3.00	490.00
Local News Visits	43,087	19.51	90.94	0.00	72.00
Days with News Visit	43,087	51.56	60.13	2.00	184.00
Days with Local News Visit	43,087	10.76	32.18	0.00	49.00
Local Visit Share	43,087	0.11	0.17	0.00	0.50
Google Referral Share	43,087	0.29	0.26	0.00	0.77
Local Share of Google Referrals	35,645	0.08	0.16	0.00	0.38
Google News Referral Share*	39,773	0.03	0.07	0.00	0.15
Local Share of Google News Referrals*	17,805	0.03	0.14	0.00	0.25
Daily Sample Data (38,648 Households)					
News Visits	7,613,656	0.37	1.38	0.00	2.00
Local News Visits	7,613,656	0.06	0.43	0.00	0.00
Probability of a News Visit	7,613,656	0.15	0.36	0.00	1.00
Probability of a Local News Visit	7,613,656	0.03	0.18	0.00	0.00
Local Visit Share	1,169,579	0.14	0.31	0.00	1.00
Google Referral Share	1,169,579	0.23	0.39	0.00	1.00
Local Share of Google Referrals	354,363	0.09	0.27	0.00	1.00
Unique Domains per Day	7,613,656	0.29	0.97	0.00	2.00
Unique Local Domains per Day	7,613,656	0.04	0.25	0.00	0.00
Unique Domains per Month	7,613,656	17.07	62.69	0.00	118.00
Unique Local Domains per Month	7,613,656	0.64	4.10	0.00	1.00

* Calculated before re-design

0.14.

The last four rows of the table show unique news domains visited per month and per day, a measure of variety in consumption. Households visit 17 news domains each month on average and 0.64 local news domains. Unique domains and unique local domains per day is substantially lower, at 0.29 and 0.04, respectively.

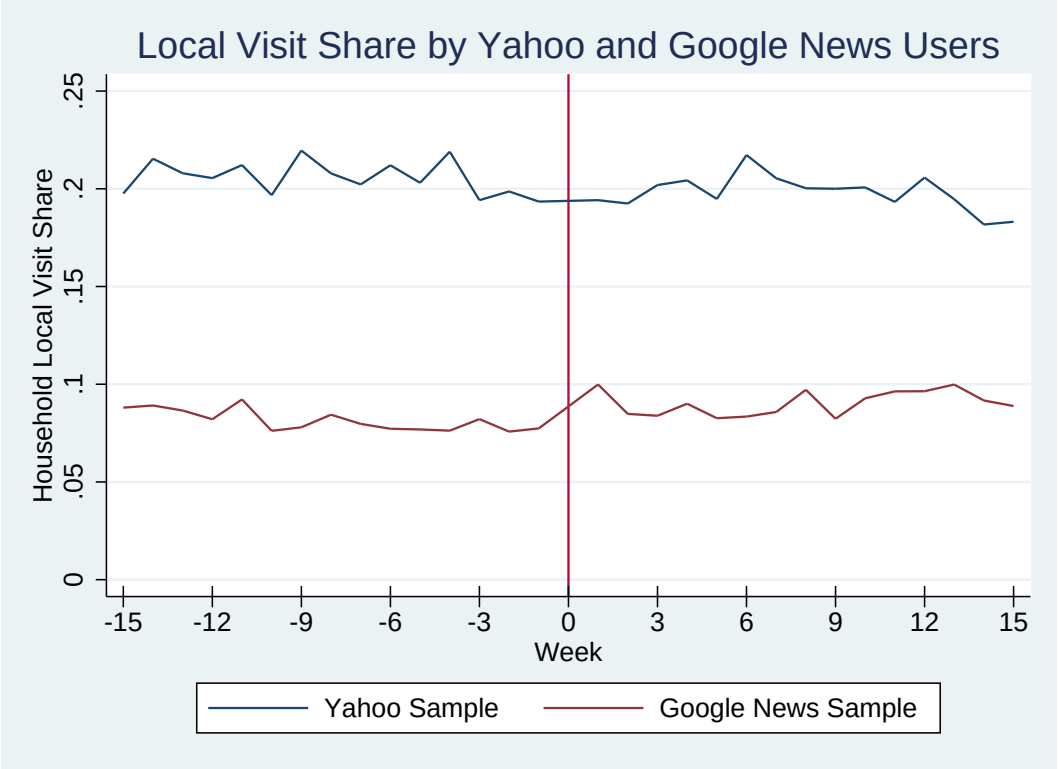
Figure 5: Non-Local and Local News Visits by Type



Note: Weekly non-local and local news visits for households above the 90th percentile in imputed *Google News* use before the June 30, 2010 redesign and users above the 90th percentile for *Yahoo!* use before the redesign.

Before turning to the empirical analysis, it is useful to consider trends in the raw data. Figure 5 shows (log) weekly non-local and local news visits for a sample of households before and after the redesign. The sample includes users above the 90^{th} percentile in *Google News* use before the redesign and users above the 90^{th} percentile for Yahoo use. Local and non-local visits are plotted over time, using a log scale to improve visualization given the relatively small number of local news visits. Figure 6

Figure 6: Local News Share by Type



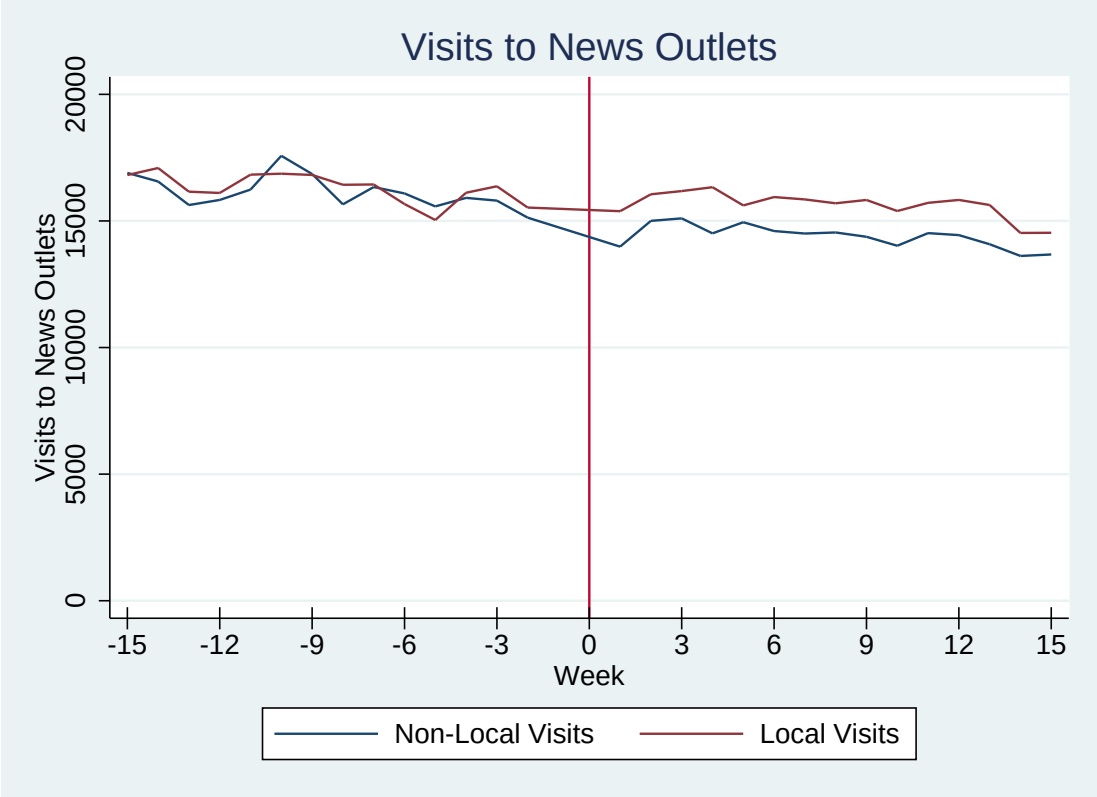
Note: Weekly local visit shares for households above the 90th percentile in imputed *Google News* use before the June 30, 2010 redesign and users above the 90th percentile for *Yahoo!* use before the redesign.

reports the local visit share for the same two groups of users.

A first point of note on the figures is that the local visit share is significantly higher for the Yahoo group (shown in the third line from the top (green)) compared to the Google group (fourth (yellow) line), with *Google News* users visiting more non-local and fewer local sites than the Yahoo users. The household visit data does not show sharp trends at the redesign, but there is some visual indication the gap between Yahoo and Google users in local visits shrinks after the redesign. In figure 6, the local visit share increases slightly for *Google News* users on the lower line at the point of the redesign. The effects in both figures appear small.

Visual patterns with the outlet data are more pronounced. Figure 7 and 8 show time trends for visits to news outlets by local and non-local users before and after the

Figure 7: Outlet Visits by Local and Non-Local Users



Note: Weekly local and non-local visits to news outlets.

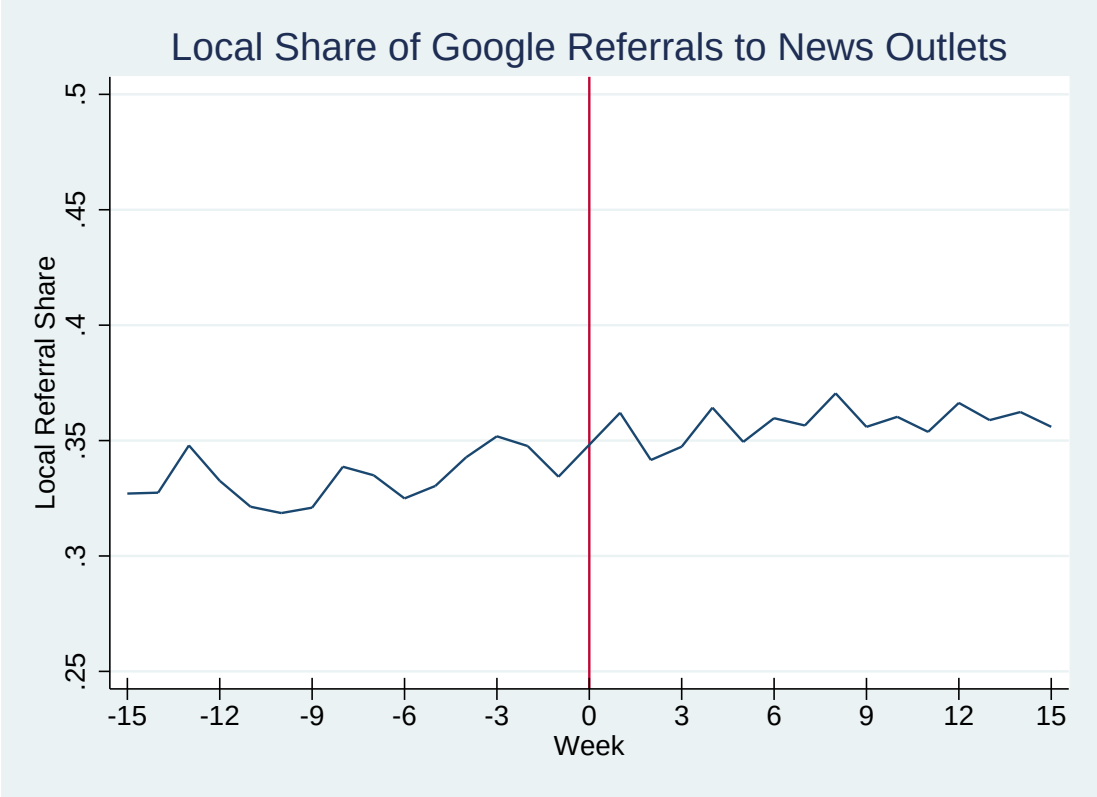
redesign. Local and non-local visits in 7 track closely before the redesign and diverge after. Figure 8 shows the local share of Google referrals (news and search) to domains over time. (The sample here is limited to outlets that have appeared on *Google News*.) There is evidence of a time trend in local referrals from Google, but also a step increase at the point of the redesign.

The next section outlines the estimation strategy for testing the patterns suggested by the graphs and summary measures of local news consumption.

4 Empirical Strategy

The goal of the analysis is to identify the effect of adding local news links to the *Google News* page on local news consumption. The baseline specification for estimation is a

Figure 8: Outlet Visits by Local and Non-Local Users



Note: Local share of Google referrals to news outlets.

standard fixed effects formulation:

$$Y_{it} = \beta_0 + \beta_1 Post + \beta_2 PostX + \tau + \gamma_i + \epsilon_{it} \quad (1)$$

where the independent variable Y captures a local news consumption measure for each household i each day t, $Post$ is a treatment dummy set to 1 after local news was added to the *Google News* site, and X is a treatment measure for households more strongly affected by changes to the *Google News* site. The error terms ϵ_{it} are assumed to be independent across users. A time trend τ and household fixed effect γ are included in all specifications, as are dummy variables for months and days of the week (not shown). Equation (1) is estimated with three measures of local news consumption (Y) and two treatment specifications (X). The independent variables are the number of local

news visits (with log transformation, $lnLvisits$), the probability of a local news visit ($ILvisit$), and the share of news visits to local sites ($Lshare$). The first two measures address the amount local news consumption, the third the share of attention to local relative to non-local outlets.

The choice of modeling approach for count data is not straightforward, especially when much of the variation is captured in the difference between zero (no local news visit) and one. In the tables that follow, we use a semi-log specification and a linear probability model to capture changes in local news consumption. Because of the large number of users who make no local news visits on any given day, local news visits are transformed as $lnLvisits = \ln(Lvisits + 1)$. With this specification, the coefficients can be interpreted as the percent change in local visits after the redesign. A second approach is to use a linear probability model, where the coefficient estimates show the effect of the redesign on the likelihood a treatment or control household makes a local news visit. The final specification measures the local share of all household news visits. All variables are measured at the household-day level.

Two identification strategies are adopted for each set of consumption measures. The first uses a treatment group and control group. The treatment group is comprised of frequent *Google News* users in the first half of 2010. The control group consists of frequent Yahoo users in the first half of 2010. We use the 90^{th} percentile as the cutoff for heavy users in both cases. Adding local links to *Google News* should affect the treatment group more than the control group.[5] In terms of equation (1), X is set equal to one for the *Google News* users and to zero for the control group of Yahoo users. A positive coefficient β_2 supports the hypothesis that the effect of the redesign is greater for the *Google News users*.

The ideal treatment group would be chosen to resemble the control group in every way but for the use of *Google News*. As in Chiou Tucker (2011), Yahoo users are chosen as a comparison group because they are expected to be more similar to *Google News* users than individuals who do not use intermediaries. However the number of intense Google and Yahoo users in the sample is small, about 7,500 households. To make better use of available data, we adopt a second strategy that identifies the effect of adding local links to the *Google News* site from the intensity of imputed *Google News* use

[5]Because *Google News* intensity is measured from referrals, individuals who use *Google News* but do not click through to individual web pages would not be recorded as frequent users. For this reason, the effect of the redesign on the control group may not be zero but should be less than effects on the treatment group.

Table 4: Summary Statistics for Treatment and Control Groups

	Google News Sample			*Yahoo Sample*		
	Households	Mean	Std. Dev.	Households	Mean	Std. Dev.
News Visits	3,593	0.255	1.13	3,885	0.259	0.99
Local News Visits	3,593	0.022	0.21	3,885	0.052	0.34
Local Visit Share	3,593	0.085	0.25	3,885	0.195	0.36
Pr. News Visit	3,593	0.112	0.32	3,885	0.128	0.33
Pr. Local News Visit	3,593	0.015	0.12	3,885	0.034	0.18
Hispanic Share	3,593	0.072	0.26	3,885	0.072	0.26
Black Share	3,593	0.127	0.33	3,885	0.133	0.34
High Income Share	3,593	0.187	0.39	3,885	0.122	0.33
Household Size	3,593	2.629	1.53	3,885	2.582	1.47

prior to the redesign. Intensity is measured as the share of all news visits referred from *Google News* (*GNVshare*) prior to the redesign, and in these specifications replaces the treatment dummy as X in equation (1). A positive coefficient β_2 then indicates the effect of adding local links to *Google News* for more avid *Google News* users.

The treatment group is comprised of households in the top decile of news visits referred by *Google News* in the first half of the year, a cutoff of 10% of news visits referred. The control group is comprised of users in the top decile of news visits referred by Yahoo, a cutoff of 33% of news visits referred prior to the redesign. Results are not highly sensitive to the cutoff, but the cutoffs do affect the sample size and the magnitude of measured effects.

Sample statistics for the *Google News* treatment group and Yahoo control group are shown in table 4. Looking first at visit counts, total news consumption by the two groups are similar, with an average of .26 news visits per day for each group. Local visits and especially the local visit share is higher for Yahoo users, 0.085 versus .195 for the Yahoo group. The two groups are similar in demographic characteristics, though the share of high income households is higher in the Google sample (19% vs 12%)

The time period of analysis is fourteen weeks before and after the June 30^{th} redesign when local links were added to the *Google News* site. While it is possible to estimate equation (1) over a longer time period, changes in local and non-local news consumption patterns into the fall season introduce considerable noise into the data and complicate causal links to the *Google News* redesign. Incremental change to Google search algorithms in early 2010 and in the fall further limit the ability to study longer

Table 5: Do Local Links Increase Local Visits Among *Google News* Users?

	Log Local Visits (1)	Pr. Local Visit (2)	Local Visit Share (3)
Post Treatment	.0014	.0008	.0215**
	(.001)	(.002)	(.008)
Post x GNews	.0025**	.0029**	.0033
	(.001)	(.001)	(.004)
Time Trend	–.0040**	–.0039**	–.0161**
	(.001)	(.001)	(.004)
Constant	.0167**	.0200**	.1356**
	(.001)	(.001)	(.005)
Households	7,635	7,635	7,478
N	1,504,095	1,504,095	177,440

Dependent variable in column 1 is transformed log of local news visits. Dependent variable in column 2 is probability of a local news visit. Dependent variable in column 3 is the share of visits to local outlets. All specifications include month, weekday and household fixed effects, standard errors clustered by household: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

time periods with specifications akin to equation (1).

A key assumption in the error structure of (1) is that the timing of the *Google News* redesign is uncorrelated with unobserved trends in local news consumption. The effect of *Google News* is also restricted to impact the mean of local consumption at a constant level. In all specifications, we include a continuous time trend as well as month and weekday fixed effects to account for temporal variation in news consumption.

5 Results

5.1 Basic Specifications

Estimates of equation (1) for the treatment and control group are shown in table 5. In the first row, *Post* (β_1) reports the effect of adding local links to *Google News* on the control group while the interaction $Post \times GNews$ (β_2) reports the result for *Google News* users. The first column measures the volume of local news visits and the second column the likelihood of a local news visit. The third column reports the effects on the local visit share. After the Google redesign, *Yahoo* users see no change in the number of local news visits while those with a high value of imputed *Google News* use make

Table 6: Does the Effect of Local Links Increase with *Google News* Use?

	Log Local Visits (1)	Pr. Local Visit (2)	Local Visit Share (3)
Post Treatment	–.0003	–.0001	.0073**
	(.001)	(.001)	(.003)
Post x GNews Share	.0111**	.0113**	.0222+
	(.003)	(.003)	(.013)
Time Trend	–.0026**	–.0027**	–.0045**
	(.000)	(.000)	(.002)
Constant	.0252**	.0271**	.1357**
	(.000)	(.000)	(.002)
Households	38,648	38,648	37,862
N	7,613,656	7,613,656	1,169,579

Dependent variable in column 1 is transformed log of local news visits. Dependent variable in column 2 is probability of a local news visit. Dependent variable in column 3 is the share of visits to local outlets. All specifications include month, weekday and household fixed effects, standard errors clustered by household: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

more visits on more days. Significance tests show the difference in effects across the two groups to be positive and statistically significant, indicating that local visits increased after the redesign among intense *Google News* users relative to intense Yahoo users. The magnitude of the visit effect is small, indicating an increase in local news visits by only 0.25%. The estimated effect on the likelihood of a local visit is larger, 0.0029 on a baseline of .015, an increase of 20% for the *Google News* group. The coefficient estimate in column (3) indicate the redesign increases the local visit share among the treatment group by 0.0033 on an average of 0.14 or 4%, but the effect in this case is not statistically significant.

Turning to the second set of estimates, table 6 reports results for the full sample of households where X is the imputed share of visits referred from *Google News* prior to the redesign. The post treatment indicator β_1 shows the baseline effect of the redesign for households that do not use *Google News* (*GNewsShare = 0)*. The interaction term *Post x GNewsShare* (β_2) shows the incremental effect for more intensive use. Recall that the average *Google News* referral share before the redesign is 3% with standard deviation of 7%. The referral share at the 95^{th} is 15%.

As in table 5, the top row shows effects of the redesign on control households. Now the second row shows the interaction of the post period and the share of all news

Table 7: Does the Effect of Local Links Increase with *Google News* Use? (Conditional Sample)

	Log Local Visits (1)	Pr. Local Visit (2)	Local Visit Share (3)
Post Treatment	.0058+	.0070*	.0073**
	(.003)	(.004)	(.003)
Post x GNews Share	.0579**	.0519**	.0222+
	(.019)	(.019)	(.013)
Time Trend	–.0080**	–.0068**	–.0045**
	(.002)	(.002)	(.002)
Constant	.1877**	.2001**	.1357**
	(.002)	(.002)	(.002)
Households	37,862	37,862	37,862
N	1,169,579	1,169,579	1,169,579

Dependent variable in column 1 is transformed log of local news visits. Dependent variable in column 2 is probability of a local news visit. Dependent variable in column 3 is the share of visits to local outlets. The sample is restricted to users-days where the household has made at least one news visit. All specifications include month, weekday and household fixed effects, standard errors clustered by household: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

visits referred by *Google News* prior to the change. Also as above, the first two columns consider the effect of adding local links on the level of local news consumption while the third column considers the local share. Both transformed local visits and the likelihood of a visit are estimated at about -0.001-.003, indicating baseline effects on non-*Google News* users close to zero.

Results indicate that the effect of the redesign increases with *Google News*. For a user with the average *Google News* referral share of 3%, the redesign increases local visits by $0.03x0.0111 = 0.03\%$. For *Google News* users at the 95^{th} percentile (15%), the redesign raises local visits by $0.15x0.0111 = 0.17\%$. As above, estimates with the linear probability model in column (2) are higher. For users with average *Google News* intensity, the probability of a local visit increases by 0.0003 or 1.1%. For users at the 95^{th} percentile, the probability of a local visit rises by 0.0017 or 6%.

Results for the effect of the redesign on the local visit share are shown in column(3) of table 6. In the full sample the redesign does increase the local share at a statistically significant level. At the *Google News* referral mean of .03, the redesign increases the local share by .006 on an average of 14%, or .48%. At the 95^{th} percentile of *Google News* intensity the redesign increases the local share by .00333 or 2.4%.

The overall effect in tables table 5 and table 6 are modest, especially given the low baseline. Thinking more deeply about the mechanism by which the *Google News* redesign affects consumption, it is likely that geo-targeted links on the page increase consumption on days when a user is consuming news, in other words on days when $Visits$ is non-zero and the local share variable is not missing. Restricting the visit and linear probability specifications to days with non-zero news visits increases the estimated coefficients considerably, by a factor of 5. Results are shown in table 7. With this approach, local visits for an average *Google News* user increase after the redesign by 0.17% and for a 95^{th} percentile user by 0.87%. The probability of a local visit given at least one news visit is similarly $0.03x0.0519 = 0.0016$ and $0.15x0.0519 = 0.0078$ for a mean and 95^{th} user, or 0.8% and 3.9% after adjusting for the higher baseline likelihood of a local visit.

We consider next a robustness check on this result. Recall that our definition of local news classifies a visit as local if the MSA for the household matched the home MSA for the outlet, with outlets receiving less than a 15% local visit share classified as non-local for all users. Rather than adopting a cutoff, we calculate a continuous measure of local consumption for each household-day by weighting all news visits by the local share of visits received by each outlet. The intuition here is that outlets with a very small local share contribute little to the weighted total, while visits to outlets of largely local interest contribute more heavily.

Estimates of equation (1) with weighted local consumption are shown in table 8. The first and third columns repeat the first column of table 5 and 6. The second and forth column show weighted local news consumption. Coefficients for the redesign, shown in the interaction term, are similar to the baseline estimates. These results suggest that the measured effects of the redesign are robust to the local definitions used in section 5.

Taken together, results above indicate that the *Google News* redesign did increase local news consumption overall and also shift attention from non-local to local news sources, with the largest effects among most intense users. The magnitude of the increase is small, with local news visits increasing less than a percent and the probability of a local visit by 4-6% even for intense *Google News* users. The effect of the redesign on the local share is about 2.4% for intense *Google News* users, indicating that adding geo-targeted links to the *Google News* page does shift attention on the margin in ways that can impact competition. The impact on local news outlets is considered in more

Table 8: How Local? News Visits Weighted by Outlet Local Share

	Local Visits (1)	Wtd. Local Visits (2)	Local Visits (3)	Wtd. Local Visits (4)
Post Treatment	.0014	.0015	–.0003	–.0005
	(.001)	(.001)	(.001)	(.001)
Post x Google	.0025**	.0016*		
	(.001)	(.001)		
Post x GNews Share			.0111**	.0090**
			(.003)	(.002)
Time Trend	–.0040**	–.0029**	–.0026**	–.0019**
	(.001)	(.001)	(.000)	(.000)
Constant	.0167**	.0123**	.0252**	.0197**
	(.001)	(.001)	(.000)	(.000)
Households	7,635	7,635	38,648	38,648
N	1,504,095	1,504,095	7,613,656	7,613,656

Dependent variable in columns 1 and 3 is transformed log local news visits. Dependent variable in columns 2 and 4 is transformed log of all news visits weighted by outlet local share. All specifications include month, weekday and household fixed effects, standard errors clustered by household: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

detail in Section 6.

5.2 Aggregators and Consumption Variety

The analyses above examined changes in the level of local news consumption. Another important dimension of consumption is the role of the aggregator in diffusing or concentrating attention across outlets. Source variety is one of the key elements of regulatory policy toward media markets because of its association with viewpoint diversity. Theoretically, the lower transaction costs associated with adding geo-targeted links can either increase or decrease variety. This can happen for two reasons. First, the convenience of aggregated links can reduce search efforts and focus attention on the aggregator page. If the aggregated local links come from a small set of sources, consumption variety might fall. If the aggregator instead searches and scans more widely than individual readers, consumption variety might increase. Second, geo-targeted links reduce search time, which can be used to visit additional local or non-local outlets.[6] The effect of aggregation on consumption variety is thus an empirical question that is

[6] George and Hogendorn (2012) consider the differential impact of aggregation and search on demand, with implications for targeted advertising.

Table 9: Do Local Links Increase Consumption Variety?

	Local Outlets/Day		Local Outlets/Month	
	(1)	(2)	(3)	(4)
Post Treatment	.0010	–.0002	–.0015	–.0044
	(.001)	(.001)	(.008)	(.003)
Post x GNews	.0025**		.0038	
	(.001)		(.007)	
Post x GNews Share		.0090**		–.0124
		(.002)		(.016)
Time Trend	–.0035**	–.0021**	–.0122**	–.0061**
	(.001)	(.000)	(.005)	(.002)
Constant	.0155**	.0212**	.0982**	.1250**
	(.001)	(.000)	(.004)	(.002)
Households	7,635	38,648	7,635	38,648
N	1,504,095	7,613,656	53,445	270,536

Dependent variable in columns 1 & 2 is unique local news outlets visited per day (log transform). Dependent variable in columns 3 & 4 is unique local news outlets visited per month (log transform). All specifications include month, weekday and household fixed effects, standard errors clustered by household: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

difficult to test without an exogenous shock to aggregation such as the *Google News* redesign.

To test whether the redesign increased variety in local news consumption, we re-estimate equation (1) with two new dependent variables, the number of unique local news outlets visited per day and the number of unique local news outlets visited per month. Table 9 reports the relationship between the *Google News* redesign and the number of unique local news sites visited per day and per month. The first two columns consider unique local news outlets per day and the third and forth consider unique outlets per month. The first column in each pair reports results for the treatment and control sample while the second reports results for the full sample with *Google News* use intensity.

The table reveals a distinct difference between the daily and monthly effects. All specifications produce effects close to zero for households that do not use *Google News*. For more intense *Google News* users, the number of unique local outlets visited per day increases after the redesign, but the redesign has no effect on the number of unique local outlets visited per month. The estimates suggest that increases in local news consumption come about by more frequent visits to a familiar set of outlets rather

than through introduction of new outlets into the household set. As suggested by figures 3 4, aggregators appear to concentrate attention rather than foster interest in the long tail of media sources.

The lack of a positive link between aggregation and consumption variety over an extended time horizon suggests that aggregators might play less of a role in product discovery than sometimes argued. However because local outlets are smaller in number and perhaps more familiar than non-local outlets, changes in local aggregation is not the ideal setting for testing the broader role of aggregators in discovery. The relationship between aggregators and consumption variety along different news dimensions is a topic worth of further analysis.[7]

6 Outlets

With evidence above that geo-targeted links on the *Google News* page increase local news consumption and the local news share by a small amount, we turn now to the question of how this shift in attention affects outlets.

As a first step, we consider the gross effect of an aggregated link on visits to outlets, estimated as:

$$Visits_{jt} = \alpha_0 + \alpha_1 Link + \tau + \gamma_j + \epsilon_{jt} \tag{2}$$

where $Visits$ is the number of visits to outlet j on day t, and $Link$ is an indicator for whether outlet j appeared on the *Google News* page on day t. We include outlet fixed effects γ_j and a time trend τ as well as weekday dummies and month dummies. Results are presented in table 10. A link on the *Google News* page is associated with a 12% increase in outlet visits from outside the home market and a 2% increase in visits from within the local area. The local visit share drops by .004, about 1.2%, though the precision falls to just below standard confidence levels.

Because the specifications in table 10 include outlet fixed effects, the coefficient estimates are robust to static differences across outlets such as topic, size, style, etc.

[7] It is worth noting that the specifications in table 9 can be used to study the role of the *Google News* redesign on all the overall number of unique outlets visited per day or month. The re-design appears to *reduce* the number of unique non-local outlets per month for heavy *Google News* users. This result, though beyond the scope of this paper, suggests that the increasingly tailored news environment such as the "News for You" feature introduced on the *Google News* page do serve to increase consumption but also to channel attention to fewer sources.

Table 10: Google News Placement and Local Visits

	Log Non-Local Visits (1)	Log Local Visits (2)	Local Visit Share (3)
Google News Link	.122**	.023**	–.004
	(.012)	(.005)	(.002)
Time Trend	–.024*	–.036**	–.006
	(.012)	(.009)	(.006)
Constant	1.105**	.364**	.277**
	(.022)	(.017)	(.011)
Outlets	832	832	832
N	219,648	219,648	164,469

Dependent variable in columns 1 & 2 are non-local and local visits to news outlets per day (log transform). Dependent variable in column 3 is local share of visits to news outlets per day (log transform). All specifications include outlet, month, and weekday fixed effects. Standard errors clustered by outlet: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

However it is also likely also the case that both a link on the *Google News* page and direct outlet visits are driven by unobserved "newsworthiness" of writing on any particular day, so the coefficients in table 10 will tend to overestimate the effect of aggregation on outlet visits. Nonetheless, the large magnitude of the shock suggests that some overall complementarity between aggregated links and outlet visits is likely. Perhaps more important for competition, a *Google News* link alters the audience for the media outlet toward viewers outside the home market.

We would like to test directly whether geo-targeting on *Google News* shifted attention to local outlets in an economically meaningful way. We are hindered in this by the structure of available data: after the redesign, the *Google News* page shows different local links in every market, but our scrapes record only the layout of the page as seen in the San Francisco metro area.[8]

With this limitation, we tackle the question in two ways. First, we look at changes to the *Google News* page after the redesign, testing whether local outlets are less likely to appear on the main *Google News* page after the redesign when they were moved to the sidebar. This test allows us to consider whether non-local visits to outlets would be expected to fall as local visits rise. Results are shown in table 11. The dependent

[8]We observe a handful of observations in other markets, but not enough for systematic evaluation. We also have only 83 days of data from July-December as compared to 183 days from January 1-June 30. It appears that it took some time for the Wayback Machine to adjust to the reformatted *Google News* page, though we cannot be sure the redesign caused the drop in scrape frequency.

Table 11: *Google News* Redesign and Local Outlets

	Probability of a Google News Link	
	(1)	(2)
Post Redesign	.007	.002
	(.006)	(.006)
Post x HomeShare	–.041**	–.039**
	(.006)	(.008)
Post x TotalVisits		.001
		(.001)
Post x HomeShare x TotalVisits		.007
		(.008)
Time Trend	.007+	.007+
	(.004)	(.004)
Constant	.043**	.043**
	(.002)	(.002)
Outlets	832	832
N	219,648	219,648

Dependent variable is likelihood of a link on the *Google News* front page. All specifications include month, weekday and outlet fixed effects, not shown. Standard errors clustered by outlet: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

variable in this regression is an indicator for a link on the *Google News* page, and we interact the post redesign indicator with both the home market share of the outlet and total visits (in millions) before the design change. (The specifications include outlet fixed effects, so the baseline impact of outlet home share or size cannot be estimated.) Coefficient estimates indicate that after the redesign outlets that had a greater share of visits from within the home market become less likely to appear on the Google News front page. The test confirms that the redesign did indeed shift geo-targeted local news away from the main page into the sidebar.

We look next at the effect of the redesign on outlet visits directly. Although we cannot construct a treatment and control group as readily as for the households analysis, we can still make use of the fact that referrals from Google should change relative to referrals from Yahoo. It is useful for this comparison to aggregate the visit data for each outlet to the weekly level and restrict attention to outlets with referrals from both aggregators in a given week. We can then estimate the trends documented in figures 7 and 8 directly. More specifically, we estimate equation (3), where Y is a measure of demand for outlet j in week t, and $Post$ is an indicator for the redesign. Given the

Table 12: *Google News* Redesign and Local Google Referrals

	NLVisits (1)	LVisits (2)	LShare (3)	Google LShare (4)	Yahoo LShare (5)
Post Treatment	–.041*	.042*	.019**	.020+	.006
	(.016)	(.021)	(.006)	(.010)	(.011)
Time Trend	–.004*	–.007**	–.001	.000	–.000
	(.002)	(.002)	(.001)	(.001)	(.001)
Post x Trend	.005+	.005	–.000	–.002	.000
	(.002)	(.003)	(.001)	(.001)	(.001)
Constant	2.094**	2.191**	.521**	.401**	.467**
	(.013)	(.015)	(.004)	(.007)	(.008)
Outlets	1,968	1,968	1,968	1,968	1,968
N	15,725	15,725	15,725	15,725	15,725

Dependent variable in column 1 is weekly non-local visits (log transform). Dependent variable in column 2 is weekly local visits (log transform). Dependent variable in column 3 is the weekly local visit share. Dependent variables in columns 4 and 5 are the weekly local share of Google and Yahoo referrals, respectively. Sample restricted to local outlets with positive Google and Yahoo referrals for the week. All specifications outlet fixed effects. Standard errors clustered by outlet: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

evidence of underlying trends in the figures, we include an interaction term to evaluate whether the redesign impacts the trends in the audience measure. composition. Table 12 presents results. The first two columns show the effect of the redesign on non-local and local outlet visits (log transform). Visits from outside the home market for the outlet fall by about 4% and visits from inside the home market rise also by 4%, with an increase in the local share of 0.019 on an average of 0.52, or 3.7%. The last two columns in the table show the share of local referrals from Google and Yahoo (news and search) before and after the redesign. The local share rises for Google referrals by 0.02 (5%) and remains unchanged for Yahoo referrals, providing further evidence of a causal link. The interaction term in most specifications is not statistically significant, though the estimate in column (1) indicates that the decline in non-local news consumption associated with the redesign shrinks over time.[9]

$$Y_{jt} = a_0 + a_1 Post + \tau + a_3 Post \cdot \tau + \gamma_j + \epsilon_{jt} \tag{3}$$

As a final look at the nature of the redesign on local news outlets, we study the

[9]The results in table 12 are robust to placebo specifications of the timing for the *Google News* redesign.

Table 13: Geo-targeting in San Francisco

	Google		Non-Local		Local		Non-Local		Local	
	News Links		Referrals		Referrals		Visits		Visits	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
San Jose Mercury News	0.58	0.82	2.38	1.89	0.25	0.29	10.44	7.44	1.96	1.76
San Francisco Chronicle	0.65	0.67	7.15	5.32	0.34	0.86	22.42	18.26	5.65	4.56

Note: Links, referrals and visits after July 1 are taken from 82 scraped days from July - December 2010.

case of the San Francisco metropolitan area, where due to the IP address of the Way-back Machine we observe geo-targeted *Google News* links after the redesign. The *San Francisco Chronicle* and *San Jose Mercury News* dominate the local news sidebar, appearing in 82 out of the 90 links scraped after the re-design.[10] Returning to the daily data, table 13 reports the daily likelihood of appearing anywhere on the *Google News* page before and after the redesign along with the daily non-local and local referrals, and also non-local and local visits. The summary statistics indicate that both sites have a higher likelihood of appearing on *Google News* after the redesign due to their prominence on the geo-targeted bar. As in the overall data in table 12, Google referrals from outside the San Francisco area fall and referrals from inside the region increase for both outlets after the redesign. By comparison, total non-local and local visits to the both outlets fall slightly after the redesign. Regression estimates show the differences in the table to be statistically significant.

The results in this section highlight that aggregation can affect not only the number of visits to an outlet, but also the composition of the audience. Geo-targeting shifts attention to more local sources, which reduces non-local and increases local visits to outlets referred by the aggregator. Although the percentage drop in non-local attention is estimated to be similar to the increase in local users, the net effect on outlets becomes a question of the relative value of different user types. Figures 3 and 4 show that aggregators play a bigger role in demand for some outlets than others, so the economic significance of the audience shifts measured here are likely to vary substantially across outlets. The broad rise of paywalls suggests that both consumer willingness to pay

[10]The other six markets where limited geo-targeted links are available are not so highly concentrated, for example the 43 links scraped in Kentucky region cover 27 outlets, and 17 links scraped in Indiana span 11 outlets. In both of these markets the geo-targeted links are taken from radio, television and print outlets. The sample of scrapes after the redesign is generally too small to make inference about the composition of geo-targeted links on the *Google News* page from this study, but the topic warrants continued research.

and advertiser value are greater for local visits, but as yet no research can identify this effect.

7 Conclusion

This paper exploits a rare opportunity to study the effect of aggregation on the market for news, an overhaul of the *Google News* site that added geo-targeted local news links and headlines to the front page. Results indicate that adding local news links increases both the amount and share of local news consumption among active *Google News* users. Although the magnitude of the estimated effects are modest, there is no evidence in our sample that aggregation reduces demand for original content. We find that adding geo-targeted links increases the number of different local outlets visited per day, but not the number of unique sites visited per month, suggesting that increases in local news consumption arise from more frequent visits to familiar news outlets rather than visits to additional news providers. The result implies that aggregators play less of a role in product discovery than is sometimes argued, at least in the narrow context of local news. We also find evidence that the redesign raises visits from within the home market and reduces visits from elsewhere. Aggregation can alter not only the size but the composition of the audience for local media.

Our findings are limited by the nature of available data, which provides only top-level domains rather full article links to content. For this reason we cannot fully distinguish search from aggregator referrals, and though our identification strategy relies on changes to *Google News* that do not affect search, much could be learned from further analysis with data collected by Google and other digital firms. We also cannot identify the geo-targeted local links posted on Google News after the redesign in more than a handful of markets, which would illustrate more clearly the winners and losers from aggregation. Many interesting avenues for future study remain.

References

Alaouiy, Larbi and Fabrizio Germano. *Time Scarcity and the Market for News* Working paper, November 20132.

Athey, Susan and Markus Mobius. *The Impact of News Aggregators on Internet News*

Consumption: The Case of Localization. Working paper, February 2012.

Cameron, A. C. and Pravin K. Trivedi (1998*), Regression Analysis of Count Data*, Econometric Society Monograph No.30, Cambridge University Press, 1998.

Chiou, Lesley and Catherine Tucker. *Copyright, Digitization, and Aggregation.* NET Institute Working Paper No. 11-18, December 2011.

Chrysanthos Dellarocas, Zsolt Katona, and William Rand. Media, Aggregators And The Link Economy: Strategic Hyperlink Formation In Content Networks. Working Paper, February 2012.

George, Lisa M. and Christiaan Hogendorn (2012). "Aggregators, Search and the Economics of New Media Institutions." *Information Economics and Policy,* Vol. 22: Issue 1 pp. 40–51.

George, Lisa M. (2008) "The Internet and the Market for Daily Newspapers." *The B.E. Journal of Economic Analysis & Policy,* Vol. 8, No. 1 (Advances), Article 26.

George, Lisa M. and Joel Waldfogel (2006). "The *New York Times* and the Market for Local Newspapers." *American Economic Review,* Vol. 96, No. 1, pp. 435–477.

George, Lisa M. and Joel Waldfogel (2008). "National Media and Local Political Participation: The Case of the *New York Times*" in Roumeen Islam, ed., *Information and Public Choice: From Media Markets to Policymaking.* Washington, DC: World Bank Publications, pp. 33–48.

Jeon, Doh-Shin and Nikrooz Nasr Esfahani. *News Aggregators and Competition Among Newspapers in the Internet.* Working paper, June 2012.

Rutt, James. *Aggregators and the News Industry: Charging for Access to Content* Working paper, October 2011.