Who Decides What News You Read: Search Engines and Media Markets[1]

Joan Calzada[2] Nestor Duch-Brown[3] Ricard Gil[4]

January 2025

Abstract

Search engines are important access channels to the growing amount of content and information available in the internet. Yet, we know little about how search algorithms match consumer queries to content providers. Using daily traffic data for 606 news outlets from 15 European countries, we analyze whether Google can influence organic search visits by exploiting quasi-exogenous variation in news outlets' indexation caused by nine core algorithm updates rolled out between 2018 and 2020. We show that Google core updates overall reduced the number of keywords (queries) for which news outlets occupy one of the top 10 organic search results positions. Given the positive impact that the number of top keywords has on traffic, these updates led on average to a decrease in the number of search visits and overall visits to news outlets. Our findings are consistent with the algorithm prioritizing "expertise" over "authority" (as defined by industry practitioners) when ranking web pages for given a query. We also find that Google core updates have no impact, on average, on media market concentration, although the effects of different core updates range between a reduction and an increase of 1% per core update.

[1] We are grateful to Luis Aguiar, Rodrigo Belo, Luis Cabral, Gerard Llobet, Estrella Gómez, Tim Meyer, Lisa George, and Joan Santalo for their comments and suggestions, as well as to seminar and conference participants at the Toulouse Business School, Universidad Complutense de Madrid, the Paris 13th Conference on Digital Economics, Munich Summer Institute 2022, XXXVI Jornadas de Economia Industrial, Workshop on Digital Markets (Madrid, 2023), Queen's Workshop on the Economics of Media (Kingston), IIOC in Washington DC, CESifo Area Conference on the Economics of Digitalization (Munich), 12th Lisbon Meetings, the Third Catalan Economic Society Conference (Barcelona), the 38th EEA Meetings in Barcelona and the Joint Research Center (EU). We gratefully acknowledge financial help from Center for Technology, Innovation and Competition (CTIC) from the University of Pennsylvania and the Knight Foundation as well as the Spanish Ministerio de Ciencia e Innovación (MCI), Agencia Estatal de Investigación (AEI) and the Fondo Europeo de Desarrollo Regional (FEDER) (Grant: PID2021-128237OB-I00). The usual disclaimer applies.

[2] Universitat de Barcelona, Spain; calzada@ub.edu.
[3] Joint Research Centre (JRC), Spain; Nestor.DUCH-BROWN@ec.europa.eu.
[4] IESE Business School, Spain; rgils@iese.edu.

1. Introduction

The digital revolution has radically changed the way consumers access and consume media content such as music, movies, and books, among others (Waldfogel, 2017). On the one hand, the amount and variety of "niche" contents available to consumers has increased. On the other hand, the digitalization of media is directly responsible for the emergence of search engines, aggregators, and recommender systems, which help consumers navigating through the information clutter. While the former has expanded the "long tail" of products available to consumers (Anderson, 2004; Peltier et al., 2016; Zhang, 2018; Goldfarb and Tucker, 2019), the increasing importance of aggregators and curators might have facilitated the creation of "superstar" firms that dominate some markets (Fleder and Hosanagar, 2009; Brynjolfsson et al. 2011, Martin et al., 2024).[5]

The news media market is a paradigmatic example of the transformation caused by new digital technologies. In this context, the substitution of traditional media (printed newspapers, radio, and TV) for online media (digital newspapers, social media, news aggregators) has had a wide and diverse range of effects, with potential implications for the creation and quality of journalistic content as well as for the development and functioning of democratic institutions. Recent studies have analyzed the effects of digitalization on competition in the media market (Athey et al., 2017, Chiou and Tucker, 2017; George and Hogendorn, 2020; Calzada and Gil, 2021), the quality of journalism (Cagé et al., 2020; Bandy and Diakopoulos, 2020), and social interactions and political polarization (Gentzkow and Shapiro, 2011; Boxell et al., 2017; Peterson et al., 2019; Zhuravskaya et al., 2020). Additionally, the expansion of digital newspapers has brought along the development of algorithm-based platforms that offer readers tailored recommendations of news contents. For instance, nowadays in the European Union, only 45% of news outlets' visits comes from consumers that directly browse the news sites' address when looking for news contents. The remaining visits split between organic search traffic from search engines (35%), visits from social media (12%), and visits from paid search and advertising display.[6] Despite the relevance of these changes, little is known about the effects that search engines and social platforms might have in media markets and their media power (Prat, 2018; Allcott et al., 2024). Some exceptions (Sismeiro and Mahmood, 2018; Cagé, Hervé and Mazoyer, 2020; Dujeancourt and Garz, 2023; Germano et al., 2022; Chiou and Tucker, 2024; Guess et al., 2024) examine how Facebook and Twitter have modified news production and readers' engagement as well as their potential effects on misinformation and polarization.

In this paper, we investigate whether and how search engines have the capacity to shape the media industry by influencing how users discover and consume online news. We do so by examining how changes in Google Search's indexation activity affects the volume of organic search visits both at the individual news outlet level and at the market level. Search engines are designed to provide users with the most accurate and useful results

[5] New research has also shown that the presence of dominant firms, such as Amazon, might favor the specialization in narrow niches (Bondi and Cabral, 2022).
[6] Own calculations, based on SimilarWeb data.

possible for their queries. When a query is made, Google's algorithms determine the ranking of the links to the news outlets in the search engine results pages (SERP). Google Search facilitates the discovery of news outlets and can broaden the variety of contents available to readers. However, the ability of small and local news outlets to appear in top positions in Google's SERP depends on how it balances the "expertise" and proximity of small and niche outlets against the "authority" of larger outlets. Even though search engines' algorithms are expected to be free of any bias towards particular individuals, organizations, or viewpoints,[7] search engines index news outlets according to their technological performance and the consumer engagement, which likely biases results in favor of outlets that are better at identifying the interest of large audiences and that have a stronger domain reputation (Leung and Stumpf, 2023). Thus, our paper examines whether recent changes in Google's core algorithms have prioritized the "authority" of major news outlets, creating a "superstar" effect, or if they have favored the "expertise" of local and niche outlets, thereby enhancing the "long tail" effect as documented in the literature in media markets. Moreover, since changes in Google's indexation can significantly modify where consumers obtain news, our analysis can be interpreted as an empirical test of Google's media power, beyond its market power.

Our paper examines the impact of the nine global core algorithm updates rolled out by Google between 2018 and 2020. According to Google, these updates were global in that they affected all regions and languages worldwide and they did not target specific search queries or website. Instead, on each occasion, these updates were designed to improve Google's ability to evaluate content, ensuring that search results deliver the most relevant information to users. We leverage variation in indexation from these quasi-natural experiments to examine how changes in news outlets' indexation affect organic search visit traffic and the distribution of traffic across outlets.[8] Specifically, we analyze whether Google core algorithm updates reduced the skewness of the distribution of search traffic across news outlets, or whether they contributed to a thicker "long tail" of niche and specialized content.

Our study draws from a rich data set obtained from SimilarWeb containing information for 606 news outlets in 15 European countries. This data set includes daily information about news outlets' direct, organic search and social network visits, and can distinguish between desktop and mobile traffic. It also contains monthly information about traffic generated by each search keyword in each news outlet in our sample. We complement these data with information from Ahrefs (a well-known search engine optimization tool) on the daily number of keywords (each keyword reflects a consumer query) for which a news outlet occupies one of the first 10 positions (or 11-100 positions) in Google's organic search results.

---

[7] Several studies have examined potential sources of bias in search algorithms, including studies on racial, gender, and political bias (Barocas and Selbst, 2016; Noble, 2018 and Tolan, 2019).

[8] According to Google, there is nothing site owners can do to increase their search traffic or to recover their position after an update. "Sometimes, we make broad changes to our core algorithm. We inform about those because the actionable advice is that there is nothing in particular to "fix," and we don't want content owners to mistakenly try to change things that aren't issues." See https://t.co/ohdP8vDatr (Google Search Liaison @searchliaison, October 11, 2018).

Our paper makes two main contributions. First, we use a sound identification strategy to econometrically isolate the effect of Google's search algorithm core updates on the number of organic search visits received by news outlets. One important difficulty in studying how Google's indexation affects search visits is that these visits may be correlated with relevant, yet unobserved, characteristics of new outlets or the contents they publish. News sites compete for high-traffic keywords and invest considerable resources in optimizing their search results.[9] To deal with this endogeneity problem, and given the absence of a valid control group, we adopt a two-stage identification strategy. Specifically, we use Google's core algorithm updates from 2018 and 2020 as exogenous shifters for changes in the number of keywords that news outlets had on top organic search positions. Because these updates directly impact the indexation of news sites, these episodes provide a source of exogenous variation in the number of keywords in top SERP positions. Then, in the second stage, we use the exogenous shift in the number of keywords to estimate the true relationship between search traffic and number of keywords appearing in the top 10 positions of organic search queries. This approach allows us to test whether Google has power to influence visit traffic patterns of media outlets.

Because Google (as any other search engine) is rather secretive about how its algorithm updates balance authority and expertise in evaluating news outlets,[10] we first validate our hypothesis that a shift towards more weight on authority (expertise) may benefit larger (smaller) outlets. For this purpose, we precede our main empirical analysis with a simple correlational exercise focusing on two updates for which there is less opacity about Google's objectives: the August 2018 core "medic" update and the May 2020 core "pandemic" update. According to industry sources, the "medic" update aimed to protect users from harmful information by prioritizing web pages with higher domain authority, thereby benefiting the indexation of websites with high-quality content, such as news outlets. By contrast, the "pandemic" update aimed to ensure that, after the breakout of the COVID-19 pandemic, local information and news were not hidden below larger websites offering similar content. As a result, it prioritized the "expertise" of local news outlets over the "authority" of national news outlets. Consistently with these objectives, we find search visits to European news outlets increased after the 2018 "medic" update. In contrast, we also found that search visits decreased following the 2020 "pandemic" update, and that search visits of national outlets and regional outlets experienced a larger decline than sports and business outlets. These patterns are consistent with specialized

[9] News outlets may gather data on keyword volume and trends, analyze competitors' targeted keywords, and search for keyword and phrase combination that can increase their visits. For example, they can modify their headlines to maximize audience reach or respond to reactions on social networks (Cagé, Hervé and Mazoyer, 2020; Leung and Strumpf, 2023).

[10] Interestingly, and consistently with the current opacity of Google's search engine algorithm, Sergey Brin and Larry Page in a paper published in 1998 while still students at Stanford stated "Since it is very difficult even for experts to evaluate search engines, search engine bias is particularly insidious. A good example of this claim was OpenText, which was reported to be selling companies the right to be listed at the top of the search results for particular queries (Marchiori, 1997). As a result of its business model, OpenText became an unviable search engine. Yet, more subtle biases may be more difficult to detect and therefore be tolerated in the market for search engines (Brian and Page, 1998).

outlets having greater expertise in local queries, and therefore, being indexed more favorably as a result of the "pandemic" update.

Then, we proceed with our main empirical analysis where we estimate the average impact of the nine algorithm updates implemented during our sample period. Following SEO specialists, we categorize the nine updates into two groups: the three largest core updates as "*Big*" updates and the remaining six smaller updates as "*Not Big*". Our results show that core updates had on average a negative causal effect on the number of keywords in top 10 positions in organic search results. This decrease in top keywords meant a decrease in Google's organic search ranking and, consequently, a reduction in the visibility of news outlets in Google results pages (relative to all other pages indexed by Google). The magnitude of this effect varies across different types of news outlets (e.g. national or regional), different features (national rank, domestic traffic, or traffic from Google), and different countries.

Our second contribution is our analysis of how Google' core algorithm updates affect media market concentration across European countries. We find that the three "*big"* core updates reduced market concentration by 1%, while the rest of updates increased it by 0.8%, nearly offsetting the overall impact. At the individual country level, the effects were mixed: marked concentration decreased in Finland, Germany, and Greece but increased in Netherlands and Portugal. Moreover, we find the salience of national outlets decreased with *Big* updates. Overall, our results suggest that the impact of Google's updates on market concentration, and thus competition in the media market, is specific to the country and update. These findings are at odds with current policy debate in Europe (and elsewhere) claiming that the Google search engine consistently and gradually favors larger outlets over smaller outlets. Our findings do not support these claims.

Additionally, using monthly data from Spanish news outlets, we provide two additional findings that complement our primary analysis. First, we examine how Google´s core updates affect concentration of search visits at the keyword level. While "*big*" core updates reduced search visit concentration at the keyword level by 4% (followed by a quick increase in concentration within two months), consistently with our site level analysis, "*big*" core updates increased the concentration of visits from keywords in top positions by 6%. This implies that Google's updates increased market concentration for keywords likely to generate more search traffic to news outlets. Second, using monthly search-engine traffic data for each outlet in our sample, we show that core updates reduced market concentration of Google Search by 2% on the month of the updates (relative to other engines such as Bing, DuckDuckGo and Yahoo), paired with an increase in market concentration two months later.

In summary, our analysis and findings have important policy implications. We demonstrate that algorithm updates can drastically reduce visits to news outlets and impact market competition. The role of search engines in media markets are far more important than in other industries. In fact, where consumers and voters obtain their news and information may tilt their political attitudes and voting intentions, alter their perceptions and opinions, and reinforce stereotypes (Prat, 2018; Bandy and Diakopoulos,

2020). News sources can also affect how voters come to be informed during elections and which problems are perceived more relevant for the public opinion. Our analysis is important for understanding how search engines shape media markets.

Additionally, our findings have implications for ongoing competition policy regarding the influence of digital platforms that rely on indexation algorithms and the potential introduction of biases in their organic search results. We show that changes in Google's indexation do not always increase media market concentration, and, thus, competition policy should be case-specific. As a matter of fact, Google Search has faced intense antitrust scrutiny from both U.S. and European authorities (Yun, 2018). At the beginning of the 2010s, the U.S. Federal Trade Commission (FTC) investigated Google for alleged biases in search results, but finally closed the case without action. In 2015, the EC fined Google $2.7 billion for abusing its dominance in Google Shopping (Scott, 2017). The Commission found that Google had given an unfair advantage to its comparison-shopping service by placing its results above those of competitors, effectively diverting traffic away from them (European Commission, 2017). The EC concluded that no alternative sources of traffic could compensate for the loss of visibility in Google's generic search results.[11]

Finally, acknowledging the fact that our empirical analysis focuses solely on Google Search updates, it is crucial to consider the external validity of our findings. Between 2018 and 2020, Google Search held a 93% market share in Europe, suggesting our results may be broadly applicable within that context. Furthermore, our sample encompasses 606 news outlets, representing a significant portion of news readership across 15 countries. This robust sample enhances the generalizability of our findings. Therefore, we believe concerns regarding the external validity of our research are mitigated.

The article is structured as follows. In section 2, we review the existing literature. Section 3 outlines the main features of Google Search and explain how Google updates its indexation algorithms. We present our data and empirical strategy in section 4. Section 5 and 6 shows our findings on the impact of Google's core algorithm updates on search, desktop, and mobile traffic and media market concentration, respectively, in European news media. Section 7 provides additional results at the keyword-month level and at the search engine level. Finally, section 8 concludes.

2. Literature review

This paper contributes to several streams of literature. First, we build on and contribute with our empirical findings to a theoretical literature examining the existence of bias in search engines (Belleflamme and Peitz, 2018 and 2021). Prior theoretical work has shown that search engines can adjust the quality of their organic results to favor sponsored search from which they obtain larger profits (Xu, Chen, and Whinston, 2012; Taylor, 2013; and White, 2013). In a similar line, Cornière and Taylor (2014) and Burguet et al. (2015)

[11] Decarolis, Li and Paternollo (2023) examine several interventions taken in the European Economic Area, Russia, and Turkey to limit its default use in the mobile devices.

analyze biases in search results when search engines are vertically integrated with a seller. Other papers have shown that search engines may strategically degrade the quality of their search results (Chen and He, 2011; Eliaz and Spiegler, 2011). Hagiu and Jullien (2011) show that intermediaries can introduce some noise in the search process (i.e., to divert search) to increase the number of consumers' searches and to influence the strategic choices (i.e., pricing) of affiliated sellers. Some recent theoretical papers have also examined the incentives for digital platforms to bias results in "recommendation systems" and on product market competition (Bourreau and Gaudin, 2022; Bourreau et al., 2021; Drugov and Jeon, 2024; Calvano et al., 2023).

Our paper also contributes to the empirical literature analyzing biases in algorithm-based platforms. Chiou (2017) examines the effects of Google's acquisition in 2011 of Google Flights (compares airlines fares) and Zagat (rates and reviews restaurants). She shows that after the acquisition of Google Flights, clicks in Google for the "travel" keyword declined for competing online fares comparators. By contrast, the integration of Zagat into Google increased the number of clicks to other sites, as Zagat provides information about the quality of restaurants, but also gives visibility to them. Hunold et al. (2017) investigate the default hotels' rankings offered by Booking.com and Expedia.com to their consumers, which differ from the rankings that would be obtained with the hotel prices or reviewer ratings. Aguiar et al. (2021) analyze potential biases in Spotify's rankings and the impact in the songs streaming success. They find that Spotify's New Music Friday rankings favor music from independent labels and female artists. Huang and Xie (2023) examine platforms that favor some sellers by offering consumers repetitive information about their products. Using data from food delivery platforms, they show that this behavior leads to higher average prices and more skewed revenue distributions among sellers. Finally, Chiou and Tucker (2024) show that changes in Facebook's algorithm demoting engagement bait decreased referral traffic from Facebook for recently established news relative to Twitter.

Our work is also related to the literature examining the impact of algorithmic recommendation systems on diversity and product discovery (Fleder and Hosanagar, 2009; Pathak et al., 2010; Brynjolfsson et al., 2011; Oestreicher-Singer and Sundararajan, 2012; Datta et al., 2018; and Aguiar and Waldfogel, 2020).[12] There is mixed evidence that recommendation systems promote long tail products and encourage sellers participation by making their products more accessible to niche consumers. Oestreicher-Singer and Sundararajan (2012) analyzed Amazon's co-purchase links, which suggest titles often bought together, and found that these links triple the influence of complementary books on each other's demand. Consistent with the long tail theory, niche books perform better, while popular books perform relatively worse in categories where recommendations are more significant. Hosanagar et al. (2014) use data from an online music service to examine whether recommender systems fragment users. They obtain that

[12] A related stream of literature analyzes the discriminatory effects of algorithms used in automate decision-making. Cowgill and Tucker (2019) survey studies on algorithmic bias and fairness, while Sweeney (2013) and Datta et al. (2015) focus on algorithm discrimination in advertising. Lambrecht and Tucker (2019) conduct a field experiment investigating bias in algorithms that deliver job opportunities ads.

recommendation systems make networks of users more homogeneous. Lee and Hosanagar (2019) analyze collaborative filtering recommender algorithms used by e-commerce firms. Using data from a randomized field experiment in an American top retailer, they demonstrate that collaborative filters reduce sales diversity. Absolute sales and views for niche items increase, but their gains are smaller than for popular items.

Finally, our paper complements the literature exploring how media firms influence public opinion by shaping the information landscape. A few papers have tried to identify the sources of media bias (Gentzkow and Shapiro, 2010; Duggan and Martinelli 2011; Oliveros and Vardy, 2015). Others have focused on the effects of media bias on the political process (Gentzkow and Shapiro, 2008, 2010, 2011; Duggan and Martinelli, 2011; Oliveros and Vardy, 2015; Piolatto and Schuett, 2015; Battaglini, 2017; Giovanniello, 2017; Buechel and Mechtenberg, 2019; Campbell et al., 2019; Pogorelskiy and Shum, 2019; Enikolopov et al., 2020). Our paper extends this literature by investigating whether algorithm-based platforms can influence media markets and, therefore, whether they hold media power (Prat, 2018; Allcott et al., 2024).

3. Google's search algorithm

Search engines such as Google, Bing, and Yahoo use bots to crawl web pages, moving from site to site, collecting information and indexing them. Googlebot, Google's web crawler, scans accessible web pages, analyzes their content, and indexes them. When consumers perform a query about a specific keyword, Google uses its algorithms to rank pages and display the results on its SERP. Organic search results are ranked based on their relevance to the query according to the algorithms criteria. The SERP also shows paid ads from firms aiming to attract consumers searching for specific keyword.

The success of news outlets in organic search depends on how well they rank relatively to their closest competitors, and more generally on how Google's algorithms balance and weigh "domain authority" and "content expertise" in queries that are relevant to them. The "domain authority" reflects features of the pages that are independent of the query, such as the page-speed, page-length, use of unique images, number of images, inclusion of original and updated contents, language, number of links targeting at the website, compression ratio of text, among others. The "authority" of websites determines their *static ranking* (Chandra, Suaib, and Beg, 2015; Baye, de los Santos and Wildenbeest, 2016). By contrast, the "domain expertise" shows how relevant are the contents of the news outlets for the query. This is influenced by various factors, including the query's keywords, the time and date of the search, the user's query history, the user's location, and the page's geographic origin, among others. The "expertise" of news outlets on a particular query determines their *dynamic ranking,* which can change over time. Google continuously evaluates the ranking of news outlets, considering factors such as click-through rates, publisher trustworthiness, story relevance based on the reader's location, and the freshness of the content. Interestingly, while news outlets with lower authority may struggle to attract traffic for highly competitive keywords, they can still rank well

for queries related to their region or niche market, where their expertise is more pronounced.

A useful document for understanding how Google ranks the web pages is its *Search Quality Rater Guidelines,* first published in 2015 to instruct human evaluators in assessing web pages.[13] The information raised by these raters is used to evaluate the performance of its algorithms.[14] Google explains that web pages are ranked according to the EEAT criteria, which assess *Experience, Expertise*, *Authoritativeness* and *Trustworthiness*. Pages are evaluated considering three dimensions: the quality of the website; the quality of the main content on the page; and the quality of the authors of the main content.

According to its guidelines, Google considers as high-quality webpages those with the primary objective of providing valuable assistance to users. Google pays special attention to "Your Money or Your Life" (YMYL) web content, which include pages (or topics) that could potentially impact a person's future happiness, health, financial stability, or safety. These could be, for example, websites that offer financial or medical advice. Google includes in this group news on important topics such as international events, business, politics, science, and technology. Despite this, not all news stories are considered YMYL. For example, sports, entertainment, and everyday lifestyle topics are generally not classified as YMYL. In its guidelines, Google instruct raters to assign low valuations to YMYL pages that present inaccurate, untruthful, or deceptive content.

Nowadays, SEO software firms like Moz, Majestic and Ahrefs offer tools to help websites increase their visibility in search engines and boost traffic. SEO is a crucial part of digital marketing, as search engines are an essential distribution channel for many businesses. Google does not share externally any indexing criteria that can help websites or advertisers to attract visitors. However, SEO companies use reverse engineering to identify the factors used by Google to index websites and have developed metrics to approximate websites' ranking or "domain authority".[15]

3.1 Google's Core and non-Core Updates

Google continuously updates its search algorithms and systems to improve the quality of search results. Most of these changes are minor in scale and scope and so they go unnoticed by users. However, a few times a year, Google implements large "core updates" that significantly modify the way it indexes and ranks websites. These updates can have

[13] Google employs around 10,000 people as "quality raters" worldwide. Rater data is not used directly by Google in its ranking algorithms, rather they use them as a mechanism to test if their systems work well. See Danny Sullivan in https://www.pi-datametrics.com/blog/google-e-a-t-ultimate-guide/.

[14] Google explained the relevance of these aspects in 2011, following the roll out of its "Panda update", which significantly changed its ranking algorithm A description of the new EEAT criteria can be found in the Google's guidelines for its reviewers. See https://developers.google.com/search/blog/2022/12/google-raters-guidelines-e-e-a-t and https://developers.google.com/search/blog/2011/05/more-guidance-on-building-high-quality.

[15] The concepts of "domain authority" and "domain trust" are based on the concept "PageRank", developed in the late 1990s under Google's search patents. PageRank assesses a website's authority on a topic and helps rank webpages after a query, reflecting both the number and quality of links to a page.

a significant impact on traffic. According to Google, these changes "*are designed to ensure that overall, we're delivering on our mission to present relevant and authoritative content to searchers.*"[16] In what follows, we describe how global core updates differ from minor (regular and routine) updates.

Global Core Updates. Core updates are rolled out globally, affecting all Google search regions and languages. They do not focus on specific types of search queries or on particular websites characteristics. These updates generate fluctuations in search rankings in the weeks following their adoption and can significantly alter the search traffic that websites receive. In recent years, Google has started to publicly announce its core algorithm updates to prevent speculation about their release. However, the company is reluctant to provide information about the changes introduced. Table 1A presents the nine core updates confirmed by Google during the period 2018-2020, which we analyze in this paper. Figure 1A shows Google's public announcements for each of these core updates, while Figure 1B presents Google's announcement on Twitter about its May 4th, 2020 core update and some of the immediate reactions from small newspapers.

Recovering website traffic after a core update can be extremely challenging. According to Google, there is nothing that specific website owners can do to immediately regain search ranking after these updates.[17] Google's algorithm is designed to deter spam and prevent sites from using unorthodox methods to increase their ranking. However, Google does offer advice and guidelines to webmasters on how to design and organize their pages to improve search results.[18]

It is often difficult to understand the causes driving Google's updates. In the past, it was sometimes possible to identify specific changes and determine which sites benefited or suffered from them. Nowadays, multiple hypotheses exist to explain the updates. One explanation is that Google is adjusting the level of EEAT (Experience, Expertise, Authoritativeness, Trustworthiness) required for a website to rank in the top SERP positions. Another explanation is that Google is using new technologies such as "Natural Language Processing" to understand whether the content is relevant and trustworthy to rank in high positions. As Google improves their ability to understand language, it gets better at identifying the contents that are truly helpful and authoritative. As a result, it may put less emphasis on PageRank when it comes to determining authoritativeness.[19] Changes in the consumers' search interest might also affect the rank of a website. For example, video content has become more important in the last years, and this has required machine learning to match queries with relevant content.

---

[16] See https://www.performics.com/2020/01/22/january-2020-google-core-algorithm-update

[17] See the announcement by Google Search Liaison (@searchliaison) on Oct. 11, 2018 "*Sometimes, we make broad changes to our core algorithm. We inform about those because the actionable advice is that there is nothing in particular to "fix," and we don't want content owners to mistakenly try to change things that aren't issues…"*. See also https://blog.searchmetrics.com/us/google-update-november-2019/.

[18] Google Webmaster Blog shows different actions that can be adopted after being affected by Core Updates (https://developers.google.com/search/blog/2019/08/core-updates). It also publishes a "Webmaster Guidelines", showing how it index and rank web site, and the treatment of elicit content. See https://developers.google.com/search/docs/advanced/guidelines/webmaster-guidelines.

[19] See https://www.mariehaynes.com/eat-and-semantic-seo/

Regular updates of search algorithms. In addition to the core updates, Google regularly launches other specific algorithm changes that, while smaller in scale, can still affect many sites. An example of such type of updates was the introduction of the Mobile-First Index in November 2016, which shifted Google's focus from desktop to mobile versions of web content for indexing and ranking. The rollout of this update started in 2018 and was completed in 2021. Considering this gradual migration, it is difficult to assess which has been its impact on news outlets search traffic.

Another example was the global rollout of the BERT algorithm (Bidirectional Encoder Representations from Transformers) on December 9, 2019. This is a neural network-based technique for natural language processing (NLP) that helps Google better discern the context of words with queries. It can improve the accuracy of the results in 10% of searches across 70 languages.

In addition to these changes, Google makes daily minor changes to its algorithms to improve the search results and to correct different types of bugs. For example, several changes have corrected indexing and canonical bugs (Google confirmed these adjustments on August 10, September 29 and October 12, 2020). Canonical tags are crucial when content is syndicated, as they indicate the original page, ensuring that the primary source is prioritized in search results when its content is reproduced elsewhere. Google also launched "passage indexing" on February 10, 2021, to enable its algorithm to index specific sections of a page, potentially affecting 7% of search queries across all languages.

3.2. A Story of Two Tales

Our goal in this paper is to understand the impact of Google's global core updates on search traffic and market concentration. Yet, Google's opacity about the reasons behind its core updates is an obstacle for our and anybody's research aiming to understand the impact and efficacy of search algorithms. Fortunately, Google disclosed the objectives behind two specific updates, **the August 2018 core "medic" update** and **the May 2020 core "pandemic" update**. Next, we briefly describe these two particular updates and their likely impact on the media market. Appendix A provides extensive detail about their rollout.

The August 2018 "Medic" update. This update was released in August 1, 2018, and its rollout was gradual taking several weeks. While it affected a wide range of websites, it focused on medical sites and thus its "medic" nickname. The update aimed to increase the protection of users from harmful information by modifying the algorithms that assess the relevance and authority of web page content. In fact, the main goal of the update was to better understand and identify "Authority" and "Expertise" of website content.[20] According to industry experts, the "medic" update increased the ranking of websites offering accurate and reliable information, with a strong reputation and a high level of

[20] In its Search Quality Rating Guidelines of March 2013, Google described its search quality rankings were based on the concepts of Expertise, Authoritativeness and Trustworthiness (EAT).

trust among users. After the update, websites with high-quality, well-written content that was relevant to the user search queries were more likely to rank higher in the search results. While the update affected the ranking of sites that compete for health, medical, legal, and finance keywords, it also affected newspapers, public information sites and pages dedicated to ecommerce.

The May 2020 "Pandemic" update. From May 4 to May 18, 2020, Google Search released its first core algorithm update after the start of the coronavirus pandemic. Industry experts consider this update as one of the largest to date. The aim of the update was to adjust the algorithm to better serve the new user behavior patterns in this period. After March 2020, digital newspapers were registering unprecedented levels of visits due to the large demand for news about COVID-19. In addition, there was a huge increase in the interest about local news. As a response, Google adjusted its SERP by introducing a COVID-19 left panel with information about the pandemic and including a local news box. People searching for coronavirus information could see both a "top stories carousel" as well as a "local news carousel" with coronavirus information (see Appendix Figure A1). According to Google, one of the objectives of the May update was to ensure that local news was not buried down below larger national newspapers covering the same topics and to increase the visibility of publishers with high authority on local matters.

In summary, while the "medic" update emphasized Expertise and Authority, the "pandemic" update highlighted the relevance of Expertise over Authority. Accordingly, we would expect the "medic" update to increase search visits in general news outlets relative to regional and niche news outlets, and the opposite result for the "pandemic" update. More generally, when we evaluate the joint impact of the nine core updates released between 2018 and 2020, we can interpret a relative increase in search visits to national outlets as an indication of Google's growing preference for Authority, while a relative increase in visits to regional and specialized outlets suggest a higher preference for Expertise. We take these predictions to data in the next sections.

4. The Data and Empirical Strategy

4.1. The Data

Our analysis uses information at the domain-day level from SimilarWeb, a web measurement company providing traffic data and user-engagement statistics. This firm collects data on browsing behavior from rich and diversified panels of consumers in several countries. The information collected for our study covers the time period from October 1, 2017, to December 31, 2020, including the nine Google core updates occurred during that period.

To examine the effect of Google core updates on news outlets search traffic, we consider 606 news outlets from 15 different European countries. To reach our final sample, we selected a first list of relevant and sizable online news outlets using national rankings from Alexa (www.alexa.com) and SimilarWeb (www.similarweb.com) as well as other websites (e.g., Wikipedia) and publicly available news outlets lists for each country. It is

important to note that these sources also included webpages from TV and radio stations offering news contents. Finally, we were constrained by the fact that SimilarWeb does not provide daily information for domains with traffic levels below 5000 daily visits. Thus, we dropped from our initial list those smaller outlets failing to meet this requirement.[21] The final data set includes the daily visits from desktop and mobile devices, except for Denmark, for which daily mobile data is not available. Mobile data for Belgium, Finland, Ireland, Netherlands, and Sweden starts on January 1, 2018. Overall, we aimed to have a well-balanced sample of news outlets.

The main variable of interest in our analysis is the domain's *Daily Desktop Search Visits*. This variable is defined as the daily organic visits to a news outlet originated in a search engine. In our dataset, more than 95% of the organic search traffic is originated in Google Search. Data on mobile organic search is only available at the monthly level. We analyze two additional outcome variables, the *Daily Total Desktop Visits* and the *Daily Total Mobile Visits*, which reflect the total visits that news outlets obtain from these two distribution channels, respectively. Thus, *Daily Total Desktop Visits* are the sum of *Daily Desktop Search Visits* (our main variable of interest) and other traffic sources such as direct visits, visits from social media, paid search, and display ads. The same applies for *Daily Total Mobile Visits*.

Our data set also details other traffic sources for newspapers. *Daily Desktop Direct Visits* is the daily traffic to a news outlet from a different web domain or from the beginning of an empty browsing session. In other words, these visits are from users who know the outlets and intentionally type their names to access them. We use this variable to control for daily changes in the visits of news outlets that are related to the content they publish or country-specific events driving the overall number of visits. Figure 2 shows the evolution of daily desktop and mobile visits between January 2018 and October 2020. The red vertical lines in the figure show the dates of Google's algorithm core updates. During our sample period, mobile traffic grew at a higher rate than desktop traffic. The figure also shows that desktop and mobile visits dramatically increased after the WHO declared the coronavirus a global pandemic on March 11, 2020. Figure 3 presents the evolution of the shares of direct, organic search and social media traffic. Our data set also provides information on the *Daily Desktop Paid Visits* and the *Daily Desktop Ads Visits*, which are the traffic originated from paid search ads on search engines and the traffic from display and video ads via an ad-serving platform (i.e., GDN, Doubleclick), respectively. [22] We use these two variables in section 5.3 to examine whether news outlets react to Google's core updates by increasing their advertising effort in Google Search and other websites.

[21] The final list of outlets in our sample is distributed across 15 European countries as follows (number of outlets per country in parenthesis): Austria (35); Belgium (24); Denmark (25); Finland (32); France (43); Germany (49); Greece (50); Ireland (34); Italy (54); Netherlands (42); Poland (52); Portugal (27); Spain (65); Sweden (37); and UK (37). Table 1B presents the complete list of the domains.

[22] Our main analysis focusses on organic search, and therefore it abstracts from news sites competition to attract sponsored keywords (Choi et al. 2020; Simonov et al, 2018; Decarolis and Rovigatti, 2021).

When exploring outlet heterogeneity, we classify news outlets in our sample according to different criteria. First, we classify outlets according to their specialization (*National, Regional, Business, Sports* or *TV/Radio*) following verbal descriptions from several sources such as Alexa, SimilarWeb and Wikipedia. This classification is informative of the outlet scope and degree of specialization. Second, we divide outlets according to their national rank. Specifically, we classify a news site as *Top Rank* (*Bottom Rank*) if their national rank is above (below) the median in their own respective country in our sample. This way of grouping outlets is based on the total number of visits they receive relative to the visits to other outlets of the same country in our sample. Third, we classify outlets as *Top Domestic* (*Bottom Domestic*) if the share of domestic visits is above (below) the median in their own respective country in our sample. This distinction is meaningful in that news outlets receiving a larger share of visits from abroad are likely to have more authority and to be less sensitive to changes in the search algorithm. Fourth, outlets are classified as *Top Google* (*Bottom Google*) when the percentage of the total "search visits" originated in Google Search is above (below) the median in their country in our sample. This last classification is informative in that it separates outlets more dependent on Google's search traffic from those less dependent.

SimilarWeb also collects information on the *monthly traffic* generated by each keyword used in Google Search's queries for each outlet in our sample. Using this information for our sample of Spanish outlets, we use this variable in section 7.1 to estimate the impact of Google updates on the concentration of search traffic at the keyword level. These data show the number of months each keyword has been used in consumer queries, and how often news outlets have been ranked in top 10 organic search results by Google for each keyword.

Our study also uses data on website performance from Ahrefs and Majestic, two renowned SEO service firms. As explained above, Google has modified its algorithm repeatedly over the last few years to reflect changes in its EEAT criteria. As a result, SEO companies have developed their own software to monitor websites' SEO health over time. We have collected daily information of two variables that can be seen as proxies of the EEAT criteria. The first variable is *Ahrefs' Domain Rating,* which measures the strength of a website's backlink profile compared to those of other websites. Specifically, it considers the quality and quantity of domains linking to an entire website. This variable is measured on a logarithmic scale from 0 to 100, with the latter being the strongest.[23] According to Ahrefs, this metric works in a similar way to the original PageRank calculation, although it ranks websites and not web pages.[24] Similarly, *Majestic Trust Flow* is a score between 0 to 100 that indicates the perceived trust of a website based on the quality of backlinks that a website receives. The higher the score, the higher the level of perceived trust of a website.

[23] See https://ahrefs.com/blog/seo-metrics/#section7 and https://ahrefs.com/blog/domain-rating/.

[24] Google stopped using PageRank (PR) to assess websites in 2013, although it still uses this metric internally in its web positioning algorithm. One problem of the PR was that it was relatively easy to increase it by buying sponsored articles, commenting on blogs, or getting links on high PR sites. See https://ahrefs.com/blog/google-pagerank/.

Finally, an important variable for our analysis is *Words Top 10*, which is a metric developed by Ahrefs that shows the number of keywords (queries) for which news outlets occupy one of the top 10 positions in Google's organic search results .[25] Specifically, this variable shows the frequency a news outlet ranks in the top 10 search results for any of the 605 million keywords Ahrefs has in its database. This is a country-specific variable. Note that the number of organic keywords news outlets have in top positions can change over time, not only because they can rank higher or lower in search queries, but also because the Ahrefs' database is growing. To simplify the presentation of our results, most of our analysis will focus on the variable *Words Top 10.* However, appendix Tables A1 and A2 use the variables *Words Top 100* and *Words Top 11-100,* which account for the number of keywords (queries) that news outlets have in the top 100 and the top 11-100 organic search results in Google, respectively. According to industry reports, the results in the first page of Google's SERP capture around 71% of search traffic clicks, and the results in the second page capture less than 5.5% of the clicks.[26] Thus, the number of keywords in the top 10 search results is a crucial factor for news outlets to obtain search traffic, even when they might have thousands of keywords in top 11-100 positions. Also, note that, after a first query, users can redefine their search keywords and phrases to obtain more accurate results. Figure 4 shows an example of the 10 first search results for "US Election 2021," which are in the first page of the SERP. The first link for a news outlet is for CNBC, in the sixth position. Previous results are for Wikipedia and for several institutional sites. In this example, the first page of Google's SERP includes "zero-click searches", which are answers to queries that do not include a link to third-party websites. Google uses its *Direct Answer Box* to offer answers to many consumers' queries, such as queries for celebrities, geography, history, the weather or stock market prices. It is estimated that around 50 percent of searches currently end without a click on an organic search result (zero-click searches). Table 2 shows summary statistics for all the variables used in our analysis.

4.2. Empirical Strategy

Our empirical model examines how Google Search affects the visits received by European news outlets. As previously discussed, Google's algorithms index news outlets and determines their ranking in search results pages when users perform queries. The higher a news outlet ranks, the greater the probability that users will click-through on its links and generate visits. This means that we should observe an empirical relationship between the search visits to news outlets and the number of keywords they have in top 10 search results. Therefore, our baseline specification is as follows,

$$\ln[visits_{it}] = \alpha_i + \beta \ln[Words\ Top\ 10_{it}] + \gamma X_{it} + \delta_t + u_{it}, \qquad (1)$$

[25] See https://ahrefs.com/blog/seo-metrics/#section6.
[26] See https://moz.com/blog/google-organic-click-through-rates-in-2014.

where $\ln[visits_{it}]$ is the natural logarithm of the number of visits (desktop organic search visits, total desktop visits and total mobile visits), to news outlet *i* in day *t*, and $\ln[Words\ Top\ 10_{it}]$ is the natural logarithm of the number of keywords that the news outlet *i* has in the top 10 organic search results in day *t*. At this point, it is important to note, from summary statistics detailed in Table 2, that the number of keywords in the top 10 organic search results are bounded away from zero with an average of 88148. Moreover, X_{it} is a set of variables varying across news outlets and days such as the number of direct visits, and α_i and δ_t are news outlet and year, week, and day-of-week fixed effects, respectively. The usual iid assumptions apply to the error term u_{it}.

To account for unobserved heterogeneity at the news outlet level and the potential incidence of non-stationarity in the data, we use first differences of equation (1) such that,

$$\Delta\ln[visits_{it}] = \alpha + \beta\,\Delta\ln[Words\ Top\ 10\ _{it}] + \gamma\Delta X_{it} + \Delta\delta_t + \Delta u_{it}, \qquad (2)$$

where we difference out the term α_i and we take care of potential autocorrelation in the error term. All other variables are the result of differences between the contemporaneous variable with realizations of the variable four days before such that $\Delta\ln[y_{it}] = \ln[y_{it}] - \ln[y_{it-4}]$. While we chose to implement our empirical strategy of 4-day gaps to capture immediate results after Google Search core updates were announced, our results are qualitatively and statistically equivalent when using 7-day differences. Thus, our identification of the coefficient of interest β rests on the assumption that $\mathrm{cov}(\Delta\ln[Words\ Top_{it}], \Delta u_{it}) = 0$.

Regardless of the use of first differences and the exogeneity assumption, it may still be the case that news outlets invest more heavily in keywords that can generate more visits when there are contemporaneous events (unobserved by the econometrician) that can attract the attention of consumers. News outlets can gather data on keyword volume and trends as well as information on keywords targeted by competitors, with the goal of using combinations of keywords and phrases that generate more visits. This behavior would potentially increase their audience, online traffic, and consequently their ad revenues.[27]

To address this endogeneity problem, we cannot use a traditional diff-in-diff approach because updates are global and, consequently, there is no control group unaffected by Google's core updates. Instead, we pursue a quasi-experimental two-stage identification strategy. For this, we need a variable (the "instrument") that is correlated with the number of keywords a news outlet has in Google's top search position but that does not directly influence the outlet's search visits or the probability that readers will search for its content, other than indirectly through the number of keywords in top positions. Our

[27] Baye et al. (2016) show that SEO efforts, such as optimizing anchor texts, headings, meta tags, robot.txt files, and page titles, improve a site's relevance for product searches. Hagar and Diakopoulus (2019) describe the use of A/B headline testing to increase traffic to stories. Leung and Strumpf (2023) show that the New York Times and the Wall Street Journal adjust headlines and their effect on performance metrics.

"instrument" for the number of keywords in top SERP positions is the occurrence of Google's core updates, which directly modify the news outlets' indexation for each consumer's query, and thus, are a source of exogenous variation for the news outlets' visits. We estimate our first-stage model such that,

$$\Delta \ln[Words\ Top\ 10_{it}] = \theta_0 + \theta_1 CoreUpdatePlus7_t + \theta_2 \Delta X_{it} + \Delta\omega_{it}. \quad (3)$$

The instrument *Core Update Plus* 7 is a dummy variable that takes value 1 if day *t* is the day in which Google rolled up a core update or it is within seven days after that, and 0 otherwise. Thus, in the event-study regression (3), the treatment period is the seven-day period following any of the nine core updates during the three-year period in our sample. The implicit assumption behind this approach is that the impact on the number of keywords will be more pronounced during the seven-day period when the instrument dummy takes value 1. On all other days in our sample, outside these nine seven-day periods, the dummy takes a value of 0. As mentioned above, it is important to emphasize that we are examining global core updates, which impact all outlets regardless their individual characteristics and thus precluding a traditional diff-in-diff approach. Note also that our first stage results are robust and qualitatively equivalent when using 4-day and 14-day windows for the Core Update dummy variable.

Our analysis considers the nine core algorithm updates launched by Google between October 2017 and December 2020. This dummy variable is an instrument for the independent variable $\Delta \ln[Words\ Top\ 10_{it}]$ under the assumption that $\text{cov}(Core_update_plus7, \Delta u_{it}) = 0$. This implies that Google core updates are orthogonal to changes in visits (search or total) to any news site *i* in our sample. Indeed, we assume that no site anticipates the announcement of a global update and that these updates are not triggered by any increase or decrease in the number of keywords for which news outlets *i* appears in the top ten search results. That is, we assume Google rolls out its global core updates without considering traffic changes to the specific websites we are analyzing.

It is important to acknowledge that some outlets might have SEO teams or hire external consultants to help them react to algorithm updates. However, these SEO services cannot predict the exact date and the specific nature of a core update. While they may forecast changes in future updates,[28] adjusting to new algorithms takes time, typically ranging between two weeks and six months.[29] In most cases, "accurately forecasting SEO performance is almost impossible due to the ever-changing search engine algorithms, intense competition, and unpredictable user behavior."[30] In our analysis below, we show

[28] See https://capforge.com/seo-forecasting-why-its-important/.
[29] See for example https://twitter.com/jeremiahcsmith/status/1660359928106860544 and https://www.forbes.com/sites/joshsteimle/2015/02/07/how-long-does-seo-take-to-start-working/?sh=17a7daad464c.
[30] See https://www.linkedin.com/pulse/can-you-guarantee-1-rankings-truth-forecasting-seo-results-1e/.

the heterogenous results of core algorithm updates across different types of outlets, which may reflect variations in the expertise of their SEO services.

5. Results

5.1. Our Story of Two Tales

We begin the results section by separately examining the impact of the updates described in section 3.2: the August 2018 "medic" update and the May 2020 "pandemic" update. The objective of this analysis is to validate our earlier predictions by providing correlations between these updates, the number of top keywords and search visits for all outlets in the sample. While column 1 in Table 3A shows the 2018 "medic" update was associated to an increase in the number of top keywords, column 2 shows that the 2020 "pandemic" update was associated with a decline in the number of top keywords. These results are consistent with our predictions in section 3.2. The "medic" update modified the indexation of websites by giving more weight to their domain authority, which benefitted the ranking of the average newspapers in our sample. By contrast, the "pandemic" update emphasized on the expertise of news outlets to respond to the specific queries of users, which harmed the ranking of the average newspaper in our sample.

Table 3B explores how search visits changed with each update. Columns 1 and 2 show that search visits increased after the "medic" update, consistently with our findings in column 1 of Table 3A. Also, columns 3 and 4 in Table 3B show that search visits went down after the 2020 "pandemic" update, which is consistent with the results in column 2 of Table 3A. These results suggest that Google ranked non-news outlets (i.e. local health organizations, local governments, and large commercial online platforms) higher than news outlets for user queries relating COVID-19 or products and services shortages.

Finally, columns 3 and 4 of Table 3B examine heterogeneity on the impact of these two updates across different classifications of news outlets (national, regional, business, and sports). Column 3 shows no difference across news outlet types after the "medic" update. Thus, the relative authority of the different types of outlets in our sample did not change after the update. By contrast, column 4 shows that after the "pandemic" update national news outlets experience a larger decrease in search visits than regional outlets, and regional outlets experience a larger decrease in visits than our reference group (sports and business outlets). This is consistent with specialized outlets having larger expertise than national and regional outlets, and regional outlets having larger expertise in local queries than national outlets.

Our empirical analysis in the next section bundles all updates and examines the average effect of core updates in our sample period. Our findings in this section are helpful for interpreting the results in the sections below. When the number of top keywords increase as a result of a core update, this is because the updates, on average, favor the authority of news outlets (relative to other content domains), which likely increases the number of search visits. Conversely, if the number of top keywords in the outlets decreases, this

implies that the update favors the expertise of content domains, leading to fewer visits to national outlets compared to regional and specialized outlets.

5.2. Main Results

This section analyses the overall effects of the nine algorithm core updates confirmed by Google in the period from October 1st, 2017, to December 31st, 2020. As detailed in Table 1A, the aim of these updates was to fix various aspects of the indexing algorithms and to introduce new features that improve search accuracy. Each update is unique in its purpose and impact on online traffic. However, to simplify the exposition of our results we categorize them into two groups: the three *biggest* core updates (August 1, 2018, June 3, 2019, and May 4, 2020), as identified by SEO specialists, and the remaining six *not big* core updates.

Columns 1, 2 and 3 of Table 4 show results of running specification (2). Each column has a different outcome variable, namely, *Desktop Search Visits*; *Desktop Total Visits*; and *Mobile Total Visits*. Our independent variable *Words Top 10* reflects the number of keywords that news sites place in the top 10 search results in Google. All regressions include as a control the variable *Desktop Direct visits*, as well as day of the week, week, and year fixed effects. Standard errors are clustered at the news outlet level to allow for correlations across observations of the same outlets. All three specifications show a positive and significant correlation between traffic and the number of top 10 keywords. Specifically, the results indicate that a 1% increase in the number of top 10 keywords is associated with a 0.09% increase in the number of search visits, 0.05% increase in desktop visits, and a 0.13% increase in mobile visits.

As explained in section 4, a potential limitation of this analysis is that news outlets may strategically use keywords to increase their visits. For example, they can repeatedly use specific keywords in headlines and news stories to rank higher in frequent query results. To deal with this endogeneity problem, we pursue a two-stage identification strategy, using Google's algorithm updates as an "instrument" as shown in equation (3). This strategy leverages plausible quasi-exogenous variation in the number of keywords ranking in the top 10 search results during the core updates.

Using this identification strategy, column 4 in Table 4 shows the effect of the updates in the number of top 10 keywords that news outlets have in the top search results. We find an overall negative and significant effect of Google's updates on the number of keywords in the top 10 positions. *Big* core updates decrease the number of top 10 keywords by 0.001% and *Not Big* core updates by 0.003%. The last three columns in Table 4 show the results for the two-stage instrumental variable estimation of the linear model in equation (3), for the three outcome variables of interest, and using the variables *Big* and *Not Big Core Update Plus 7* as instruments for the variable *Words Top* 10.[31] The results of the

[31] The F test of excluded instruments is F(2, 579) = 10.76. We are able to reject the null hypothesis of model under-identification with a Kleibergen-Paap rk LM statistic of Chi-sq(2)=20.80. We are also able to reject the null hypothesis of weak instruments with a Cragg-Donald Wald F statistic = 35.72 and

second stage of our estimation show that the number of top 10 keywords has a positive and significant impact in the three outcome variables. Specifically, we obtain that a 1% increase in the number of top keywords generates a 6.6% increase in the number of search visits, and a 3.7% increase in the total number of desktop visits, and an increase of 4.1% in mobile visits. It is important to note that these coefficients are significantly larger than the OLS coefficients reported in columns 1 to 3. A potential explanation would be the heterogeneity in the impact of the core updates as exogenous shifters of the number of top keywords. Our analysis about the heterogenous effects of the updates across different types of outlets will document such heterogeneity.[32] [33]

In summary, two main conclusions follow from this section. First, the Google's algorithm updates rolled-out in the 2018-2020 period had a significant effect in the number of keywords that news outlets have in top search results. These updates affected news outlets in different directions and magnitudes, but they had an overall negative effect in the number of keywords that news outlets have in top 10 search results. This result suggests that the updates favored the "expertise" of content domains. Second, the number of keywords that news outlets have in the top 10 search results pages had a positive effect in the visits that receive news outlets. This confirms the relevance of Google's algorithm updates for the media market.

5.3. Heterogeneous Impact of Google Core Updates

Next, we investigate the heterogeneous effects of the nine core updates across national markets and different types of outlets. Table 5 repeats the estimation presented in Table 4 for each of the 15 countries in our dataset. For each country, we run first-stage regressions of first differences in log of the variable *Word Top 10* on *Big* and *Not Big* core updates dummies. Then for each country, we run the second stage estimation using the core updates as instruments for changes in the number of desktop *Search Visits*, total *Desktop Visits,* and total *Mobile Visits*. The first and second columns in Table 5 show the results of the first stage estimation for each country. Although results vary across national markets, in most countries we find evidence that *Big* and *Not Big* core updates had a negative effect in the number of keywords that news outlets had in the top 10 search

Kleibergen-Paap Wald rk F statistic = 10.76. We also apply the Montiel-Pflueger robust weak instrument test and obtain an effective F-statistic of 31.034, which allows to reject the null of weak instruments at a confidence level of 5%.

[32] Appendix Table A1 complements our analysis by examining the updates' impact on the number of top 100 and top 11-100 keywords, as well as each individual update's effect on top 100, top 10 and top 11-100 keywords. Columns 5, 6 and 7 show that each individual update has its own idiosyncratic effect (positive or negative) on keywords in top search positions. This is consistent with section 5.1, suggesting that different updates target expertise or authority. We find that updates had varying effects on the number of keywords ranked in top positions for each news outlet, with similar effects for top 10 and top 11-100 positions. Finally, Table A1 presents the effects of Google's non-core updates, showing an overall negative and significant effect on the number of keywords for which outlets ranked in top search positions.

[33] Appendix Table A2 shows the second stage estimation results of the model in equation (3), using the variables *Big* and *Not Big Core Update Plus 7* as instruments for *Words top* 100 and *Words top* 11-100 from Table A1. The results show that top 100 and top 11-100 keywords have a positive impact in the *Desktop Search Visits*. Results for *Desktop Total Visits* and *Mobile Total Visits* are not significant.

results. The three columns in the right present the results of the second stage regressions for the three outcome variables. Results for search visits are ambiguous. We find a positive relationship between *Word Top 10* and the number of visits in Demark, Poland and Spain, and a negative relationship in Greece and the UK.

Tables 6 and 7 repeat the previous analysis with different sample splits. In Table 6, news outlets are classified according to their national rank, the percentage of domestic traffic, and the percentage of their search traffic originated in Google Search. In these classifications, we divide news outlets in two groups, those above and those below the median of the variable in their respective countries. The results of the first-stage regressions show a negative relationship between the *Big* and *Not Big* core updates and the variable *Word Top 10*. The only exception is for the variable *Top Google*, which implies that the group of news outlets that receive a larger share of their search traffic from Google were not affected by the updates. These findings suggest that news outlets rank very high in terms of authority are less sensitive to core updates than those with lower authority regardless of their national rank or the percentage of foreign visits. Results for the second-stage regressions confirm that the number of keywords in top 10 search results have a positive effect in the number of *Desktop Search Visits*, and in the number of *Total Desktop Visits* and *Total Mobile Visits*.

Table 7 classifies news outlets according to their specialization, which can be *National, Regional, Business, Sports, or TV/Radio*. As above, the results of the first-stage regressions show a negative relationship between *Big* and *Not Big* core updates and number of *Keywords Top 10*, although in the case of *Big* core updates the coefficient is negative and significant only for *National* and *Regional* outlets. The findings of this table appear to indicate that outlets with specialized expertise (business, sports, TV/radio) are less sensitive to big updates. Finally, the estimates for the second-stage regression exhibits a positive relationship between the number of keywords in top 10 search results and the number of *Search Visits*, except for the case of *Sports* outlets for which the coefficient is not significant (the coefficient is significant and negative in the case of total *Mobile Traffic*).

5.4. Do News Outlets React to the Updates?

Once established that Google´s core updates, on average, decrease the number of search visits received by news outlets, we examine whether news outlets reacted to these changes in the short run. Albeit not directly observable to us as econometricians, news outlets take actions, such as hiring SEO service firms, to game the algorithm and increase their visibility.[34] Thus, it is likely that core updates trigger meetings and actions at the news outlet level that aim to understand why their traffic has been affected and to decide how best to react. To the best of our knowledge, there is not a common understanding of how

[34] Outlets may hire SEO consultants to engage in SEO forecasting (https://www.linkedin.com/pulse/can-you-guarantee-1-rankings-truth-forecasting-seo-results-1e/).

fast SEO may help outlets reestablish traffic with estimates ranging between two weeks and six months.

While we cannot observe SEO investments and meetings following core updates, we are able to analyze the traffic patterns of news outlets to detect if they reacted in specific ways after the updates. We do so in two different ways. First, Table 8 checks whether the number of top keywords increased two weeks after the release of the core updates. If short-term actions were taking place, one week after the release of the updates news outlets would increase their efforts to have more keywords in top 10 positions. Column 1 in Table 8 shows that news outlets did not react in this way, as the number of top keywords was still lower 8 to 14 days after the updates. Columns 2 to 4 show elasticities of number of keywords with respect to search, desktop, and mobile visits range between 2.5, 1.4 and 3.7, respectively.

Second, we check whether the decrease in the number of top keywords correlates with changes in the *Daily Desktop Paid Visits* and the *Daily Desktop Ads Visits* received by the outlets. The rationale behind this exercise is that a decrease in search visits may drive news outlets to increase their investment in other channels such as paid search or display ads. Table 9 shows that the variation in top keywords triggered by core updates has no impact on paid search or display ad visits. The first-stage results of these specifications are the same as in column 4 of Table 4. In a nutshell, we find no evidence that news outlets use these strategies to compensate the decrease of visits triggered by the updates in the short run.

Finally, we have emphasized in section 3 that Google's algorithm may direct more or less traffic to news outlets depending on their experience, expertise, authoritativeness and trustworthiness (EEAT criteria). Thus, we test whether the impact of the updates differs according to the domain authority of the news outlets. To do so, we use the variables *Ahrefs Domain Rating* and *Majestic Trust Flow*, which are described in section 4.1. Columns 1 and 2 in Table 10 show that outlets with a higher ranking and trust flow, respectively, are less likely to change the number of keywords in top 10 search results. However, the impact of core updates does not depend on the ranking or the level of trust flow of an outlet. Conversely, columns 3 and 4 show that core updates do not influence the *Ahrefs Domain Rating and Majestic Trust Flow*. Finally, we examine heterogeneity on the impact of core updates on *Ahrefs Domain Rating and Majestic Trust Flow* in columns 5 and 6. If anything, we find that core updates may decrease *Majestic Trust Flow* for non-national non-regional outlets but they do not impact *Ahrefs Domain Ranking* across different types of outlets.

6. Market Concentration

In the previous section we show that search engines such as Google can modify the search traffic of media outlets when they adjust their indexation algorithms through a decrease in the number of keywords that news outlets have in top positions in organic search results. In this section we investigate whether core updates have also implications for the

concentration at the market level. Although this is ultimately an empirical question, it is important to understand whether the updates have increased the visibility of small and niche news outlets, or whether they have favored the consolidation of "superstar" news outlets. This analysis is relevant to better understand the role of information technologies in product variety and market concentration (Anderson, 2004; Fleder and Hosanagar, 2009; Brynjolfsson, Hu, and Simester, 2011).

To do so, we examine whether core updates, through a reduction in top keywords at the site level, increase (reduce) market concentration. This finding would imply that the core updates implemented by Google Search have contributed to the superstar (long tail) effect as documented in the literature. We estimate the following model:

$$\Delta \ln[HHI_{ct}] = \varphi_0 + \varphi_1 CoreUpdatePlus7_{ct} + \varphi_2 \Delta X_{ct} + \Delta \varepsilon_{ct}, \qquad (4)$$

where HHI_{ct} is the Herfindahl–Hirschman market concentration index for country *c* in day *t*. We calculate this variable for each of the three outcome variables examined in our study using the market share of news outlets in their corresponding national markets. We run first differences regressions of the changes in the log of HHI for search, desktop, and mobile visits on *Big* and *Not Big* core update dummies. All specifications include month, year, day of the week fixed effects and changes in the number of direct visits as controls.

Figure 5 shows the evolution in the HHI of the three dependent variables in the period we examine. Interestingly, we observe that the variable search visits is less concentrated than total desktop visits and total mobile visits, although differences are decreasing over time. Moreover, the concentration of the search market increases importantly in periods in which there is a peak in news consumption (e.g., international football competitions, covid pandemic).

Tables 11 and 12 show the results of the estimation of equation (4) for different subsamples of outlets. Table 11 shows the aggregated effects of *Big* and *Not Big* core updates for the whole sample of news outlets and for each individual country, respectively. Focusing on the concentration of search visits, columns 1 shows that the overall result of the three *Big* core updates was a 1% reduction of market concentration. However, column 2 shows that this effect was mostly compensated by a 0.08% increase of market concentration due to the effect of the *Not Big* core updates. As shown in Table 2 above, the average and standard deviation of HHI in the search market in our sample is 800. Our findings would imply a minor 8-point difference above and below average levels and equivalent to a 1% standard deviation differences. It is important to note that such small differences in market concentration are unlikely to trigger action from regulatory agencies on their own.[35]

[35] As a reference for the magnitude of these effects, note that the Horizontal Merger Guidelines of the US Department of Justice and the Federal Trade Commission considers that mergers resulting in unconcentrated markets (HHI below 1500) are unlikely to have adverse competitive effects and ordinarily

If we now consider the effects of core updates at the individual country level, we find that results are quite heterogeneous. *Big* core updates had a negative effect on market concentration in Finland, Germany, and Greece, but a positive effect in Portugal. *Not Big* core updates had a positive effect in Finland and Netherlands. These results suggest that Google's algorithm core updates can have relevant consequences in terms on market concentration, but their effects are by no means homogeneous across European media markets.[36]

We complement this analysis investigating whether Google's updates have modified the salience of national news outlets relative to other outlets. As described in the introduction, we are interested in knowing whether algorithms updates create a "superstar" effect that reinforce the position of national outlets with a high domain authority, or whether they favor the "expertise" of local and niche outlets, and therefor strengthen the "long tail" of the media market. With this goal in mind, Table 12 shows the impact of *Big* and *Not Big* updates on the search and overall traffic received by national news outlets relative to all other outlets and relative to regional outlets in their respective country markets.

Our findings in Table 12 confirm that the *Big* core updates have, on average, reduced the share of search visits to national outlets, while relatively increasing the visibility of regional outlets. This suggest a slight strengthening of the "long tail" in the media market. It is important to emphasize that the media market is characterized by a vast number of outlets and, thus, low market concentration (Figure 5). Beyond this status quo, ensuring continued public access to local information remains a critical policy objective.

7. Additional Findings Using Monthly-Aggregated Data

Up to this point, we have investigated the impact of Google algorithm's updates using daily data. In this section, we exploit rich keyword-level data and search-engine-specific traffic data only available at the monthly level. We restrict our analysis to the Spanish media market due to the large number of Spanish newspapers in our sample (65 outlets) and to our (the authors) home-biased expertise in Spanish news media.

7.1. Concentration at the Keyword Level

Given that core updates affect the number of keywords that outlets have in top positions in search results, it seems adequate to examine whether they have affected traffic concentration at the keyword level. Note that each keyword generates visits to different

require no further analysis. However, we find that the individual effects of core updates in some national markets can be substantial. See https://www.justice.gov/atr/horizontal-merger-guidelines-08192010.

[36] Table A3 analyzes the effect of Google's core updates by type of news outlet. The results reveal that the three *Big* updates did not generate any significant aggregated effect in the concentration of *National* outlets markets. In contrast, *Not Big* updates increased market concentration of search visits for *National* news outlets, and they reduced the concentration for *Sports* outlets.

news outlets. Using *monthly data* on search visit traffic generated from Google to the news outlets at the keyword level, we analyze whether core updates increase concentration at the keyword level. Similarly to equation (4) above, our regression specification is such that,

$$\ln[\mathrm{HHI_{it}}] = \varphi_0 + \varphi_1 \mathrm{CoreUpdate_t} + \varphi_2 \mathrm{CoreUpdatePlus1_t} + \varphi_3 \mathrm{CoreUpdatePlus2_t} + \gamma \mathrm{Age_{it}} + \theta_i + \mu_t + \varepsilon_{it}, \quad (5)$$

where $\ln[HHI_{it}]$ is the natural logarithm of the HHI of keyword *i* in our sample of Spanish news outlets in month-year *t*, $CoreUpdate_t$ is a dummy variable that takes value 1 if in month-year *t* there is a Google core update, $CoreUpdatePlus1_t$ is a dummy variable that takes value 1 if in month-year *t-1* there is a Google core update, and $CoreUpdatePlus2_t$ is a dummy variable that takes value 1 if in month-year *t-2* there is a Google core update. Some specifications also include equivalent dummy variables taking value 1 when a *Big* update takes place. All specifications include keyword, month-run length, and month-year fixed effects. The usual assumption applies to the error term ε_{it}. We cluster the standard errors at the keyword level. Finally, we drop keywords that only appear in our data in one month or in one outlet.

Table 13 presents our findings. Column 1 estimates specification (5) and shows overall keyword concentration decreases by 3.4% in the month of the update, remains unchanged the following month, and then increases by 3.1% i two months after the update. Column 2 differentiates *Big* and *Not Big* core updates and shows that all the effect in column 1 comes from *Big* core updates (August 2018, June 2019 and May 2020).

The next columns analyze the existence of heterogeneous effects of core updates across different groups of keywords. Columns 3 and 4 classify keywords by the number of sites they appear in. Column 3 uses the sample of keywords appearing in 5 sites (i.e. the median number of sites a keyword appears in our sample) or more, while Column 4 uses the sample of keywords appearing in 4 sites or less. We find that our main result is driven by keywords appearing in more outlets, where the HHI drops by 8% in the month of a *Big* update, remains unaffected the following month, and then spikes by 8% two months after the update. There is no apparent impact on keywords appearing in few outlets. Columns 5 and 6 show findings after classifying keywords into those appearing for more than 8 months and those appearing for 8 months or less. Our results show opposite effects of the updates on these two groups. Column 5 shows that keywords with longer runs experience an increase in concentration after the updates, with no additional impact from a *Big* core update. In contrast, keywords with shorter runs are not affected by general updates, but they experience a 5% reduction of concentration following a *Big* update, followed by a 5% increase of concentration two months later. Finally, columns 7 and 8 separate keywords by whether they were ever a top 10 keyword by any news outlet in our sample. Here again we find interesting and important (magnitude wise) effects that differ by type of keywords. Column 7 shows that *Big* core updates increase the concentration of

keywords ever in the top 10 search results by 6.6%, and that the effects of the other updates are negligible. In contrast, the concentration of keywords never on the top 10 decreases with *Big* updates by 9.1%, but it increases with *Not Big* updates by 1.8%.

Overall, these results suggest that core updates increase market concentration for those keywords that are more frequently used in queries and that rank higher in Googles' first results page. These are precisely the keywords that are more likely to generate visits to the outlets. Moreover, the updates decrease market concentration for keywords that are used less frequently by readers and for which outlets offer a worse match with the queries.

7.2. Separating Google from Other Search Engines

Google core updates do not affect the indexation algorithms of other search engines. Thus, we can examine through a difference-in-differences analysis whether core updates changed the concentration of Google search visits across outlets, relative to the concentration of other search engines. To do this exercise, we use monthly data on the search traffic of the four largest search engines in Spain: Google (32% market share of *total* visits), Yahoo (0.3%), DuckDuckGo (0.2%) and Bing (0.1%). We estimate the following model,

$$\ln[HHI_{st}] = Google_s * CoreUpdate_t + Google_s * CoreUpdatePlus1_t, + Google_s * CoreUpdatePlus2_t + \mu_s + \partial_t + u_{st} \quad (6)$$

where $\ln[HHI_{st}]$ is the natural logarithm of the HHI of search engine *s* in month-year *t*. We calculate our main dependent variable using the market shares of the news outlets in each search engines. The variables $CoreUpdate_t$, $CoreUpdatePlus1_t$, and $CoreUpdatePlus2_t$ are defined as in model (5) above. Likewise, in some specifications we include dummy variables for *Big* Google core updates in addition to dummies for all core updates. All specifications include search engine, month, and year fixed effects. The usual assumption applies to the error term u_{st}. We cluster the standard errors at the search engine level.

Panel A of Table 14 investigates the impact of core updates in search engine concentration. Column 1 estimates equation (6) and shows no impact of the core updates. Column 2 splits *Big* updates from other updates, and shows that *Not Big* core updates reduced concentration of Google Search visits by 2% on the month of the updates, had no statistically significant impact one month after, and increased market concentration by 4.8% two months after the updates. Column 2 also shows *Big* core updates (August 2018, June 2019 and May 2020) could decrease concentration by 6.5% two months after the *Big* update takes place, but did not have an immediate impact.

Finally, panel B of Table 14 shows how core updates changed the concentration of search engine visits at the outlet level. Column 3 estimates specification (6) using $\ln[Conc_{it}]$ as

dependent variable, which is the natural logarithm of the HHI of news outlet *i* in month-year *t*. This variable is calculated using the market shares of search engines in each news outlet. Our analysis shows no impact of the updates on the concentration at the outlet level. Column 4 differentiates *Big* and *Not Big* core updates, showing that only *Big* updates affected search engine visit concentration at the outlet level. In particular, it raised concentration by 0.6% on the month of the updates, and by 0.4% one month after.

8. Conclusion

Search engines are crucial intermediaries to access content in the internet. In the case of news media, consumers frequently look for the latest news in Google, Bing or Yahoo, rather than directly visiting online news outlets. They expect search engines to provide links to breaking news, top stories, weather updates, business, entertainment, and political information. This raises the question of whether and how search engines influence citizens' access to a diverse range of high-quality news, opinion-based editorials, and information analyses from various sources. The concern is not just about how news outlets can optimize their headlines and stories to rank higher in search results, but it is also about the risk that search engines reinforce the market and the media power of some publishers.

Our paper constitutes a first step to study these questions by examining Google's ability to influence the type of news outlets accessible to readers. We have addressed two basic questions. First, we have analyzed the mechanisms that determine the number of visits news outlets receive from Google. Each time a consumer queries a specific keyword, Google ranks web pages considering how accurately they answer the query and displays them on its search results page. The rank a news outlet obtains for each keyword determines its search visits, prompting to invest important resources to secure top positions for relevant keywords. To isolate the effects of Google's indexation on search visits, we use a two-stage identification strategy, leveraging Google's core algorithm updates as a quasi-exogenous source of variation in news outlets' indexation.

Our results show that Google's core updates rolled out between 2018 and 2020 affected news outlets in different directions, with an overall negative effect on the number of keywords for which outlets ranked on top search positions. We find that the number of keywords in top 10 search results positively influences visits. These results hold across different classifications of news outlets (e.g., specialization, national rank).

The second question addressed in our paper is whether Google's core updates have increased market concentration in European media. We find that the three *Big* core updates reduced market concentration by 1%, and that this effect was mostly offset by the *Not Big* core updates, which increased concentration by 0.08%. Additionally, *Not Big* updates increased the concentration of search visits for National news outlets while decreasing it for Sports outlets. We find heterogeneity across countries: while *Big* core updates reduced market concentration in Finland, Germany, and Greece, *Not Big* updates

increased concentration in Finland and Netherlands. Overall, we find no sizable impact on market concentration from any given algorithm update.

Additionally, using Spanish keyword search data, we find that *Big* updates decreased concentration at the market-keyword level. However, these results vary depending on the importance of the keyword, its coverage across news sites, and its run length. For example, updates significantly increased market concentration for keywords that are frequently used in queries and that rank higher in Googles' search results.

Overall, our findings suggest that changes in Google's indexation algorithms can be sufficiently important to modify competition in the media market when focusing on search traffic of a particular set of outlets or specific queries, yet not important enough when using country-level concentration as primary variable of analysis. This conclusion has important implications for policy makers interested in understanding competition in online markets.

While our paper has focused on the media market, more research is needed to examine how search engines shape search traffic and competition in other markets. The European Union has recently established new regulations to improve transparency in online intermediation activities. In 2019, it approved the platform-to-business (P2B) regulation, designed to create a fair, transparent and predictable business environment for small online businesses and traders (European Commission, 2019).[37] In 2020, it developed new instruments to regulate online intermediaries, through the Digital Services Act (DSA) and the Digital Markets Act (DMA). These regulations, approved in 2022 and enacted in 2024, will transform market governance, introducing obligations for platforms to share data with regulators and researchers, and increasing transparency around algorithms and content moderation activities. In 2023, the EU designated Google and Bing as *Very Large Search Engines* that have to comply with a full set of obligations under the DSA.[38]

Our analysis does not consider how human editorial decisions in news outlets is replaced by algorithms offering personalized recommendations to readers (Agrawal et al. 2018; Claussen et al. 2021). As noted by Gentzkow (2018), "many of the deepest problems in media today stem not from an inability to give consumers what they want, but from the fact that what they appear to want is not aligned with what is good for society". As news outlets' algorithms better cater to consumers tastes, there is a risk of confining users to echo chambers, reinforcing biases through prior individual-level data (Sunstein, 2001; Boxell, Gentzkow, and Shapiro, 2017; Gentzkow, 2018; Goldfarb and Tucker, 2019; Claussen et al., 2021). Another relevant aspect not considered in our analysis is competition between news outlets and search engines to attract users and gather proprietary data for the advertising market (Prat and Valletti, 2021).

[37] This regulation, effective since July 12, 2020, requires search engines to disclose the main parameters that determine their rankings and the importance of these parameters. Platforms must also reveal if rankings are influenced by payments from business users and describe any preferential treatment given to their own goods or services compared to third-party businesses.
[38] See https://ec.europa.eu/commission/presscorner/detail/en/IP_23_2413.

Finally, our paper is relevant for understand the role of search engines on journalism and democratic institutions. Gentzkow and Shapiro (2010) note that the US regulation of news media ownership is based on the proposition that news content significantly influences politics and that unregulated media markets often lack ideological diversity. This belief has justified strict controls on market consolidation and media ownership. The rise of digital platforms and social networks poses new challenges for media regulation because, despite facilitating access to news contents, they may limit the diversity of information sources available to consumers.

References

Agrawal, A., J. Gans and A. Goldfarb (2018), Prediction Machines: The simple economics of artificial intelligence. Harvard Business Press.

Aguiar. L., J. Waldfogel and S. Waldfogel (2021), Playlisting Favorites: Measuring Platform Bias in the Music Industry, International Journal of Industrial Organization, 78, 102765.

Allcott. A., M. Gentzkow and C. Yu (2019), Trends in the diffusion of misinformation on social media, Research and Politics, April-June 2019: 1–8.

Allcott, H., J. C. Castillo, M. Gentzkow, L. Musolff and T. Salz (2024), Sources of Market Power in Web Search: Evidence from a Field Experiment, unpublished manuscript.

Anderson, C. (2004). The long tail. Wired. Issue 12.10. October.

Athey, S., M. Mobius and J. Pal (2017), The Impact of Aggregators on Internet News Consumption, Working Paper, Microsoft Research.

Bandy, J. and N. Diakopoulos (2020), Auditing News Curation Systems: A Case Study Examining Algorithmic and Editorial Logic in Apple News. Proc. International Conference on Web and Social Media (ICWSM).

Barocas, S. and Selbst, A. (2016). Big Data ' s Disparate Impact. California law review, 104(1):671–729.

Baye, M. R., B. De los Santos and M. R. Wildenbeest (2016), Search Engine Optimization: What Drives Organic Traffic to Retail Sites?, Journal of Economics and Management Strategy 25(1), 6-31.

Belleflamme, P. and M. Peitz (2018), Inside the engine room of digital platforms: Reviews, ratings, and recommendations. Working Paper.

Belleflamme, P. and M. Peitz (2021), The Economics of Platforms: Concepts and Strategy. Cambridge University Press.

Besley, T and A. Prat (2006), Handcuffs for the grabbing hand? Media capture and government accountability, American Economic Review, 96, pp. 720-736.

Bourreau, M. and G. Gaudin (2022), Streaming platform and strategic recommendation bias, Journal of Economics & Management Strategy, 31(1), 25-47.

Bourreau, M., J. Krämer and J. Hofmann (2021), Prominence-for-data schemes in digital platform ecosystems. Working Paper.

Boxell, L., Gentzkow, M., and Shapiro, J. M. (2017), Greater Internet use is not associated with faster growth in political polarization among US demographic groups." Proceedings of the National Academy of Sciences of the United States of America, 19, 1–6.

Brin, S. and L. Page (1998), The Anatomy of a Large-Scale Hypertextual Web Search Engine, Working Paper: https://www.yumpu.com/en/document/read/11535048/the-anatomy-of-a-large-scale-hypertextual-web-search-engine

Brynjolfsson, E., Y. Hu, and M.D. Smith. (2010), "Research Commentary - Long Tails vs. Superstars: The Effect of Information Technology on Product Variety and Sales Concentration Patterns" Information Systems Research, 736-747.

Brynjolfsson, E., Y. Hu, and D. Simester (2011), Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product Sales, Management Science, 57 (8), 1373-1386.

Burguet, R., R. Caminal and M. Ellman (2015), In Google we Trust?, International Journal of Industrial Organization, 39, 44-55.

Cagé, J., N. Hervé and M.H. Viaud (2020), The Production of Information in an Online World. The Review of Economic Studies, 87 (5), 2126–2164.

Cagé, J., N. Hervé and B. Mazoyer (2020), Social Media and Newsroom Production Decisions, Working Paper.

Calvano, E., G. Calzolari, V. Denicolo and S. Pastorello (2023), Artificial Intelligence, Algorithmic Recommendations and Competition, Working Paper.

Calzada, J, and R. Gil (2020), What Do News Aggregators Do? Evidence from Google News in Spain and Germany. Marketing Science, 39(1):134-167.

Campbell, A., Leister, C.M. and Zenou, Y. (2019), Social media and polarization, CEPR Discussion Paper No. DP13860.

Chandra, A., M. Suaib and R. Beg (2015), Google Search Algorithm Updates Against web Spam, Informatics Engineering, and International Journal, 3 (1).

Chiou, L. (2017), Vertical Integration and Antitrust in Search Markets, Journal of Law, Economics, and Organization, Vol. 33, 653-685, 2017.

Chiou, L., and Tucker, C. (2017), Search engines and data retention: Implications for privacy and antitrust. Working Paper.

Chiou, L., and C. Tucker (2017), Content Aggregation by Platforms: The Case of the News Media, Journal of Economics and Management Strategy, 26 (4), 782-805.

Choi, H., C.F.Mela, S. Balseiro, and A. Leary (2020), Online Display Advertising Markets: A Literature Review and Future Directions, Information System Research, 31,2.

Chiou, L., and C. E. Tucker (2024), Algorithms and Online News. Available at SSRN: https://ssrn.com/abstract=4998787.

Claussen, J., C. Peukert amd A. Sen (2021), The Editor and the Algorithm: Returns to Data and Externalities in Online News. Working Paper.

Cowgill, B. and C. Tucker (2019), Economics, Fairness and Algorithmic Bias, in preparation for the Journal of Economic Perspectives.

Datta, A., M. C. Tschantz, and A. Datta (2015), Automated experiments on ad privacy settings. Proceedings on Privacy Enhancing Technologies 2015 (1), 92–112.

Decarolis, F. M. Li and F. Paternollo (2023), Competition and Defaults in Online Search, CEPR Discussion Paper 17779.

Decarolis, F. and G. Rovigatti (2021), From Mad Men to Maths Men: Concentration and Buyer Power in Online Advertising, American Economic Review, 111 (10), 3299-3327.

Drugov, M. and D.-S. Jeon (2024), Recommendation Bias and Consumer Manipulation. Working Paper.

Duggan, J. and C. Martinelli (2011), A spatial theory of media slant and voter choice, The Review of Economic Studies, vol. 78(2), pp. 640–66.

Dujeancourt, E. and M. Garz (2023), The Effects of Algorithmic Content Selection on User Engagement with News on Twitter, The Information Society.

Edelman, B. (2011), Bias in search results: Diagnosis and response, Indian Journal of Law and Technology, 7, 16.

Enikolopov, R., A. Makarin and M. Petrova (2020), Social media and protest participation: evidence from Russia', Econometrica, vol. 88(4), 1479–514.

European Commission (2019), Platform-to-business trading practices.

European Commission (2020), Digital markets act: Ensuring fair and open digital market.

Fleder, D. and K. Hosanagar (2009), Blockbuster Culture's Next Rise or Fall: The Impact of Recommender Systems on Sales Diversity, Management Science, 55 (5): 697–712.

Gehlbach, S. and K. Sonin (2014), Government control of the media, Journal of Public Economics, vol. 118, pp. 163–71.

Gentzkow, M. (2018), Media and artificial intelligence. Working Paper.

Gentzkow, M. and J. M. Shapiro (2008), Competition and truth in the market for news. J. Econ. Perspect., 22, 133-154

Gentzkow, M., and J.M. Shapiro (2010), What drives media slant? evidence from us daily newspapers. Econometrica, 78 (1), 35–71.

Gentzkow, M. and J.M. Shapiro (2011), Ideological segregation online and offline, The Quarterly Journal of Economics, 126(4), 1799–839.

George, L.M. and C. Hogendorn (2020), Local News Online: Aggregators, Geo-Targeting and the Market for Local News. Journal of Industrial Economics, 68: 780-818.

Germano, F., V. Gomez and F. Sobbrio (2022) Crowding out the truth? A simple model of misinformation, polarization, and meangful social interactions, Working Paper.

Giovanniello, M.A. (2017), Echo chambers: voter-to-voter communication and political competition', Working Paper.

Goldfarb, A. and C. Tucker (2019), Digital Economics. Journal of Economic Literature 57(1), 3-43.

Guess, A. M., N. Malhotra, J. Pan, P. Barberá, H. Allcott, T. Brown, A. Crespo-Tenorio, D. Dimmery, D. Freelon, M. Gentzkow, S. González-Bailón, E. Kennedy, Y. M. Kim, D. Lazer, D. Moehler, B. Nyhan, C. Velasco Rivera, J. Settle, D. R. Thomas, E. Thorson, R. Tromble, A. Wilkins, M. Wojcieszak, B. Xiong, C. K. de Jonge, A. Franco, W. Mason, N. J. Stroud and J. A. Tucker (2023), Reshares on social media amplify political news but do not detectably affect beliefs or opinions. Science 381(6656), 404-408.

Hagar, Nick and Nicholas Diakopoulos (2019). Optimizing Content with A/B Headline Testing: Changing Newsroom Practices. Media and Communication 7 (1):117–127.

Hagiu, A. and B. Jullien (2011), Why do intermediaries divert search?, The RAND Journal of Economics, 42, 337-362.

Hervas-Drane, A. (2015), Recommended for you: The effect of word of mouth on sales concentration, International Journal of Research in Marketing, 32, 207-218.

Hosanagar, K., D. Fleder, D. Lee, and A. Buja (2014), Will the global village fracture into tribes? recommender systems and their effects on consumer fragmentation, Management Science, 60 (4), 805–823.

Huang, Y. and Yu Xie (2023), Search algorithm, repetitive information, and sales on online platforms, International Journal of Industrial Organization, 88, 102933.

Hunold, M., R. Kesler, and U. Laitenberger (2020), Rankings of online travel agents, channel pricing, and consumer protection, Marketing Science, 39, 92- 116.

Lambrecht, A. and C. Tucker (2019), Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of STEM career ads, Management Science, 65, 2966-2981.

Lee, D. and K. Hosanagar (2019), How Do Recommender Systems Affect Sales Diversity? A Cross-Category Investigation via Randomized Field Experiment, Information System Research, 30 (1).

Leung, T. C. and K. Strumpf (2023). All the Headlines that Are Fit to Change. Working Paper.

Marchiori, M. (1997). The Quest for Correct Information on the Web:Hyper Search Engines. Working Paper Universita di Padova.

Martin, G. J., N. Mastrorocco, J. McCrain, and A. Ornaghi (2024), Media Consolidation. CESifo Working Paper No. 11356.

Noble, S.U. (2018) Algorithms of Oppression: How Search Engines Reinforce Racism, NYU Pres.

Oestreicher-Singer, G. and A. Sundararajan (2012), The Visible Hand? Demand Effects of Recommendation Networks in Electronic Markets. Management Science, 58 (11), 1963-1981.

Oliveros, S. and F. Vardy (2015), Demand for slant: how abstention shapes voters' choice of news media', Economic Journal, 125(587), 1327–68.

Peltier, S., Benhamou, F. and M. Touré (2016), Does the long tail really favor small publishers? *Journal of Cultural Economics*, 40 (4): 393-412.

Peterson, E., S. Goel and S. Iyengar (2019), Partisan selective exposure in online news consumption: evidence from the 2016 presidential campaign, Political Science Research and Methods, 1-17.

Piolatto, A. and F. Schuett (2015), Media competition and electoral politics, Journal of Public Economics, 130, 80–93.

Pogorelskiy, K. and M. Shum (2019), News we like to share: how news sharing on social networks influences voting outcomes. Working Paper.

Prat, A. (2018), Media power, *Journal of Political Economy*, 126 (4), 1747-1783.

Prat, A. and T. Valletti (2021), Attention Oligopoly, American Economic Journal Microeconomics, forthcoming.

Schaefer, M., and G. Sapi (2020), Learning from Data and Network Effects: The Example of Internet Search. DIW Discussion Paper 1894.

Scott, M. (2017), Google Fined Record $2.7 Billion in E.U. Antitrust Ruling, N.Y. Times https://www.nytimes.com/2017/06/27/technology/eu-google-fine.html.

Sismeiro, C. and A. Mahmood (2018), Competitive Versus Complementary Effects in Online Social Networks and News Consumption: A Natural Experiment, Management Science, 64, 5014-5037.

Simonov, A., C. Nosko, and J. Rao (2018), Competition and Crowd-out for Brand Keywords in Sponsored Search, Marketing Science, 37 (2), 200-215.

Sweeney, L. (2013), Discrimination in online ad delivery. ACMQueue 11 (3), 10.

Taylor, G. (2013), Search quality and revenue cannibalization by competing search engines. Journal of Economics and Management Strategy 22, 445-467.

Tolan, S. (2019), Fair and unbiased algorithmic decision making: current state and future challenges, JRC Digital Economy Working Paper 2018-10.

Waldfogel, J. (2017), How Digitalization Has Created a Golden Age of Music, Movies, Books, and Television, Journal of Economic Perspectives, Vol. 31 (3), 195-214.

White, A. (2013), Search engines: Left side quality versus right side profits. International Journal of Industrial Organization 31, 690-701.

Xu, L., Chen, J., and A. Winston (2012), Effects of the presence of organic listing in search advertising. Information System Research 23, 1284-1302.

Yun, J.M. (2018), Understanding Google's Search Platform and the Implications for Antitrust Analysis, Journal of Competition Law & Economics, 14 (2), 311-329.

Zhang, L. (2018), Intellectual Property Strategy and the Long Tail: Evidence from the Recorded Music Industry, Management Science, 64 (1): 24–42.

Zhuravskaya, E., M. Petrova and R. Enikolopov (2020), Political Effects of the Internet and Social Media, Annual Review of Economics, 12: 415-38.

Appendix A

This appendix provides additional information on the two global core updates discussed in sections 3.2 and 5.1, respectively. We begin describing the 2018 Medic update and then we present the May 2020 Pandemic update.

The August 2018 update (Medic update). This update was released in August 1, 2018, and despite its broad effect across all types of websites, it was nicknamed "Medic" due to its significant impact on medical sites. The main goal of the update was to improve Google's ability to identify the authority and expertise of website content,[39] and to rank web pages more accurately in search results.

This update focused on those pages that can affect the users' current or future well-being, known as "Your Money or Your Life" (YMYL) pages. These include pages with information on financial products or medical treatments that can influence users' decisions. The update aimed to protect users from harmful information by refining the algorithms that evaluate the relevance and authority of web page content. Following the update, Google published a white paper in early 2019 confirming that its ranking algorithm in YMYL queries would put more weigh on factors such as Expertise, Authoritativeness, or Trustworthiness (EAT as known in the SEO industry).[40]

Essentially, Google's algorithm looks for signals in each web page that correlate with EAT. The best known of these signals is PageRank, which relates a page's authority to the number of links it receives. According to industry experts, the Medic update increased the ranking of websites offering accurate, reliable information, with a strong reputation and high trust levels among users. After the update, websites with high-quality, well-written content relevant to the user queries were more likely to rank higher in the search results.

The update also favored websites that were well-structured, with clear navigation and logical hierarchy of pages. This included factors such as the use of headings and subheadings, the presence of internal links, and the use of alt tags on images. Additionally, it penalized websites with spam content or technical issues such as broken links, slow loading times, imperfect usability on different devices, and annoying ads. These changes affected the ranking of sites competing for health, medical, legal, and finance keywords and also affected newspapers, public information sites and pages dedicated to ecommerce.

The rollout of the update was gradual and took several weeks. The first stage, starting on August 1, 2018, focused on critical elements such as evaluating the quality and expertise

[39] In March 2013, Google published its Search Quality Rating Guidelines, describing its search quality evaluation and its definition of 'High Quality Pages'. Google's rankings reward content and entities that demonstrate high levels of Expertise, Authoritativeness and Trustworthiness (EAT). In March 2014, Google first explicitly added the EAT measures to its guidelines, and it introduced EEAT in December 2022, adding an extra "E" for "Experience". Industry experts have identified several signals that Google uses to evaluate EEAT: quality of the website content, PageRank or references to the publisher, distance to trust sites in the link graph,; anchor text of backlinks, credibility of an author, name recognition of the author/publisher, sentiment around mentions or ratings and click-through rate, co-occurrences of the author/publisher with thematically relevant terms in videos, podcasts and documents and in search queries, transparency to the author/publisher via author profiles, links to own references, and agreement with common opinion and facts. See https://marketbrew.ai/a/google-medic-update#what-was-the-main-goal-of-the-google-medic-update

[40] https://kstatic.googleusercontent.com/files/388aa7d18189665e5f5579aef18e181c2d4283fb7b0d4691689dfd1bf92f7ac2ea6816e09c02eb98d5501b8e5705ead65af653cdf94071c47361821e362da55b

of websites. In the second stage, Google focused on how it assesses the relevance of websites, while the third stage modified how it evaluates the overall user experience.

The May 2020 update (Pandemic update). On May 4, 2020, Google released its first core algorithm update following the start of the Coronavirus pandemic. The aim of this update was to adjust the algorithm to new user behavior patterns. During the first months of the pandemic, digital news outlets registered unprecedented levels of traffic due to the high demand for COVID-19 related news and local news. Comscore reported that after the World Health Organization declared the state of alarm, traffic to general news outlers increased by 87%, while local news traffic increased by 158%.[41] Users sought information about online services and products, and news about the pandemic, a phenomenon known as "infodemic".[42]

Google also adjusted its SERP in response to the pandemic. In March 2020, Google introduced a COVID-19 Left Panel with information on the pandemic and by the end of April 2020, it added a local news box.[43] According to Google Search Liaison, "*last month we made a change to help local content surface better within our Top Stories box. In addition, we added a new "Local news" box to show local news content in response to coronavirus-related searches.*" As a result, people searching for coronavirus information could see both a "top stories carousel" as well as a "local news carousel" with coronavirus information (see Figure A1 below).

According to industry experts, one of the objectives of the May update was to ensure that local news were not buried down below larger national news outlets covering the same topics, and to increase the visibility of publishers with high authority (EAT). On May 4, without directly referring to the update, the Google Search Liaison clarified that local news content is "*material produced by publishers like local newspapers, that we understand specifically serve particular cities or regions. It can also be content we identify as relevant to a particular area, even if written by those outside the area....*" In addition, it stated that "*for publishers, there's no particular change to make. We already review a number of signals to understand if content seems particularly relevant to a location. Our change is simply to ensure we're better balancing those signals to match local content to those searching for it...".*

Google complemented this change with other initiatives to protect local newspapers. On April of 2020, Google News launched the "Journalism Emergency Relief Fund" to deliver urgent aid to thousands of small and medium-sized local news publishers

[41] https://www.reddeperiodistas.com/las-noticias-que-mas-se-leen-por-la-crisis-del-coronavirus/

[42] In the day of the update announcement, Google said that it "had *never seen as many searches for a single topic continue over a sustained period as is happening now with COVID-19. Many searches are for news about what's happening in local areas, such as sheltering updates or the latest on testing. Here's how we are helping*…". See https://twitter.com/searchliaison/status/1257368113676517376.

[43] On March 2, 2020 Google introduced a COVID-19 SOS alert at the top of the results and a "Help and information" and "Safety tips" box appeared between the top stories and organic listings. On March 20, the SOS alert was removed, but a tabbed COVID-19 Alert section (highlighted in red) began appearing on the right sidebar. The tabs included symptoms, prevention, treatments and an overview section with the "Help and information" box inside it. On March 23, the COVID-19 Alert section moved to the left side of the page and a statistics tab was added to it. The *Covid-19 Left Panel* offered information on the coronavirus including 'Top Stories', 'Map of cases', 'Local News', and 'Health Information' covering symptoms, prevention and treatments information from the NHS. For a detailed illustration see https://searchengineland.com/corona-to-covid-how-googles-corona-results-page-has-evolved-331245 and See https://mrs.digital/blog/google-serp-changes-2020/

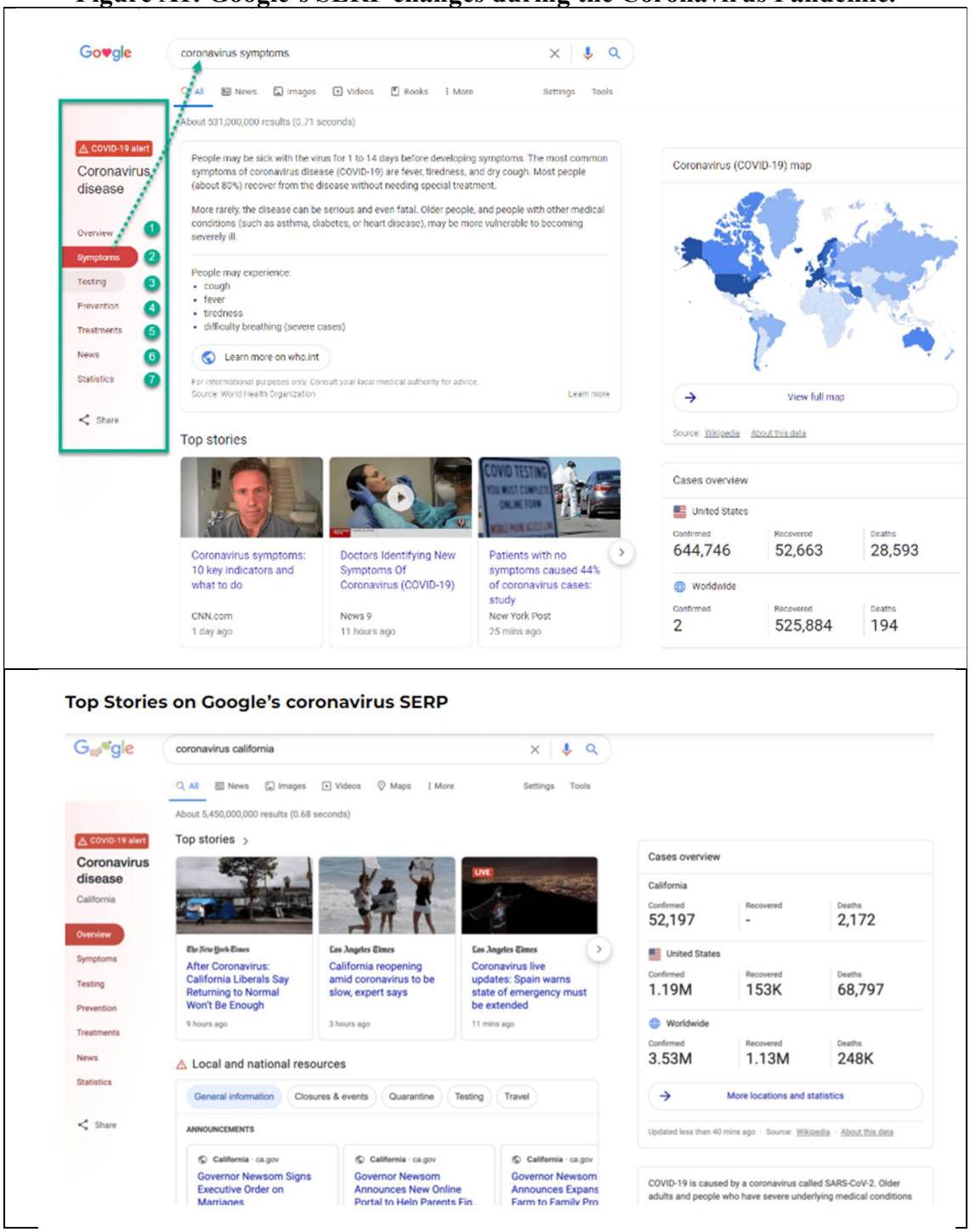
globally.[44] Google argued that the economic downturn caused by COVID-19 had significantly impacted on local publishers, leading to job cuts and layoffs.

In the months following the update, Google Search introduced new news features to help readers find content from local publishers. It expanded the carousel of local news stories initially launched for COVID searches. Additionally, local news publishers began appearing more often in general news features such as Top Stories.[45] Google also enhanced users' ability to find local topics beyond broad categories, covering more specialized subtopics. For example, users looking for basketball results in a particular city could find results for high school and college teams, not just professional teams.[46] Later on, Google also helped users find local information through tweets from local authoritative sources. These changes in the SERP allowed Google to better match users' queries without using personal information. Users can choose the type of news they want to see, local or general, reducing the need to redefine their search.[47]

The update was fully rolled out by May 18, 2020. Industry experts considered it bigger than the January 2020 update. Some consider it as an "absolute monster".

[44] https://blog.google/outreach-initiatives/google-news-initiative/global-journalism-emergency-relief-fund-local-news/
[45] https://www.reddeperiodistas.com/las-noticias-locales-permiten-colarse-en-google-discover/
[46] https://blog.google/products/news/local-news-update-census-mapper/
[47] https://www.hillwebcreations.com/google-top-stories-and-news-boxes/

Figure A1: Google's SERP changes during the Coronavirus Pandemic.



Sources: https://searchengineland.com/how-covid-19-may-change-the-serp-forever-333183; https://position1seo.co.uk/what-we-know-about-the-may-2020-google-algorithm-update/ and https://dl.acm.org/doi/pdf/10.1145/3576840.3578307

Figure 1A. Google's Core Updates Announcements: 2018-2020

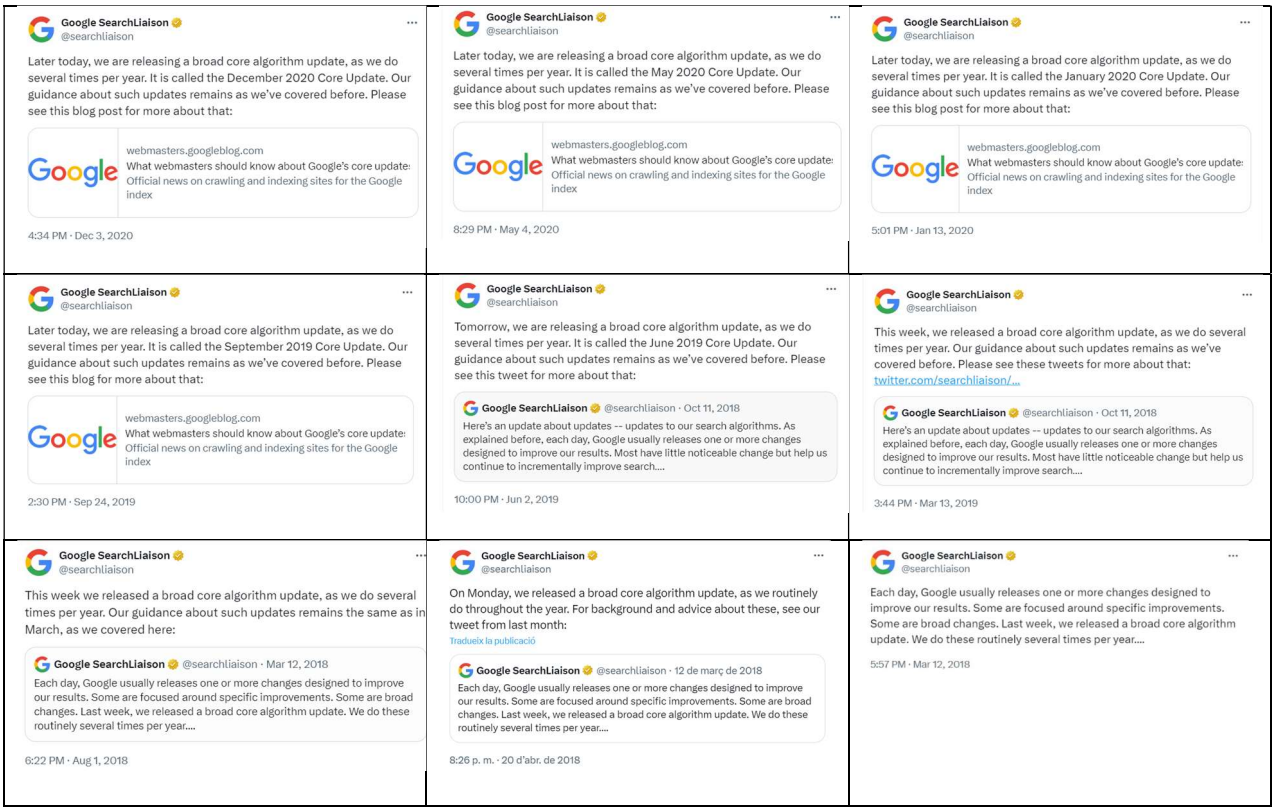


Figure 1B. Google's Core Update Announcement of May 2020

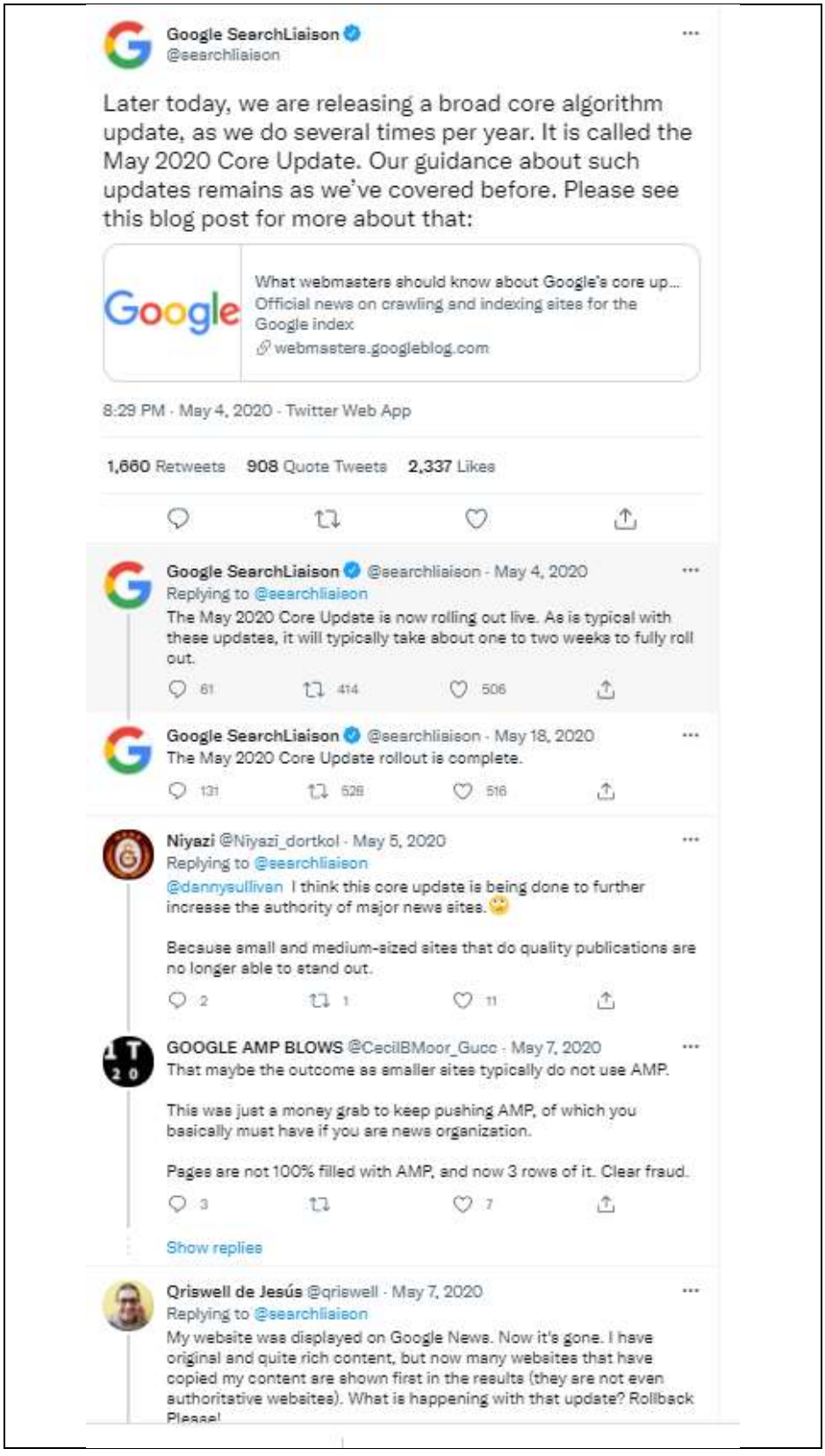
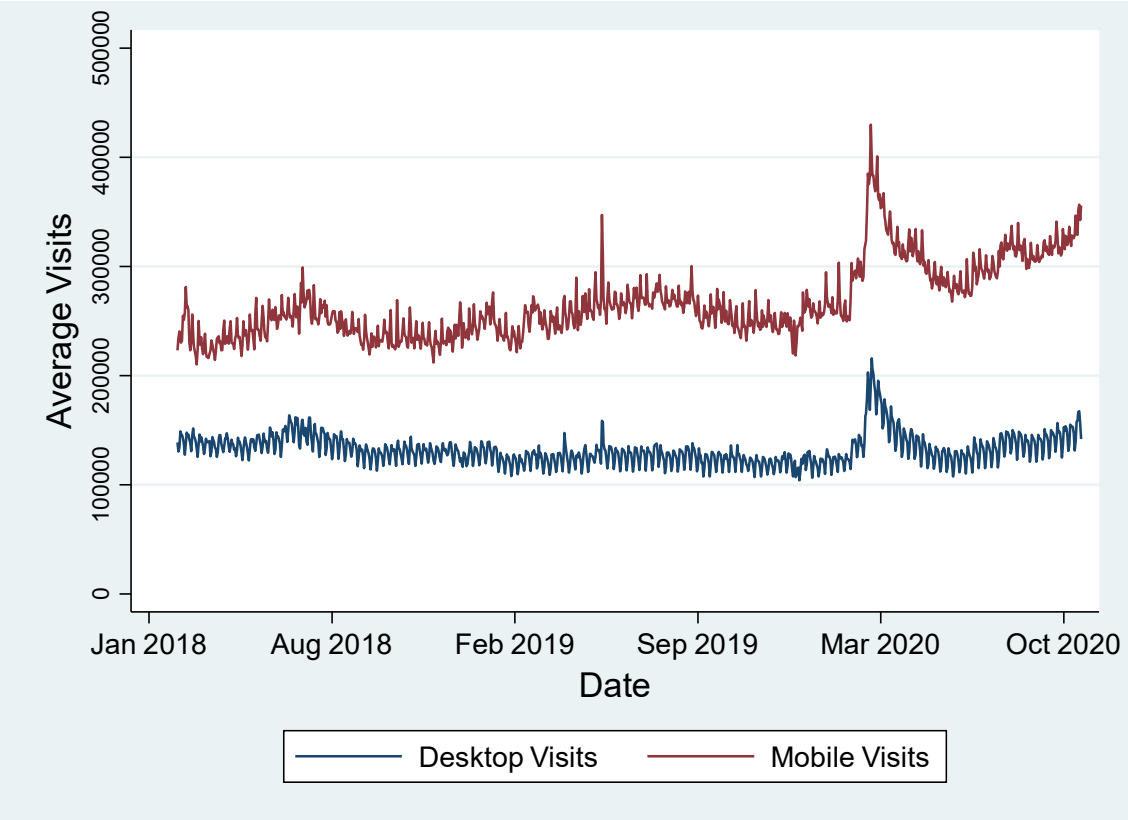
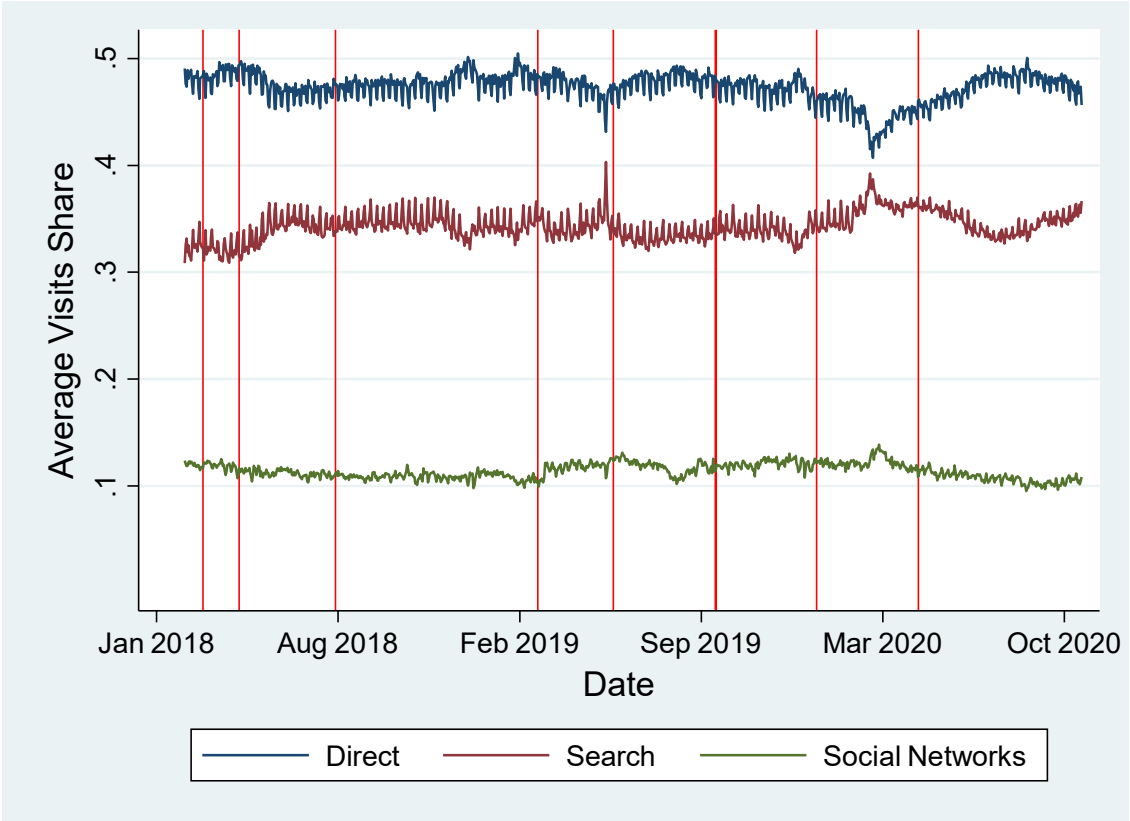


Figure 2. Desktop and Mobile Daily Visits
(January 2018 - November 2020)



This Figure shows the average number of desktop visits (blue line) and the average number of mobile visits (red line) for all outlets in our sample between January 2018 and October 2020. The Figure also shows a dramatic increase in the visits after the WHO declared the coronavirus a global pandemic on March 11, 2020.

Figure 3. Share of Desktop Direct, Search and Social Networks Daily Visits (October 2017-December 2020)



This Figure shows the average share of visits coming from direct channel (blue top line), organic search channel (red middle line), and social network channel (green bottom line) for all outlets in our sample between January 2018 and October 2020. The red vertical lines illustrate the core algorithm updates introduced by Google in this period.

Figure 4: Example of Google Search's page results

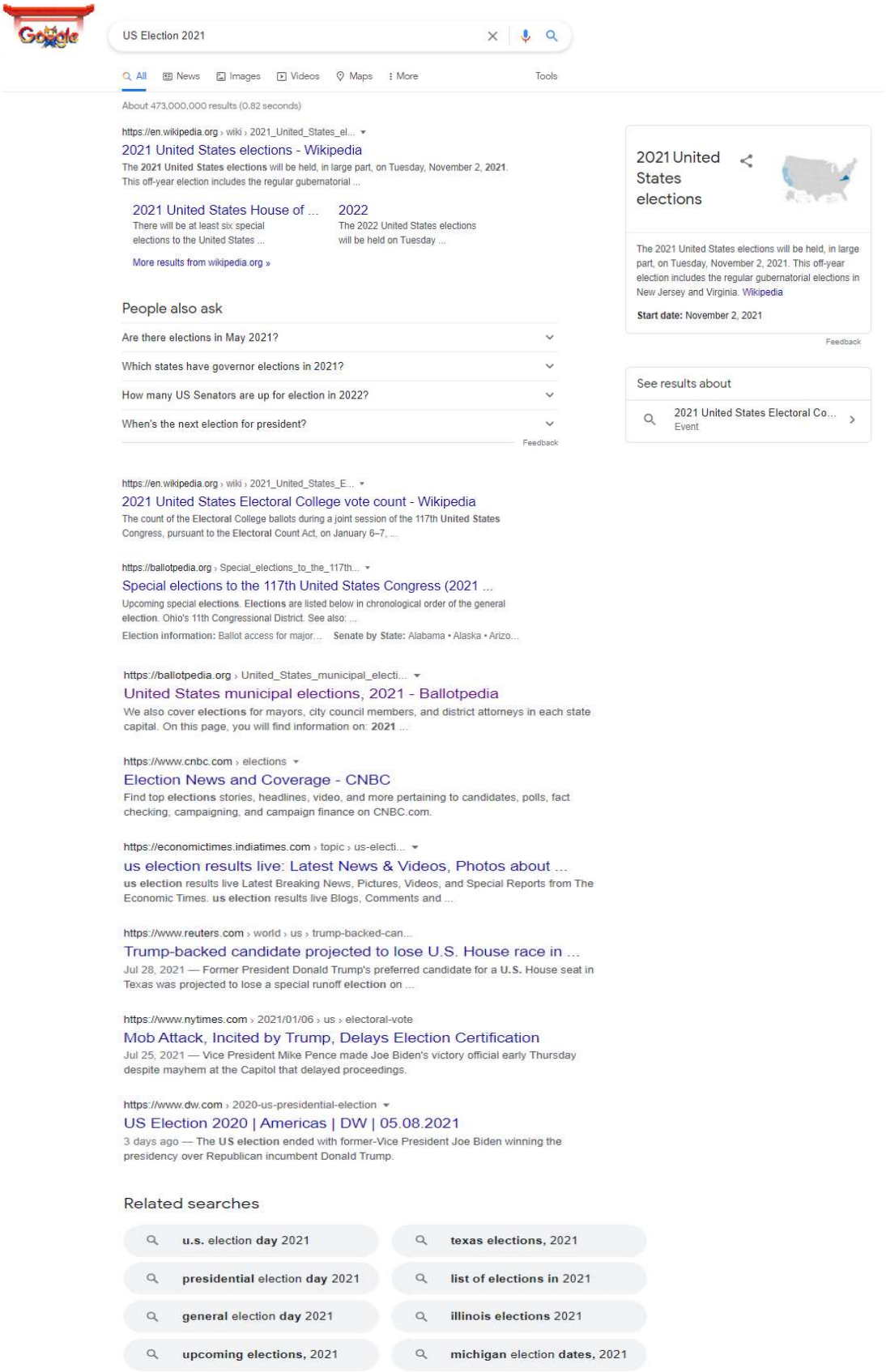
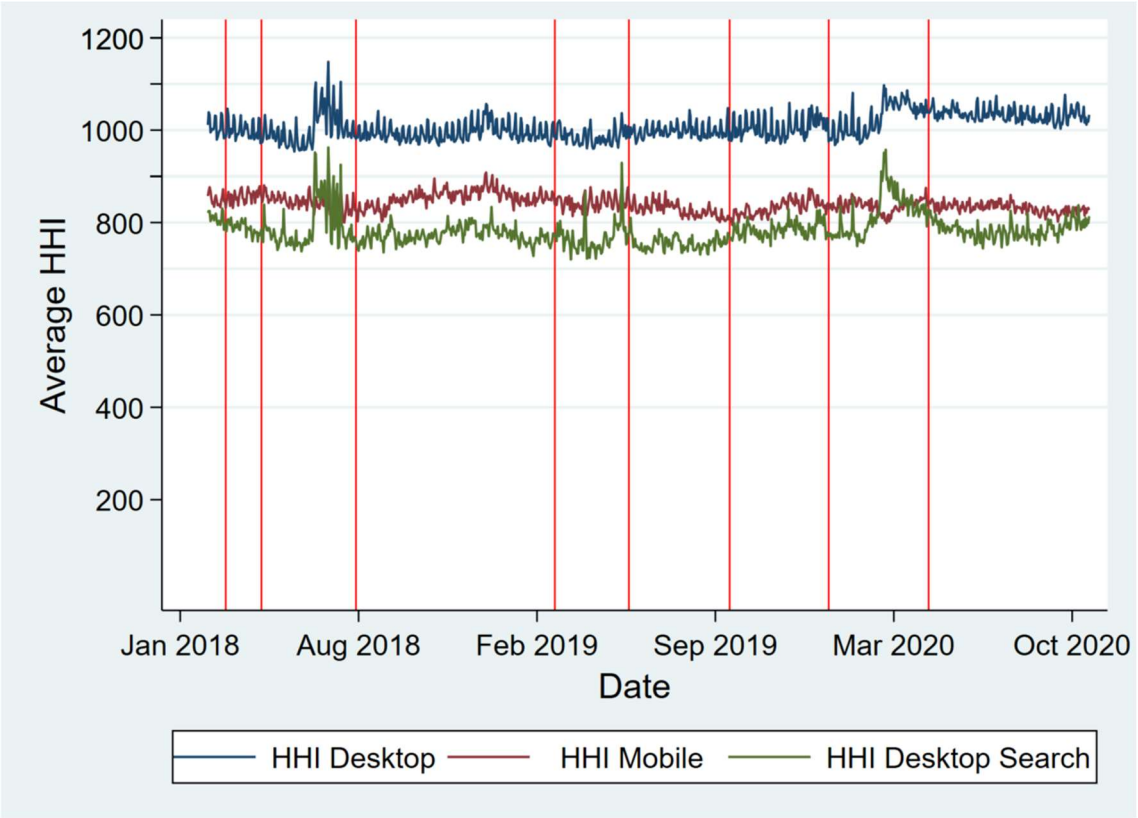


Figure 5. HHI for Desktop Search Visit, Desktop Visits and Mobile Visits (January 2018 - November 2020)



This Figure shows the average daily market concentration (measured through HHI - the Herfindahl–Hirschman Index) in our sample between January 2018 and October 2020. We measure concentration over time in three different markets, in desktop visits (blue top line), in mobile visits (red middle line), and desktop search visits (green bottom line). The red vertical lines illustrate the core algorithm updates introduced by Google in this period.

Table 1A. Google's confirmed core updates

December 2020 Core Update (December 3, 2020)	**Google's Confirmation**: https://twitter.com/searchliaison/status/1334521448074006530 Some industry experts explain that this was of the more impactful algorithm adjustments to hit the SERP over the past year or so.
May 2020 Core Update (May 4, 2020)	**Google's Confirmation**: https://twitter.com/searchliaison/status/1257376879172038656 According to Moz, this update was the second-highest Core Update after the August 2018 "Medic" update.
January 2020 Core Update (January 13, 2020)	**Google's Confirmation**: https://twitter.com/searchliaison/status/1216752087515586560 Moz considers that the effects of this core update were considered smaller than the August 2018 "Medic" core update.
September 2019 Core Update (September 24, 2019)	**Google's Confirmation**: https://twitter.com/searchliaison/status/1176473923833225221 This update focused on improvements in the content quality in the SERPs. For the second time, Google pre-announced a core algorithm update "in advance".
June 2019 Core Update (June 3, 2019)	**Google's Confirmation**: https://twitter.com/searchliaison/status/1135275028834947073 This is considered as one of the Google's most important core updates. Moreover, for the first time in the history of core updates, Google announced this update 24 hours ahead of time on Google Search Liaison Twitter channel. According to Moz, the impact was smaller than the August "Medic" update.
March 2019 Core Update (March 12, 2019)	**Google's Confirmation**: https://twitter.com/searchliaison/status/1105842166788587520 Google stated that this was the third major core update since they began using that label. The update generated ranking shifts for keywords related to health and other sensitive topics. The update affected search queries that are covered by the acronym E-A-T (Expertise, Authoritativeness, and Trust).
Medic Core Update (August 1, 2018)	**Google's Confirmation**: https://twitter.com/searchliaison/status/1024691872025833472 Expert report large impact in search results, specially for health and wellness.
Unnamed Core Update (April 17, 2018)	**Google's Confirmation**: https://twitter.com/searchliaison/status/987397051997663232 According to experts, a heavy algorithm flux that peaked on April 17 and continued for over a week. Google later confirmed a "core" update
Brackets Core Update (March 8, 2018)	**Google's confirmation**: https://twitter.com/searchliaison/status/973241540486164480 Google confirmed a "core" update on March 7th, but volatility spiked as early as March 4th, with a second spike on March 8th, and continued for almost two weeks. The "Brackets" name was coined by Glenn Gabe.

Note: According to Moz, the biggest core updates in this period are those that took place in August 1, 2018, June 3, 2019, and May 4, 2020: https://moz.com/blog/google-organic-click-through-rates-in-2014.
Source: Own elaboration and Moz.com

Table 1B. List of Domains per Country

Austria		Belgium		Denamark		Finland		France	
Site	**Classif.**	**Site**	**Classif.**	**Site**	**Classif.**	**Site**	**Classif.**	**Site**	
apa.at	N	7sur7.be	N	avisen.dk	N	aamulehti.fi	R	20minutes.fr	N
atv.at	TV/R	demorgen.be	N	berlingske.dk	N	ampparit.com	A	bfmtv.com	TV/R
boerse-express.com	B	dhnet.be	N	bold.dk	S	arvopaperi.fi	B	boursier.com	B
bvz.at	R	een.be	TV/R	borsen.dk	B	demokraatti.fi	N	boursorama.com	B
derstandard.at	N	gva.be	R	bt.dk	N	esaimaa.fi	R	capital.fr	B
dietagespresse.com	N	hbvl.be	R	dr.dk	TV/R	ess.fi	R	challenges.fr	B
falter.at	R	hln.be	N	ekstrabladet.dk	N	helsinginuutiset.fi	R	cnews.fr	TV/R
finanzen.at	B	knack.be	N	euroinvestor.dk	B	hs.fi	N	courrierinternational.co	N
fussballoesterreich.at	S	lalibre.be	N	finans.dk	B	iltalehti.fi	N	eurosport.fr	S
golf.at	S	lameuse.be	R	fyens.dk	R	is.fi	N	footmercato.net	S
kleinezeitung.at	R	lanouvellegazette.be	R	information.dk	N	jatkoaika.com	S	france24.com	TV/R
krone.at	N	lavenir.net	N	jv.dk	R	kaleva.fi	N	francetvinfo.fr	TV/R
kurier.at	N	lecho.be	R	jyllands-posten.dk	N	karjalainen.fi	R	huffingtonpost.fr	N
laola1.at	S	lesoir.be	N	kristeligt-dagblad.dk	N	kauppalehti.fi	B	journaldesfemmes.fr	N
ligaportal.at	S	levif.be	N	lokalavisen.dk	R	kouvolansanomat.fi	R	journaldunet.com	B
medianet.at	B	metrotime.be	N	nordjyske.dk	R	ksml.fi	R	ladepeche.fr	R
meinbezirk.at	N	nieuwsblad.be	N	plbold.dk	S	lapinkansa.fi	R	latribune.fr	B
nachrichten.at	R	rtbf.be	TV/R	politiken.dk	N	maaseuduntulevaisuus.fi	R	lavoixdunord.fr	R
news.at	N	rtl.be	TV/R	sn.dk	R	nimenhuuto.com	S	lci.fr	TV/R
noen.at	R	sporza.be	S	stiften.dk	R	osterbottenstidning.fi	R	ledauphine.com	R
oe24.at	N	standaard.be	N	tv2.dk	TV/R	satakunnankansa.fi	R	lefigaro.fr	N
profil.at	N	sudinfo.be	N	tv2lorry.dk	TV/R	savonsanomat.fi	R	lemonde.fr	N
puls4.com	TV/R	tijd.be	B	tv2ostjylland.dk	TV/R	seiska.fi	N	leparisien.fr	R
salzburg24.at	R	vrt.be	TV/R	tv3sport.dk	TV/R	sportti.com	S	lepoint.fr	N
salzi.at	R			tvmidtvest.dk	TV/R	stara.fi	N	leprogres.fr	R
sn.at	N					talouselama.fi	B	lequipe.fr	S
sport.orf.at	S					tilannehuone.fi	R	lesechos.fr	B
trend.at	B					tivi.fi	B	letelegramme.fr	R
tt.com	R					ts.fi	R	liberation.fr	N
tvheute.at	TV/R					uusisuomi.fi	N	lsa-conso.fr	N
vienna.at	R					verkkouutiset.fi	N	maxifoot.fr	S
vn.at	R					yle.fi	TV/R	mediapart.fr	N
vol.at	R							midilibre.fr	R
volksblatt.at	R							ouest-france.fr	R
wienerzeitung.at	N							parismatch.com	N
								rtl.fr	TV/R
								rugbyrama.fr	S
								sports.fr	S
								sudouest.fr	R
								tf1.fr	TV/R
								usinenouvelle.com	B
								zonebourse.com	B

Note: Outlets classification: N= National; R= Regional; B= Business; S= Sports; TV/R=Television.

Table 1B (cont 2). List of Domains per Country

Germany		Greece		Ireland		Italy		Netherlands	
Site	**Classif.**	**Site**	**Classif.**	**Site**	**Classif.**	**Site**	**Classif.**	**Site**	**Classif.**
3sat.de	TV/R	aek365.org	S	anglocelt.ie	R	adnkronos.com	N	ad.nl	N
abendblatt.de	R	agon.gr	R	balls.ie	S	affaritaliani.it	N	at5.nl	TV/R
ard.de	TV/R	alithia.gr	R	breakingnews.ie	N	agi.it	N	bd.nl	R
augsburger-allgemeine	R	alphatv.gr	TV/R	broadsheet.ie	N	ansa.it	N	bndestem.nl	R
autobild.de	B	antenna.gr	TV/R	businesspost.ie	B	calciomercato.com	S	businessinsider.nl	B
berliner-zeitung.de	R	avgi.gr	N	con-telegraph.ie	R	corriere.it	N	destentor.nl	R
bild.de	N	bankingnews.gr	B	connachttribune.ie	R	corrieredellosport.it	S	dvhn.nl	R
br.de	TV/R	capital.gr	B	donegaldaily.com	R	diretta.it	S	ed.nl	R
bz-berlin.de	R	contra.gr	S	dundalkdemocrat.ie	R	ecodibergamo.it	R	emerce.nl	B
computerbild.de	B	cretalive.gr	R	echolive.ie	R	fanpage.it	N	fd.nl	B
derwesten.de	R	dikaiologitika.gr	N	galwaydaily.com	R	finanzaonline.com	B	frontpage.fok.nl	N
deutsche-wirtschafts-n	B	dimokratiki.gr	R	herald.ie	N	gazzetta.it	S	geenstijl.nl	N
express.de	R	e-thessalia.gr	N	hoganstand.com	S	gds.it	R	gooieneemlander.nl	R
faz.net	N	ekathimerini.com	N	independent.ie	N	gelocal.it	R	gpupdate.net	S
finanzen.net	B	eleftheria.gr	R	irishexaminer.com	N	huffingtonpost.it	N	haarlemsdagblad.nl	R
finanzen100.de	B	ethnos.gr	N	irishmirror.ie	N	ilfattoquotidiano.it	N	iex.nl	B
finanznachrichten.de	B	euro2day.gr	B	irishrugby.ie	S	ilgazzettino.it	R	lc.nl	R
focus.de	N	filathlos.gr	S	irishtimes.com	N	ilgiornale.it	N	leidschdagblad.nl	R
fussball.de	S	fpress.gr	B	joe.ie	N	ilgiorno.it	R	limburger.nl	R
handelsblatt.com	B	gazzetta.gr	S	kilkennypeople.ie	R	ilmattino.it	R	metronieuws.nl	N
hna.de	R	iefimerida.gr	N	leinsterleader.ie	R	ilmessaggero.it	R	nhnieuws.nl	TV/R
jungefreiheit.de	N	in.gr	N	leitrimobserver.ie	R	ilmeteo.it	N	noordhollandsdagblad	R
kicker.de	S	kathimerini.gr	N	limerickleader.ie	R	ilpost.it	N	nos.nl	TV/R
ksta.de	R	kerdos.gr	B	longfordleader.ie	R	ilrestodelcarlino.it	R	nrc.nl	N
manager-magazin.de	B	makeleio.gr	N	mayonews.ie	R	ilsecoloxix.it	R	nu.nl	N
mopo.de	R	makthes.gr	R	meathchronicle.ie	R	ilsole24ore.com	B	parool.nl	R
morgenpost.de	R	naftemporiki.gr	B	politics.ie	N	ilsussidiario.net	N	pzc.nl	R
n-tv.de	TV/R	newmoney.gr	B	rte.ie	TV/R	iltempo.it	R	rd.nl	N
news.de	N	newpost.gr	N	tg4.ie	TV/R	internazionale.it	N	rijnmond.nl	TV/R
rp-online.de	R	news.google.gr	A	the42.ie	S	investireoggi.it	B	rtlnieuws.nl	TV/R
rtl.de	TV/R	news247.gr	N	thejournal.ie	N	la7.it	TV/R	rtvdrenthe.nl	TV/R
spiegel.de	N	newsbeast.gr	N	thesun.ie	N	lanazione.it	R	rtvnoord.nl	TV/R
sport.de	S	newsbomb.gr	N	tipperarylive.ie	R	lastampa.it	N	rtvoost.nl	TV/R
sport1.de	TV/R	newsit.gr	N	virginmediatelevision.i	TV/R	leggo.it	N	soccernews.nl	S
sportbild.bild.de	S	novasports.gr	S			libero.it	N	sprout.nl	B
sportschau.de	S	onsports.gr	S			liberoquotidiano.it	N	telegraaf.nl	N
spox.com	S	pelop.gr	R			milannews.it	S	trouw.nl	N
stern.de	N	pronews.gr	N			milanofinanza.it	B	tubantia.nl	R
sueddeutsche.de	R	protothema.gr	N			notizie.it	N	vi.nl	S
swr.de	TV/R	rizospastis.gr	N			palermotoday.it	R	voetbalprimeur.nl	S
tagesschau.de	TV/R	skai.gr	TV/R			panorama.it	N	voetbalzone.nl	S
tagesspiegel.de	N	sport-fm.gr	S			quifinanza.it	B	volkskrant.nl	N
taz.de	N	sport24.gr	S			quotidiano.net	N		
transfermarkt.de	S	sportdog.gr	S			rai.it	TV/R		
tz.de	R	stoxos.gr	N			rainews.it	TV/R		
welt.de	TV/R	tanea.gr	N			repubblica.it	N		
wiwo.de	B	thebest.gr	R			romatoday.it	R		
zdf.de	TV/R	tovima.gr	N			soldionline.it	B		
zeit.de	N	tvxs.gr	TV/R			today.it	N		
		zougla.gr	TV/R			transfermarkt.it	S		
						tuttomercatoweb.com	S		
						tuttosport.com	S		
						tv8.it	TV/R		
						unionesarda.it	R		

Note: Outlets classification: N= National; R= Regional; B= Business; S= Sports; TV/R=Television/R.

Table 1B (cont 3). List of Domains per Country

Poland		Portugal		Spain		Sweden		UK	
Site	**Classif.**	**Site**	**Classif.**	**Site**	**Classif.**	**Site**	**Classif.**	**Site**	**Classif.**
24kurier.pl	R	abola.pt	S	20minutos.es	N	affarsvarlden.se	B	bbc.com	TV/R
90minut.pl	S	aeiou.pt	N	abc.es	N	aftonbladet.se	N	belfasttelegraph.co.uk	R
bankier.pl	B	cmjornal.pt	N	antena3.com	TV/R	allehanda.se	R	channel4.com	TV/R
businessinsider.com.pl	B	dinheirovivo.pt	B	ara.cat	R	arbetarbladet.se	R	channel5.com	TV/R
dziennik.pl	N	dn.pt	N	as.com	S	bohuslaningen.se	R	chroniclelive.co.uk	R
dziennikbaltycki.pl	R	dnoticias.pt	N	bolsamania.com	B	corren.se	R	cityam.com	B
dziennikwschodni.pl	R	iol.pt	TV	cadenaser.com	R	di.se	B	coventrytelegraph.net	R
dziennikzachodni.pl	R	jm-madeira.pt	R	canalsur.es	TV/R	dn.se	N	dailymail.co.uk	N
echodnia.eu	R	jn.pt	N	canarias7.es	TV/R	expressen.se	N	dailyrecord.co.uk	N
expressilustrowany.pl	R	jornaldenegocios.pt	B	ccma.cat	TV/R	folkbladet.se	R	economist.com	B
fakt.pl	N	jornaleconomico.sapo.pt	B	cincodias.elpais.com	B	fotbollskanalen.se	S	edp24.co.uk	R
forbes.pl	B	n-tv.pt	TV	cope.es	TV/R	gp.se	N	express.co.uk	N
forsal.pl	B	noticiasaominuto.com	N	cuatro.com	TV/R	hn.se	R	expressandstar.com	R
gazeta.pl	N	observador.pt	N	diaridegirona.cat	R	idrottonline.se	S	ft.com	B
gazetakrakowska.pl	R	ojogo.pt	S	diariocordoba.com	R	jp.se	R	heraldscotland.com	R
gazetalubuska.pl	R	ominho.pt	R	diariodecadiz.es	R	kristianstadsbladet.se	R	huffingtonpost.co.uk	N
gazetaolsztynska.pl	R	omirante.pt	R	diariodemallorca.es	R	na.se	R	hulldailymail.co.uk	R
gazetawroclawska.pl	R	publico.pt	N	diariodenavarra.es	R	norran.se	R	independent.co.uk	N
gloswielkopolski.pl	R	record.pt	S	diariodesevilla.es	R	norrkoping.se	R	inews.co.uk	N
gol24.pl	S	rtp.pt	TV/R	diariosur.es	R	nwt.se	R	itv.com	TV/R
gp24.pl	R	sabado.pt	N	diariovasco.com	R	op.se	R	leicestermercury.co.uk	R
gs24.pl	R	sapo.pt	N	eitb.eus	TV/R	resume.se	B	liverpoolecho.co.uk	R
kurierlubelski.pl	R	sicnoticias.pt	TV	elcomercio.es	R	sla.se	R	manchestereveningnews.(	R
meczyki.pl	S	sicnoticias.sapo.pt	TV	elconfidencial.com	N	smp.se	R	metro.co.uk	N
money.pl	B	tsf.pt	R	elconfidencialdigital.con	N	svd.se	N	mirror.co.uk	N
natemat.pl	N	vidas.pt	N	elcorreo.com	R	svenskafans.com	S	pressandjournal.co.uk	R
newsweek.pl	N	zerozero.pt	S	eldiario.es	N	svt.se	TV/R	shropshirestar.com	R
niezalezna.pl	N			eldiariomontanes.es	R	sydsvenskan.se	N	skysports.com	S
nowiny24.pl	R			eleconomista.es	B	thelocal.se	N	sportinglife.com	S
nto.pl	R			elmundo.es	N	ttela.se	R	stokesentinel.co.uk	R
parkiet.com	B			elpais.com	N	tv4.se	TV/R	telegraph.co.uk	N
pb.pl	B			elperiodico.cat	R	tv4play.se	TV/R	theguardian.com	N
pomorska.pl	R			elperiodico.com	N	unt.se	R	thesun.co.uk	N
poranny.pl	R			elplural.com	N	va.se	B	thetimes.co.uk	N
przegladsportowy.pl	S			elpuntavui.cat	R	vf.se	R	uk.news.yahoo.com	A
rp.pl	N			europapress.es	N	viafree.se	TV/R	yorkshirepost.co.uk	R
se.pl	N			expansion.com	B	vlt.se	R		
sport.pl	S			heraldo.es	R				
stooq.pl	B			huffingtonpost.es	N				
telewizjarepublika.pl	TV			ideal.es	R				
tvn.pl	TV			lainformacion.com	B				
tvn24.pl	TV			laopiniondemalaga.es	R				
tvn24bis.pl	N			larazon.es	N				
tvp.info	TV			lasexta.com	TV/R				
tvp.pl	TV			lasprovincias.es	R				
weszlo.com	S			lavanguardia.com	N				
wpolityce.pl	N			laverdad.es	R				
wprost.pl	N			lavozdegalicia.es	R				
wspolczesna.pl	R			lavozdigital.es	R				
wyborcza.biz	B			levante	R				
wyborcza.pl	N			libertaddigital.com	N				
wykop.pl	N			lne.es	R				
				marca.com	S				
				mundodeportivo.com	S				
				naciodigital.cat	R				
				ondacero.es	TV/R				
				periodistadigital.com	N				
				publico.es	N				
				rtve.es	TV/R				
				sport.es	S				
				telecinco.es	TV/R				
				telemadrid.es	TV/R				
				ultimahora.es	R				
				vilaweb.cat	R				
				vozpopuli.com	N				

Note: Outlets classification: N= National; R= Regional; B= Business; S= Sports; TV/R=Television/Radio.

Table 2. Summary Statistics

Variable	Obs	Mean	Std. Dev.
Desktop Visits	676070	141479.5	257851.8
Mobile Visits	630212	288258.7	511417.2
Desktop Search Visits	674609	43466.61	79207.39
Desktop Direct Visits	675619	77498.82	164839.1
Keywords Top 1-100	653315	777894	1231113
Keywords Top 1-10	653315	88148.8	166258.9
Keywords Top 11-100	653315	689745.2	1081249
National	680641	0.298	0.457
Regional	680641	0.313	0.464
Sports	680641	0.109	0.312
Business	680641	0.116	0.320
Radio/TV	680641	0.131	0.337
Google Updates			
Core Updates +7	680641	0.049	0.217
Big Core Updates +7	680641	0.019	0.135
Not Big Core Updates +7	680641	0.031	0.174
Non Core Updates +7	680641	0.105	0.306
Concentration Measures			
HHI Mobile Visits	17117	916.9915	1127.473
HHI Desktop Visits	17117	1128.977	1077.18
HHI Search Visits	17117	831.3985	756.2952
HHI Mobile Visits per segment	96007	3490.983	2836.643
HHI Desktop Visits per segment	96007	3955.063	2720.061
HHI Search Visits per segment	96007	3543.197	2647.79

This table shows summary statistics of all variables used in our empirical analysis.

Table 3A. Impact of August 2018 and May 2020 Updates on the Number of Top Keywords

	(1)	(2)
Dependent Variable:	**Δln(Top Keywords)**	
Update August 2018 plus 7 days	0.00510***	
	(0.0009)	
Update May 2020 plus 7 days		-0.00289***
		(0.0003)
One Week After Update	0.00614***	0.00611***
	(0.0008)	(0.0004)
Constant	0.00319***	-0.0207***
	(0.0007)	(0.0008)
Observations	173,111	162,137
R-squared	0.02	0.23

All regressions include year, week and day-of-week fixed effects.
Robust standard errors clustered at the outlet/country level in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Table 3B. Impact of August 2018 and May 2020 Updates on the Number of Search Visits

	(1)	(2)	(3)	(4)
Dependent Variable:	**Δln(Search Visits)**			
update_aug_2018_plus7	0.00790**		0.0082	
	(0.0035)		(0.0058)	
update_may_2020_plus7		-0.0182***		-0.00229
		(0.0045)		(0.0061)
week_later	0.00822***	-0.00101	0.00822***	-0.00101
	(0.0031)	(0.0038)	(0.0031)	(0.0038)
update18*national			-0.00936	
			(0.0079)	
update18*regional			0.00809	
			(0.0084)	
update20*national				-0.0340***
				(0.0102)
update20*regional				-0.0184*
				(0.0106)
Δln(Direct Visits)	0.338***	0.343***	0.338***	0.343***
	(0.0302)	(0.0250)	(0.0302)	(0.0250)
Constant	0.0115	-0.0225***	0.0115	-0.0225***
	(0.0075)	(0.0057)	(0.0075)	(0.0057)
Observations	177,693	166,856	177,693	166,856
R-squared	0.25	0.28	0.25	0.28

All regressions include year, week and day-of-week fixed effects. Robust standard errors clustered at the outlet/country level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 4. First Differences OLS Regressions of Search Visits, Total Visits and Mobile Visits on the Number of Keywords and Google Core Updates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS	OLS	OLS	First Stage - IV	Second Stage -IV	Second Stage -IV	Second Stage -IV
Dependent Variable	**Δln(Search Visits) t-4**	**Δln(Desktop Visits) t-4**	**Δln(Mobile Visits) t-4**	**Δln(Keywords Top 10) t-4**	**Δln(Search Visits) t-4**	**Δln(Desktop Visits) t-4**	**Δln(Mobile Visits) t-4**
Δln(Words Top 10) t-4	0.0869***	0.0453***	0.1270***		6.6376***	3.7405***	4.1166***
	(0.0181)	(0.0122)	(0.0195)		(1.1211)	(0.7467)	(1.4006)
Big Google Core Update t+7				-0.0008*			
				(0.0004)			
Not Big Google Core Update t+7				-0.0025***			
				(0.0003)			
Δln(Desktop Direct Visits) t-4	0.3444***	0.6021***	0.3914***	0.0001	0.3436***	0.6017***	0.3910***
	(0.0257)	(0.0242)	(0.0263)	(0.0001)	(0.0258)	(0.0242)	(0.0264)
Constant	0.0332***	-0.0001	0.0997***	0.0075***	-0.0156	-0.0277***	0.0724***
	(0.0072)	(0.0050)	(0.0081)	(0.0008)	(0.0126)	(0.0079)	(0.0138)
Week FE	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES
Day of Week FE	YES	YES	YES	YES	YES	YES	YES
Observations	644469	645589	597962	645589	644463	645589	597962
R-squared	0.26	0.64	0.16	0.04		0.55	0.08

Robust standard errors clustered at the domain level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 5. Impact of Google Core Updates on Number of Search, Ddesktop and Mobile Visits per Country

	First Stage - IV		Second Stage - IV		
Dependent Variable	**Δln(Keywords Top10) t-4**		**Δln(Search Visits) t-4**	**Δln(Desktop Visits) t-4**	**Δln(Mobile Visits) t-4**
Coefficients of Interest	**β Big Core Update**	**β Not Big Core Update**	**β Δln(Keywords Top10) t-4**	**β Δln(Keywords Top10) t-4**	**β Δln(Keywords Top10) t-4**
Austria	0.0004	-0.0006	18.5679	-0.0515	1.1297
	(0.0012)	(0.0013)	(25.6457)	(10.2516)	(23.5141)
Belgium	-0.0006	-0.0046***	2.3189	2.7453**	5.1430*
	(0.0021)	(0.0011)	(1.6237)	(1.0280)	(2.5972)
Denmark	-0.0030*	-0.0042***	7.0638**	1.4451	
	(0.0016)	(0.0009)	(2.7906)	(1.8946)	
Finland	0.0014	-0.0027**	-3.2330	-4.1269**	-0.3457
	(0.0012)	(0.0011)	(1.9556)	(2.0029)	(5.0669)
France	-0.0034***	0.0019	-7.8110	-4.0528	-5.6614
	(0.0012)	(0.0032)	(5.6929)	(3.0022)	(4.8236)
Germany	-0.0005	-0.0014*	6.8452	5.7789	-1.4815
	(0.0012)	(0.0008)	(6.0689)	(3.8744)	(6.1417)
Greece	0.0085***	-0.0030***	-3.0628*	-2.3780**	-2.1888*
	(0.0020)	(0.0011)	(1.8023)	(1.0637)	(1.1690)
Ireland	-0.0074***	-0.0066***	-0.0312	-0.3740	1.1909
	(0.0021)	(0.0016)	(0.5429)	(0.8609)	(1.2613)
Italy	0.0002	-0.0001	46.9613	16.7190	-66.1106
	(0.0012)	(0.0009)	(204.4139)	(74.5345)	(293.4514)
Netherlands	-0.0034***	-0.0032***	0.7125	1.9045**	-3.0322
	(0.0012)	(0.0008)	(1.3368)	(0.8993)	(2.5607)
Poland	-0.0032***	-0.0034***	10.3908***	6.0095***	7.8332***
	(0.0010)	(0.0009)	(2.1928)	(1.5756)	(2.2501)
Portugal	0.0019	-0.0041***	2.6860	1.8328	6.1433
	(0.0018)	(0.0013)	(2.4840)	(1.5616)	(4.2944)
Spain	0.0013	-0.0047***	6.9571***	3.2958***	2.3239
	(0.0009)	(0.0006)	(1.0232)	(0.7012)	(1.4320)
Sweden	-0.0045***	-0.0008	0.5642	0.7596	-1.9122
	(0.0015)	(0.0008)	(1.3304)	(1.5677)	(2.3194)
United Kingdom	-0.0030**	-0.0024**	-4.4395*	-3.4895*	1.3677
	(0.0014)	(0.0009)	(2.3095)	(1.8522)	(3.2620)

This table contains results of 59 different regressions. For each country, we run first-stage regressions of first differences in log of number of keywords in top 10 positions on Big core updates and Not Big core update dummies. Then for each country, we run second stage using Google core updates as instruments for changes in the number of search visits, desktop visits and mobile visits. All specifications include week, year, day of the week FE and changes in the number of direct visits as controls.
Robust standard errors clustered at the news outlet level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 6. Impact of Google Core Updates on Number of Search, Desktop and Mobile Visits per News Outlet Type

	First Stage - IV		Second Stage - IV		
Dependent Variable	**Δln(Keywords Top10) t-4**		**Δln(Search Visits) t-4**	**Δln(Desktop Visits) t-4**	**Δln(Mobile Visits) t-4**
Coefficients of Interest	**β Big Core Update**	**β Not Big Core Update**	**β Δln(Keywords Top10) t-4**	**β Δln(Keywords Top10) t-4**	**β Δln(Keywords Top10) t-4**
Top Rank	0.0001	-0.0026***	5.0845***	2.9693***	2.9184*
	(0.0006)	(0.0005)	(1.4868)	(0.8491)	(1.6065)
Bot Rank	-0.0015***	-0.0024***	5.4388***	2.8425***	2.8790
	(0.0006)	(0.0004)	(1.3257)	(1.0117)	(1.8926)
Top Domestic %	-0.0013**	-0.0028***	5.7484***	3.4065***	5.9952***
	(0.0005)	(0.0004)	(1.2311)	(0.8461)	(1.7329)
Bot Domestic %	-0.0002	-0.0022***	7.1389***	4.0163***	0.4720
	(0.0006)	(0.0006)	(2.1115)	(1.4005)	(2.1136)
Top Google %	-0.0020	-0.0016	-7.0346	5.4355	-6.6891
	(0.0032)	(0.0021)	(6.3987)	(10.5390)	(13.8042)
Bot Google %	-0.0007*	-0.0025***	6.8018***	3.8415***	4.3679***
	(0.0004)	(0.0003)	(1.1433)	(0.7589)	(1.4259)

This table contains results of 24 different regressions. For each type of news outlet (top and bottom national rank, top and bottom domestic visit percentage, and top and bottom google visits %), we run first-stage regressions of first differences in log of number of keywords in top 10 positions on Big core updates and Not Big core update dummies. Then for each type of news outlet, we run second stage regressions using Google core updates as instruments for changes in the number of search visits, desktop visits and mobile visits. All specifications include week, year, day of the week FE and changes in the number of direct visits as controls.
Robust standard errors clustered at the news outlet level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 7. Impact of Google Core Updates on Number of Search, Desktop and Mobile visits per News Outlets Type

	First Stage - IV		Second Stage - IV		
Dependent Variable	**Δln(Keywords Top10) t-4**		**Δln(Search Visits) t-4**	**Δln(Desktop Visits) t-4**	**Δln(Mobile Visits) t-4**
Coefficients of Interest	**β Big Core Update**	**β Not Big Core Update**	**β Δln(Keywords Top10) t-4**	**β Δln(Keywords Top10) t-4**	**β Δln(Keywords Top10) t-4**
National	-0.0016**	-0.0028***	7.9725***	5.4994***	6.2104***
	(0.0007)	(0.0004)	(1.6319)	(1.0500)	(1.9389)
Regional	-0.0018***	-0.0037***	5.3915***	3.4021***	5.6872***
	(0.0006)	(0.0005)	(1.0139)	(0.7939)	(1.4540)
Business	0.0005	-0.0017**	12.4077**	3.9189	10.8587
	(0.0013)	(0.0007)	(5.8336)	(3.3803)	(8.3448)
Sports	0.0017	-0.0017**	-6.8526	-3.8886	-11.2317*
	(0.0014)	(0.0008)	(4.3187)	(2.5973)	(6.4024)
TV/Radio	0.0012	-0.0018**	10.9817**	5.2162**	7.8365
	(0.0012)	(0.0007)	(5.0367)	(2.5761)	(6.8450)

This table contains results of 20 different regressions. For each type of news outlet (national, regional, business, sports, TV/Radio), we run first-stage regressions of first differences in log of number of keywords in top 10 positions on Big core updates and Not Big core updates dummies. Then for each type of news outlets, we run second stage regressions using Google core updates as instruments for changes in the number of search visits, desktop visits and mobile visits. All specifications include week, year, day of the week FE and changes in the number of direct visits as controls. Robust standard errors clustered at the news outlet level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 8. Impact of Google Core Updates Two Weeks After

	(1)	(2)	(3)	(4)
Dependent Variable	**Δln(Keywords Top10) t-4**	**Δln(Search Visits) t-4**	**Δln(Desktop Visits) t-4**	**Δln(Mobile Visits) t-4**
Δln(Keywords Top 10) t-4		2.5431***	1.4547***	3.7283***
		(0.4055)	(0.2859)	(0.6376)
Google Core Update t to t+7	-0.0021***			
	(0.0003)			
Google Core Update t+8 to t+14	-0.0035***			
	(0.0003)			
Δln(Desktop Direct Visits) t-4	0.0001	0.3441***	0.6019***	0.3910***
	(0.0001)	(0.0258)	(0.0242)	(0.0264)
Constant	0.0074***	0.0149*	-0.0106*	0.0751***
	(0.0008)	(0.0084)	(0.0057)	(0.0102)
Week FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Day of Week FE	YES	YES	YES	YES
Observations	645.589	644.463	645.589	597.962
R-squared	0.04	0.22	0.63	0.10

Robust standard errors clustered at the news outlet level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 9. Paid Search Visits and Display Ads Visits

	(1)	(2)
Dependent Variable	**Δln(Paid Search Visits) t-4**	**Δln(Display Ads Visits) t-4**
Δln(Keywords Top10) t-4	0.0375	0.0723
	(0.0893)	(0.5421)
Δln(Desktop Direct Visits) t-4	0.0008*	0.0107
	(0.0004)	(0.0089)
Constant	-0.0004	0.0036
	(0.0010)	(0.0049)
Observations	644,463	405,805

Standard errors clustered at the news outlets level in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Table 10. Domain Authority and Core Updates

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	**Δln(Keywords Top10) t-4**	**Δln(Keywords Top10) t-4**	**Δln(Trust Flow) t-4**	**Δln(Domain Ranking) t-4**	**Δln(Trust Flow) t-4**	**Δln(Domain Ranking) t-4**
Google Core Update t+7	-0.0020**	-0.0042**	-0.0008	-0.00003	-0.0012*	-0.0001
	(0.00079)	(0.00194)	(0.0005)	(0.0001)	(0.0007)	(0.0001)
Trust Flow	-0.00002**					
	(0.00001)					
Update * Trust Flow	0.00000					
	(0.00001)					
Domain Ranking		-0.0001***				
		(0.00003)				
Update*Domain Ranking		0.00003				
		(0.00003)				
Google Core Update t+7 * National					0.0009	0.0001
					(0.0010)	(0.0001)
Google Core Update t+7 * Regional					0.0005	0.0001
					(0.0009)	(0.0001)
National					0.0000	0.0000
					(0.0001)	(0.0000)
Regional					0.0000	0.0001
					(0.0001)	(0.0001)
Δln(Desktop Direct Visits) t-4	0.00013	0.00010	-0.0012***	0.0000	-0.0012***	-0.00003
	(0.00010)	(0.00010)	(0.0004)	(0.0000)	(0.0004)	(0.00004)
Constant	0.0085***	0.0129***	0.0005	0.0006***	0.0005	0.0006***
	(0.00105)	(0.00219)	(0.0007)	(0.0001)	(0.0007)	(0.0001)
Observations	638,407	644,466	659390	662598	659390	662598
R-squared	0.05	0.04	0.002	0.001	0.002	0.001

Standard errors clustered at the news outlets level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 11. Impact of Core Updates on HHI of Search, Desktop and Mobile Visits

Dependent Variable	Δln(HHI Search Visits) t-4		Δln(HHI Desktop Visits) t-4		Δln(HHI Mobile Visits) t-4	
Coefficients of Interest	**β Big Core Update**	**β Not Big Core Update**	**β Big Core Update**	**β Not Big Core Update**	**β Big Core Update**	**β Not Big Core Update**
All	-0.0110*	0.0086**	-0.0014	0.0019	-0.0081	0.0021
	(0.0054)	(0.0029)	(0.0035)	(0.0023)	(0.0052)	(0.0026)
Austria	-0.0105	0.0150	-0.0126	0.0036	-0.0180	-0.0022
	(0.0174)	(0.0164)	(0.0101)	(0.0109)	(0.0165)	(0.0126)
Belgium	0.0170	0.0080	0.0081	0.0021	-0.0018	0.0043
	(0.0178)	(0.0100)	(0.0094)	(0.0077)	(0.0081)	(0.0058)
Denmark	-0.0024	-0.0003	0.0146	-0.0049	-	-
	(0.0394)	(0.0106)	(0.0176)	(0.0044)	-	-
Finland	-0.0186*	0.0142*	-0.0042	-0.0038	-0.0116*	-0.0074
	(0.0108)	(0.0082)	(0.0051)	(0.0038)	(0.0069)	(0.0056)
France	-0.0115	0.0025	-0.0054	-0.0041	-0.0146	0.0130
	(0.0077)	(0.0075)	(0.0051)	(0.0053)	(0.0113)	(0.0105)
Germany	-0.0477***	0.0103	-0.0040	0.0041	-0.0345**	0.0122
	(0.0131)	(0.0082)	(0.0057)	(0.0045)	(0.0138)	(0.0108)
Greece	-0.0433***	-0.0091	0.0096	0.0065	-0.0269***	0.0109
	(0.0147)	(0.0134)	(0.0078)	(0.0093)	(0.0101)	(0.0076)
Ireland	-0.0208	0.0096	-0.0223	-0.0045	-0.0015	-0.0076
	(0.0222)	(0.0172)	(0.0154)	(0.0121)	(0.0159)	(0.0120)
Italy	0.0043	0.0038	0.0237	0.0410**	0.0246	-0.0055
	(0.0126)	(0.0128)	(0.0218)	(0.0199)	(0.0160)	(0.0101)
Netherlands	0.0052	0.0287*	0.0059	0.0061	-0.0007	-0.0090
	(0.0146)	(0.0172)	(0.0069)	(0.0063)	(0.0084)	(0.0081)
Poland	-0.0041	0.0042	0.0044	0.0058	0.0235**	0.0074
	(0.0124)	(0.0098)	(0.0075)	(0.0057)	(0.0115)	(0.0114)
Portugal	0.0243*	-0.0060	0.0035	0.0074	0.0026	0.0146
	(0.0143)	(0.0129)	(0.0103)	(0.0074)	(0.0087)	(0.0098)
Spain	-0.0001	0.0164	-0.0009	-0.0084	-0.0164	0.0017
	(0.0111)	(0.0124)	(0.0119)	(0.0082)	(0.0133)	(0.0079)
Sweden	-0.0069	0.0264	0.0046	-0.0026	0.0111	0.0007
	(0.0148)	(0.0163)	(0.0100)	(0.0073)	(0.0089)	(0.0079)
United Kingdom	-0.0078	0.0169	-0.0056	0.0133	-0.0071	0.0175*
	(0.0094)	(0.0110)	(0.0079)	(0.0091)	(0.0095)	(0.0102)

This table shows results of 47 different regressions. The rows show the sample of countries used in each regression, all countries or each country individually. The big three columns show the result for each dependent variables, namely the first differences of logarithm of search visits, desktop visits and mobile visits 4 days apart. Within each dependent variable, we report the coefficient attached to Big and Not Big core updates. All regression specifications include first differences of the log of direct visits four days apart at the country level. * 0.1 significance, ** 0.05, *** 0.01.

Table 12. Impact of Core Updates on National Outlets Traffic at the Country Level

	(1)	(2)	(3)	(4)
Dependent Variables	**Δln(National Search Share) t-4**	**Δln(National Total Share) t-4**	**Δln(Ratio National/Regional Search Visits) t-4**	**Δln(Ratio National/Regional Total Visits) t-4**
Big Core Update	-0.0088	-0.0051**	-0.0140*	-0.0027
	(0.0056)	(0.0023)	(0.0067)	(0.0057)
Not Big Core Update	0.0042	-0.0025	0.0034	-0.0019
	(0.0059)	(0.0017)	(0.0060)	(0.0037)
Δln(Desktop Direct Visits) t-4	0.4241***	-0.0899***	0.5531***	0.2357*
	(0.1382)	(0.0290)	(0.1452)	(0.1155)
Constant	0.0412**	-0.0351***	0.0080	-0.0009
	(0.0179)	(0.0091)	(0.0166)	(0.0140)
Observations	16,845	16,965	16,845	16,845
R-squared	0.13	0.15	0.21	0.07

Note: The dependent variable in column (1) is the ratio of search traffic to all national outlets divided by the total amount of visits to all outlets at the country level. The dependent variable in column (2) is the ratio of all traffic to all national outlets divided by the total amount of visits to all outlets at the country level. The dependent variable in column (3) is the ratio of all search traffic to all national outlets divided by all search traffic to all regional outlets at the country level. The dependent variable in column (4) is the ratio of all traffic to all national outlets divided by all traffic to regional outlets at the country level. All dependent variables are first differences of the log between values four days apart. All specifications include year, month, day of week fixed effects.
Robust standard errors in parentheses clustered at the country level. *** p<0.01, ** p<0.05, * p<0.1

Table 13. Impact of Google Core Updates on HHI per Keyword in Spain

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep Var: ln(HHI traffic)								
Core Update	-0.0342***	0.0064	0.0042	0.0064	0.0346**	0.0034	-0.0106	0.0185**
	(0.0067)	(0.0066)	(0.0174)	(0.0050)	(0.0148)	(0.0072)	(0.0080)	(0.0090)
Core Update Plus 1	-0.0057	0.0023	0.0269	-0.0001	0.0334**	-0.0021	-0.0164**	0.0267***
	(0.0065)	(0.0067)	(0.0173)	(0.0051)	(0.0150)	(0.0072)	(0.0081)	(0.0092)
Core Update Plus 2	0.0310***	-0.0079	-0.0061	0.0095*	0.0301**	-0.0140*	0.0143*	-0.0023
	(0.0065)	(0.0067)	(0.0159)	(0.0050)	(0.0149)	(0.0072)	(0.0081)	(0.0090)
Big Core Update		-0.0406***	-0.0780***	-0.0070	-0.0190	-0.0502***	0.0664***	-0.0911***
		(0.0066)	(0.0172)	(0.0049)	(0.0139)	(0.0075)	(0.0078)	(0.0090)
Big Core Update Plus 1		-0.0080	-0.0276	0.0009	-0.0140	-0.0102	0.0868***	-0.0587***
		(0.0065)	(0.0171)	(0.0049)	(0.0140)	(0.0072)	(0.0078)	(0.0089)
Big Core Update Plus 2		0.0389***	0.0819***	0.0050	0.0005	0.0486***	0.0833***	0.0062
		(0.0063)	(0.0158)	(0.0049)	(0.0138)	(0.0070)	(0.0078)	(0.0084)
Constant	-0.7630***	-0.7630***	-1.2252***	-0.5945***	-0.8710***	-0.7302***	-0.6992***	-0.8409***
	(0.0035)	(0.0035)	(0.0078)	(0.0025)	(0.0064)	(0.0041)	(0.0045)	(0.0046)
Month Run Length FE	YES	YES	YES	YES	YES	YES	YES	YES
Keyword FE	YES	YES	YES	YES	YES	YES	YES	YES
Month-Year FE	YES	YES	YES	YES	YES	YES	YES	YES
Sample	All	All	> 4 sites	< 5 sites	> 8 months	< 9 months	Top10 Kword	Not Top10 Kword
Observations	3,644,630	3,644,630	1,003,707	2,640,923	945.559	2,699,071	1,863,610	1,781,020
R-squared	0.61	0.61	0.49	0.49	0.68	0.54	0.56	0.63

An observation is the HHI of traffic for each keyword in each month between november 2017 and november 2020 for our sample of 65 news outlets in Spain. The specifications of this table only include those keywords with visits and those keywords with more than one site at any given point during its length. Columns 3 and 4 separate our initial sample by keywords with more and less than 4 sites during their length. Columns 5 and 6 separate keywords by whether their run was longer or shorter than 8 months during our sample. Columns 7 and 8 separate keywords by whether they were ever Top 10 Keywords by any news outlet. Robust standard errors in parentheses clustered at the keyword level.

*** p<0.01, ** p<0.05, * p<0.1.

Table 14. Impact of Google Core Updates on Search Engine visits HHI

Panel A	**(1)**	**(2)**	**Panel B**	**(3)**	**(4)**
Dep Var: ln(HHI st)			**Dep Var: ln(Conc it)**		
Google * Core Update	-0.0109	-0.0228*	**Core Updatew**	0.0007	-0.0007
	(0.0083)	(0.0078)		(0.0008)	(0.0009)
Google * Core Update Plus 1	0.0014	0.0050	**Core Updatew Plus 1**	0.0005	0.0000
	(0.0033)	(0.0045)		(0.0006)	(0.0006)
Google * Core Update Plus 2	0.0289	0.0483**	**Core Updatew Plus 2**	0.0007	0.0011
	(0.0072)	(0.0093)		(0.0010)	(0.0013)
Google * Big Core Update		0.0264	**Big Core Update**		0.0065***
		(0.0188)			(0.0012)
Google * Big Core Update Plus 1		-0.0221	**Big Core Update Plus 1**		0.0043***
		(0.0203)			(0.0014)
Google * Big Core Update Plus 2		-0.0645**	**Big Core Update Plus 2**		0.0030
		(0.0113)			(0.0020)
Constant	6.3437***	6.3473***	**Constant**	9.1881***	9.1887***
	(0.0240)	(0.0239)		(0.0022)	(0.0022)
Search Engine FE	YES	YES	**News Outlet FE**	YES	YES
Month - Year FE	YES	YES	**Month - Year FE**	YES	YES
Observations	144	144	**Observations**	2334	2334
R-squared	0.44	0.45	**R-squared**	0.62	0.63

Panel A´s standard errors are clustered at the search engine level. Panel B's standard errors are clustered at the news outlet level.
*** p<0.01, ** p<0.05, * p<0.1.

Table A1. Differences Across Core Updates and Non-Core Updates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent Variables	**Δln(Keywords Top 100) t-4**	**Δln(Keywords Top 11-100) t-4**	**Δln(Keywords Top 100) t-4**	**Δln(Keywords Top 11-100) t-4**	**Δln(Keywords Top 100) t-4**	**Δln(Keywords Top 10) t-4**	**Δln(Keywords Top 11-100) t-4**
Big Google Core Update t+7	0.0017***	0.0017***	0.0019***	0.0020***			
	(0.0004)	(0.0004)	(0.0004)	(0.0004)			
Not Big Google Core Update t+7	-0.0009***	-0.0008***	-0.0011***	-0.0010***			
	(0.0003)	(0.0003)	(0.0003)	(0.0003)			
Core Update December 2020 t+7					0.0038***	0.0122	0.0029***
					(0.0015)	(0.0083)	(0.0010)
Core Update May 2020 t+7 #					-0.0010	-0.0080***	-0.0004
					(0.0008)	(0.0008)	(0.0008)
Core Update January 2020 t+7					-0.0039***	-0.0081***	-0.0037***
					(0.0005)	(0.0006)	(0.0005)
Core Update September 2019 t+7					0.0136***	0.0245***	0.0126***
					(0.0007)	(0.0008)	(0.0007)
Core Update June 2019 t+7 #					0.0071***	0.0040***	0.0074***
					(0.0005)	(0.0006)	(0.0006)
Core Update March 2019 t+7					-0.0103***	-0.0161***	-0.0101***
					(0.0006)	(0.0007)	(0.0007)
Core Update August 2018 t+7 #					0.0001	0.0028***	-0.0007
					(0.0008)	(0.0008)	(0.0007)
Core Update April 2018 t+7					-0.0012	-0.0019*	(0.0010)
					(0.0011)	(0.0010)	(0.0011)
Core Update March 2018 t+7					-0.0048***	-0.0131***	-0.0041***
					(0.0004)	(0.0005)	(0.0005)
Non-Core Google Update			-0.0022***	-0.0023***	-0.0025***	-0.0024***	-0.0026***
			(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Δln(Desktop Direct Visits) t-4	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Constant	0.0023***	0.0016**	0.0024***	0.0018**	0.0024***	0.0075***	0.0017**
	(0.0008)	(0.0008)	(0.0008)	(0.0008)	(0.0008)	(0.0008)	(0.0008)
Week FE	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES
Day of Week FE	YES	YES	YES	YES	YES	YES	YES
Observations	645597	645597	645597	645597	645597	645589	645597
R-squared	0.02	0.02	0.02	0.02	0.03	0.05	0.03

represent the Google Core Updates considered as Big by SEO experts.

Robust standard errors clustered at the domain level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A2. Second-Stage Regressions Using Number of Keywords in top 100 and top 11-100

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	**Δln(Search Visits) t-4**	**Δln(Desktop Visits) t-4**	**Δln(Mobile Visits) t-4**	**Δln(Search Visits) t-4**	**Δln(Desktop Visits) t-4**	**Δln(Mobile Visits) t-4**
Δln(Keywords Top 100) t-4	4.5169***	1.6302	1.7529			
	(1.4600)	(0.9994)	(1.4963)			
Δln(Keywords Top 11-100) t-4				3.8913***	1.2498	1.4390
				(1.4179)	(0.9965)	(1.4756)
Δln(Desktop Direct Visits) t-4	0.3439***	0.6019***	0.3911***	0.3440***	0.6019***	0.3912***
	(0.0257)	(0.0242)	(0.0263)	(0.0257)	(0.0242)	(0.0263)
Constant	0.0235***	-0.0035	0.0981***	0.0275***	-0.0018	0.0996***
	(0.0088)	(0.0056)	(0.0088)	(0.0082)	(0.0053)	(0.0084)
Week FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Day of Week FE	YES	YES	YES	YES	YES	YES
Observations	644469	645597	597968	644469	645597	597968
R-squared	0.13	0.63	0.15	0.16	0.63	0.15

First Stage of columns 1, 2 and 3 is column 1 in Table A1. First Stage of columns 4, 5 and 6 is column 2 in Table A1.
Robust standard errors clustered at the domain level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A3. Impact of Core Updates on HHI per News Outlet Type and Country

Dependent Variable	Δln(HHI Search Visits) t-4		Δln(HHI Desktop Visits) t-4		Δln(HHI Mobile Visits) t-4	
Coefficients of Interest	**β Big Core Update**	**β Not Big Core Update**	**β Big Core Update**	**β Not Big Core Update**	**β Big Core Update**	**β Not Big Core Update**
National	-0.0081	0.0095***	0.0005	0.0039**	-0.0036	0.0017
	(0.0047)	(0.0021)	(0.0024)	(0.0017)	(0.0040)	(0.0032)
Regional	-0.0055	0.0046	-0.0024	-0.0018	-0.0052	0.0085
	(0.0041)	(0.0028)	(0.0038)	(0.0027)	(0.0076)	(0.0082)
Business	0.0028	-0.0036	0.0062	0.0015	0.0061	-0.0069
	(0.0065)	(0.0050)	(0.0046)	(0.0045)	(0.0076)	(0.0062)
Sports	0.0008	-0.0067**	0.0077	0.0003	0.0027	-0.0008
	(0.0045)	(0.0026)	(0.0058)	(0.0028)	(0.0049)	(0.0048)
TV/Radio	0.0056	0.0064	0.0015	0.0007	0.0010	-0.0034
	(0.0064)	(0.0041)	(0.0048)	(0.0029)	(0.0061)	(0.0067)

This table contains results of 15 different regressions. For each type of news outlet (national, regional, business, sports, TV/Radio), we run first differences regressions of the changes in the log of HHI for search, desktop and mobile visits on Big Core Updates and Not Big Core Update dummies. All specifications include week, year, day of the week FE and changes in the number of direct visits as controls.
Robust standard errors clustered at the news outlets level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.