CESifo WORKING PAPERS

10671
2023

September 2023

Do Search Engines Increase Concentration in Media Markets?

Joan Calzada, Nestor Duch-Brown, Ricard Gil

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

Do Search Engines Increase Concentration in Media Markets?

Abstract

Search engines are important access channels to news content of traditional newspapers with Google alone responsible for 35% of online visits to news outlets in the European Union. Yet, the effects of Google Search on market competition and information diversity have received scant attention. Using daily traffic data for 606 news outlets from 15 European countries, we analyze Google's capacity to influence organic search visits by exploiting exogenous variation in news outlets' indexation caused by nine core algorithm updates rolled out by Google between 2018 and 2020. We find Google core updates overall reduced the number of keywords (queries) for which news outlets occupy one of the top 10 organic search results positions. Therefore, given the positive impact that the number of top keywords have on traffic this led to the decrease in the overall number of news outlets' visits. Finally, when studying the impact of Google core updates on media market concentration, we find the three "big" core updates identified in this period reduced market concentration by 1%, but this effect was offset by the rest of the updates. Similarly, in the context of Spain, we find the three "big" core updates reduced monthly keyword concentration by 4%.

JEL-Codes: D430, L500, L820, M310.

Keywords: search engines, market concentration Google, news sites, Europe.

Joan Calzada
Universitat de Barcelona / Spain
calzada@ub.edu

Nestor Duch-Brown
Joint Research Centre (JRC)
Seville / Spain
Nestor.DUCH-BROWN@ec.europa.eu

Ricard Gil
Smith School of Business, Queen's University
Kingston / Canada
ricard.gil@queensu.ca

September 2023
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), and the 38th EEA Meetings in Barcelona. 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.

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).[1]

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 network (12%), and visits from paid search and advertising display.[2] 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). Some exceptions are Sismeiro and Mahmood (2018), Cagé, Hervé and Mazoyer (2020) and Dujeancourt and Garz (2023), who examine how Facebook and Twitter have modified news production and readers' engagement. Germano et al. (2022) investigate how the modification of ranking algorithms to increase users' engagement in social networks might have detrimental effects in terms of misinformation and polarization.

In this paper, we investigate the capacity that search engines have to shape the media industry and how they can influence the way users discover and consume news online. Search engines are designed to provide users with the most accurate and useful results possible for their queries, and they are expected to design algorithms free of any bias

[1] 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).
[2] Own calculations, based on SimilarWeb data.

towards particular individuals, organizations, or viewpoints.[3] Despite this, search engines index news outlets according to their technological performance and the consumers' engagement, which is likely to bias search results in favor of outlets that are better at identifying the interest of large audiences and that have a higher domain reputation (Leung and Stumpf, 2023).[4] The purpose of our paper is to examine how Google Search's indexation activity affects the volume of organic search visits of news outlets, and ultimately, the concentration of news consumption in the European media markets. To the extent that changes in Google Search's indexation can significantly modify where news consumers obtain their information, ours is an empirical test of Google's media power, beyond its market power.

Google Search facilitates the discovery of unknown news outlets and has increased the variety of contents available for readers. However, the capacity of small and local news outlets to appear in top positions in Google's organic search results page depends on how Google trades-off the accuracy and proximity that can offer small and niche outlets and the expertise and authoritativeness of large and established outlets. Thus, the novelty of our paper is to examine whether the technological changes introduced in Google's core algorithms in the past few years have increased the visits to "superstar" news outlets, or whether they have thickened the "long tail" by giving more visibility to less popular, niche, and local newspapers.

Google is the ubiquitous search engine for desktop computers and mobile devices.[5] It uses bots to crawl news outlets pages and collect information about their contents. When a consumer has a query, Google uses algorithms to determine the order in which the links to the news pages appear in the search engine results pages (SERP hereafter). Google ranks news outlets pages according to two main criteria: the relevance of the contents for the query (dynamic ranking) and the authoritativeness of the news outlets (static ranking). A *dynamic ranking* is calculated at search time and depends on the search query, the user's location, the location of page, day, time, and query history, among others. A *static ranking* reflects features of the pages that are independent of the query (length of the page, frequency of keywords, number of images, compression ratio of text, among others), and it is calculated before the time of indexing (Chandra, Suaib, and Beg, 2015; Baye, de los Santos and Wildenbeest, 2016).[6] Thus, news outlets with a low static ranking (low domain authority) might find it difficult to obtain traffic for largely requested keywords, but they can rank high in specific queries that affect their region or their niche market. The success of news outlets in the organic search market depends on how well they rank relatively to

[3] 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).

[4] Brynjolfsson et al. (2010) describe non-technological drivers for the creation of "superstar" products such as increasing returns to quality and to advertising, or consumers' preferences for social interaction with other consumers.

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

[6] In addition, Google's top stories box shows up at the top of search results and presents a number of news articles relevant to the query. The algorithm reviews content automatically, looking for indicators of quality such as the number of clicks that it has attracted the trustworthiness of the publisher, the relevance of the story according to the reader's geographical location and the freshness.

their closest competitors, and more generally on how Google's algorithms weigh domain authority and content accuracy in queries that are relevant to them.

Our paper examines the impact of nine core algorithm updates rolled out by Google between 2018 and 2020. According to Google, these core updates are global, affect all Google search regions and languages, and do not focus on specific types of search queries or on particular website characteristics. The updates are designed to improve the way Google's system assess content, and to ensure that the service offers overall relevant and authoritative content to searchers (see Appendix Table A1 for a brief description of each of these core updates). We exploit these quasi-natural experiments to examine how changes in news outlets' indexation affect the organic search visits they obtain and the distribution of traffic across outlets.[7] Specifically, we analyze whether Google core algorithm updates have reinforced the skewness of the distribution of search traffic across news outlets, or if they have made the "long tail" thicker.

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 considers monthly data about the traffic generated by each search keyword in each news outlet in our sample. We complement these data with information from Ahrefs on the keywords (each keyword reflects a consumer query) that can generate search visits to the new outlets. These data show the daily number of keywords for which a news outlet occupies one of the first 10 positions (or 11-100 positions) in Google's organic search results.

The main contributions of the paper are twofold. First, we use a sound identification strategy to econometrically isolate the effects that Google's search algorithm has on the organic search traffic received by European news outlets, and thus, we provide a test of its media power. One important difficulty when studying how Google's indexation affects news outlets search traffic is that the visits to news outlets can be correlated with relevant, yet unobserved, news sites characteristics, or with the contents they publish. News sites compete for the keywords that generate more traffic and invest important resources to optimize their search results: they gather data on keyword volume and trends, keywords targeted by competitors, and search for combinations of keywords and phrases that increase their visits. For example, they can modify their headlines to maximize their audiences or as a response of the reaction of social networks (Cagé, Hervé and Mazoyer, 2020; Leung and Strumpf, 2023). To deal with this endogeneity problem, our paper adopts an instrumental variable identification strategy. Specifically, we use Google's core algorithm updates rolled out between 2018 and 2020 as an instrument for changes in the number of keywords that news outlets had on top organic search positions. The updates

[7] 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). Despite this, there are economic incentives to manipulate search engines listings, and search engines adapt their ranking algorithms continuously to mitigate the effect of spamming tactics on their search results (Chandra, Suaib, and Beg, 2015).

have a direct effect in news sites' indexation and are a source of exogenous variation for the outlets' search visits. To simplify the presentation of our results we divide the updates in two groups, the 3 biggest core updates according to SEO specialists, and the remaining 6 not big core updates.

Our results show that the core algorithm updates in our sample period had a negative causal effect on the number of queries (keywords) where news outlets occupy one of the top 10 organic search results positions. Overall, these findings imply that core updates reduced the visibility of news outlets in Google results pages, as they lost positions in organic search results. The magnitude of this effect is heterogeneous and robust across different types of news outlets (national, regional, business, sports, TV/radio), or when we group outlets according to different features (national rank, domestic traffic, traffic from Google). It is also important to note that *Big* and *Not Big* core updates exhibit different results across countries, although in most cases they have a negative effect on the number of keywords that news outlets have in top 10 positions.

The second contribution of the paper is to analyze the effect of Google' core algorithms updates beyond individual outlets, and onto the concentration of the media market across European countries. We find that the three *Big* core updates reduced market concentration by 1%, but that this effect was mostly compensated by a 0.8% increase generated by the rest of core updates. At the individual country level, the effect of the updates on the concentration of the search visits is heterogeneous. While they have reduced the concentration of the market in Finland, Germany, and Greece, they have increased it in Netherlands and Portugal. It is also interesting to note that Google's core updates have increased the market concentration among national generalist news outlets. In summary, we found evidence that Google's technological updates are not innocuous regarding competition in the media market, yet their influence is specific to each country and update.

Additionally, using monthly data only for Spanish news outlets, we examine how Google´s core updates affect concentration of search visits at the keyword level. Consistently with our market level results, we find that *Big* core updates reduced the concentration of the visits originated by each individual keyword by 4% within the month of the core update, followed by a quick increase in the concentration of the visits within two months after the core update. This result is driven by keywords for which at least 4 outlets occupy top 10 positions in Google's results page and with a run length shorter than nine months. Interestingly and in line with our site level analysis, *Big* core updates increased the by 6% the concentration in the visits originated by keywords for which Spanish outlets occupy a top 10 position in search results. This result implies that Google's core updates increased market concentration for the keywords that are more likely to generate more search traffic to the news outlets.

Finally, 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.

Our analysis and findings have important policy implications. First, our evidence quantifies Google's media power by showing how updates in its algorithm can significantly reduce outlets' visits and affect media market concentration. Second, we find that changes in Google Search`s indexation do not always lead to increases in media market concentration. Competition authorities worldwide are concerned about the current level of market and media power held by digital platforms that are based on indexation or recommendations algorithms. It is unclear which biases these platforms can introduce in their activities and how they can affect competition. Google Search has been subject to intense antitrust scrutiny from the US and European competition authorities (Yun, 2018). At the beginning of the 2010s, the U.S. Federal Trade Commission (FTC) investigated several antitrust allegations including the use of bias in search results, but the FTC ultimately closed its investigation. In 2015, the European Commission (EC) also investigated Google alleging search bias, and in 2017, the EC fined Google $2.7 billion for abuse of dominance in Google Shopping (Scott, 2017). According to the European Commission (2017), Google had abused its market dominance as a search engine by giving an illegal advantage to its own comparison-shopping service. Specifically, Google's comparison-shopping results were placed above Google's generic search results, and this allegedly diverted traffic from its competitors to Google. The EC found that none of the alternative sources of traffic available to competitors could effectively replace the generic search traffic from Google.

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. As such, it is important to understand the effects that search engines and new aggregators have on the shaping of media markets. Our findings constitute a first step in that direction.

The article is structured as follows. Section 2 reviews the literature closely related to our paper. Section 3 describes the main features of Google Search and explains how Google updates its indexation algorithms. Section 4 presents the data and our empirical strategy. Section 5 examines the impact of Google's core algorithm updates on the number of search, desktop, and mobile visits of European news outlets. Section 6 analyzes the effect of Google's core updates on media market concentration for organic search visits and overall visits. Section 7 examines the impact of core updates on market concentration 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 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) analyze biases in search results when search engines are vertically integrated with a seller. De Cornière and Taylor (2014) show that the integrated search engine can bias its search results to favor its own website and obtain more ad revenues. However, the search engine can also benefit by offering high quality search results that increase customers' participation and its ad revenues. Burguet et al. (2015) examine a search engine that provide consumers with organic search results that help them find online content providers, and sponsored search results that facilitate interaction with merchants selling offline products. The engine's organic search service attracts consumers who then can use the sponsored search results, but the engine and the content providers compete in the advertising market. The model shows that the integration of the engine with a fraction of content providers internalizes vertical externalities and increases organic and sponsored reliability, but it affects horizontal competition and can reduce social welfare.

Other papers have shown that search engines may strategically degrade the quality of their search results. Chen and He (2011) and Eliaz and Spiegler (2011) show that they can reduce their quality to relax sellers' competition and extract higher profits. 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 examined the incentives of digital platforms to bias results in "recommendation systems." Bourreau and Gaudin, (2022) examine a monopoly streaming platform that offers access to differentiated content providers. The platform can use recommendations to reduce the market power of content providers, and hence be able to set higher fees to consumers. Bourreau et al. (2021) consider a model where content providers can offer to a platform data (rather than money) about their consumers to obtain a prominent position in search results. Drugov and Jeon (2017) examine a vertically integrated platform that can bias recommendations towards its own content when consumers' utility in the long-run is shaped by their short-run usage. Calvano et al. (2023) analyze algorithmic recommendations on product market competition, showing that they lead to higher market concentration and prices. They also examine the potential for manipulation of recommendations and its impact on competition.

Our paper contributes to the empirical literature examining the existence of 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. In 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. They find that hotels ranking position are lower when they are also announced in a rival platform, at a lower price. 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 independent-label music as well as music by female artists. Huang and Xie (2023) consider platforms that favor some group of sellers by offering consumers repetitive information about their products. Using data from food delivery platforms, they show this behavior generate higher average prices and more skewed revenue distributions for sellers.

Our work is also related to the empirical 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).[8] There is ambiguous evidence that recommendation systems favor products in the long tail and encourage sellers' participation, as these products become more accessible for niche consumers. Oestreicher-Singer and Sundararajan (2012) collect information on the co-purchase links shown in Amazon when consumers look at a particular book (links on titles that other consumers bought together with each book). They explain that co-purchase links triple the influence that complementary books have on each other's demand. In addition, consistent with the theory of the long tail, niche books perform better, and popular books perform relatively worse in book categories where recommendations are more important. Hosanagar et al. (2014) use data from an online music service to examine whether recommender systems fragment users. They obtain that a network of users becomes more homogeneous after the introduction of a recommendation system. 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, this paper also contributes to the literature that investigates the role of media in the provision of information to the public and the shaping of political outcomes. 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; Gentzkow and Shapiro, 2010; Gentzkow and Shapiro, 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 contrasts with this literature in that we show how algorithm-based platforms can affect news outlets' market and media power (Prat, 2018).

[8] A related stream of the literature analyzes the discriminatory effects of the algorithms that automate decision-making. Cowgill and Tucker (2019) survey the theoretical and empirical literature examining algorithmic bias and fairness. Sweeney (2013) and Datta et al. (2015) study algorithm discrimination in advertising. Lambrecht and Tucker (2019) investigate in a field experiment the bias introduced by an algorithm that delivered ads promoting job opportunities. The advertisement campaign was intended to be gender-neutral, but the ad was shown to over 20% more men than women. The explanation is that younger women are a prized demographic group, and it is more expensive to target them.

3. Google's search algorithm

Search engines such as Google, Bing, and Yahoo use bots to crawl web pages, going from site to site, collecting information about the pages and indexing them. Googlebot is the robot of Google that crawls accessible web pages, sees and classifies their content, and indexes them. When consumers have a query, Google use its algorithms to rank pages and show the results in the Search Engine Results Page (SERP). Organic search results are ranked by their relevance to the query according to the algorithms criteria. The algorithms consider the match of the contents with the information needed by the consumer and other aspects, such as page-speed, use of unique images, inclusion of original and updated contents, the language, or the number of links targeting at the website. The SERP also shows sponsored results, which are paid by firms that want to attract to their websites consumers making specific keyword queries.

Google ranks web pages according to the EEAT criteria, which consider their *Experience, Expertise*, *Authoritativeness* and *Trustworthiness*. Specifically, 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.[9] Google explained the relevance of these aspects in 2011, after rolling out the "Panda update" of its algorithm.[10] Furthermore, in 2015 Google published its EEAT guidelines to explain to its human evaluators[11] how to assess web pages. The information raised by these evaluations is used to assess the performance of its algorithms.[12]

According to Google's guidelines, websites and pages with the primary objective to provide valuable assistance to users are considered high-quality webpages. High-quality pages not only need to fulfill their intended purpose, but they should also prioritize a user-centered approach. Google pays special attention to "Your Money or Your Life" (YMYL) web content. YMYL pages (or topics) are those 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 content about important topics such as international events, business, politics, science, and technology. Despite this, not all news articles are considered YMYL. For example, sports, entertainment, and everyday lifestyle topics are generally not YMYL. In its guidelines, Google asks its 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 websites to increase their visibility in search engines and increase their visits. SEO is a fundamental part of digital marketing because search engines are an essential distribution channel for

[9] https://www.pi-datametrics.com/blog/google-core-update-december-2020/

[10] https://developers.google.com/search/blog/2011/05/more-guidance-on-building-high-quality

[11] 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. Google uses rater feedback and other input data to shape relevant algorithms. Danny Sullivan, Public Liason for Google Search. See https://www.pi-datametrics.com/blog/google-e-a-t-ultimate-guide/

[12] A description of the new EEAT criteria can be found in the Google's guidelines for its reviewers: https://developers.google.com/search/blog/2022/12/google-raters-guidelines-e-e-a-t

many firms. Google does not share externally any scoring or indexing criteria that can help websites or advertisers to attract visitors. However, SEO companies apply reverse engineering to identify the factors used by Google to index websites and have created metrics that try to approximate the ranking or "domain authority" of websites.[13]

3.1 Google's Core Updates

Google continuously updates its search algorithms and systems to improve the quality of search results. Most of these changes are unnoticeable. However, a few times a year Google makes large "core updates" that generate significant modifications in the way it indexes and ranks websites. These updates can have a significant impact on traffic. According to the firm, these changes "*are designed to ensure that overall, we're delivering on our mission to present relevant and authoritative content to searchers.*"[14]

The rollout of core updates is global, affects all Google search regions and languages, and it is not focused on specific types of search queries or on particular websites characteristics. However, updates generate fluctuations in search rankings throughout the next days and weeks after their adoption and can affect different types of websites in different ways. In recent years, Google has started to publicly announce its core algorithm updates to prevent speculation about their release. However, the firm is reluctant to provide information about the changes introduced. Figure 1 shows as an example the announcement on twitter of Google's May 4th, 2020, core update, and some of the immediate reactions of small newspapers. Moreover, Table 1A presents the nine core updates confirmed by Google during the period 2018-2020, which are the updates analyzed in our empirical analysis.

Traffic recovery can be extremely challenging after a core update. According to Google, there is nothing website owners can do to recover their search traffic after the updates.[15] Indeed, the Google's algorithm is designed to avoid spam and unorthodox methods that could be used by the sites to get their content at the top of the ranking. At the same time, Google offers advice and guidelines to webmasters on how to design and organize their pages to improve search results.[16]

It is often difficult to understand the causes driving Google's updates. Some years ago, it could have been possible to identify some of the changes introduced by the firm and to

[13] The concept of "domain authority" or "domain trust" is based on the concept "PageRank" developed at the end of the nineties within one of Google's search patents. The "PageRank" aims at describing the website's authority on a topic and it is used, among other aspects, to rank webpages after the query of a consumer. It reflects the number and quality of links to a page.

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

[15] "*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*…. https://t.co/ohdP8vDatr (Google SearchLiaison (@searchliaison) Oct. 11, 2018). See https://blog.searchmetrics.com/us/google-update-november-2019/

[16] Google Webmaster Blog suggests different actions that can be adopted after being affected by Core Updates (https://developers.google.com/search/blog/2019/08/core-updates). Google also publishes their "Webmaster Guidelines", showing how they index and rank web sites, and outlines some of the illicit practices that may lead to a site being removed entirely from the Google index (https://developers.google.com/search/docs/advanced/guidelines/webmaster-guidelines).

determine which sites won and lost with them. Nowadays, there are multiple hypotheses to explain the rollout of different updates. One explanation is that Google is changing the level of EEAT that a website needs to rank in the top SERP positions. Another one is that Google uses new technologies such as "Natural Language Processing" to understand whether the content is relevant and trustworthy to rank high. As Google improves their ability to understand language, they get better at understanding which contents are truly helpful and authoritative. In doing so, they may be able to put less emphasis on PageRank when it comes to determining authoritativeness.[17] Changes in the consumers' search interest might also affect the rank of a website. For example, video content is becoming increasingly important, and this requires machine learning to match queries with relevant content.

3.2 General updates of search algorithms

In addition to the core updates, Google regularly launches other specific algorithm changes that can affect many sites. In November 2016, Google modified the method for crawling websites and launched its Mobile-First Index, which means Google predominantly uses the mobile version of the content for indexing and ranking. Historically, Google primarily used the desktop version of a page's content when evaluating the relevance of a page to a user's query. However, as nowadays most users make their search with a mobile device, Googlebot primarily crawls and indexes the mobile version of web pages. Google announced the roll out of the Mobile-First Index on March 26, 2018 and completed its deployment in 2021. Considering this gradual migration, it will be difficult to assess how this specific update affected news outlets indexation and search traffic.

In addition to these changes, every day Google releases one or more changes to its algorithm to improve the search results for consumers and to correct different types of bugs.[18] Thus, for example, Google can correct indexing and canonical bugs. If a site owner decides to syndicate content (they allow their content to be republished on another site), then canonical tags are used to show search engines whether a URL is the original content page. This helps the site that originally provided the content to still rank in the SERPs when its content is reproduced elsewhere. Some changes in the algorithm of Google are used to fix incidences with the indexing or the canonical tags. Thus, for example, Google confirmed this type of adjustments on August 10, September 29 and October 12, 2020. Another example of an update is Google's introduction of "passage indexing" on February 10, 2021, to index specific passages, not just the overall page. Google considered that passage-based indexing can affect 7% of search queries across all languages.

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

[18] Some changes in the algorithms might not be announced. In some occasions, rank tracking tools and webmaster chatter suggest the existence of unconfirmed updates by Google (phantom updates), although these can be temporary and disappear after some days or weeks. See the previous footnote.

Another more recent change has been the inclusion of the BERT algorithm (Bidirectional Encoder Representations from Transformers). This is a neural network-based technique for natural language processing pre-training. It helps Google to better discern the context of words in search queries and to offer results that are more accurate. BERT began rolling out in Google's search system on October 22, 2019, for English-language queries, including featured snippets. On December 9, 2019, Google confirmed that the BERT algorithm was rolling out internationally, in 70 languages. The firm considers that BERT can affect 10% of searches.

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 covers the period from October 1, 2017, to December 31, 2020, which includes the 9 Google core updates examined in the paper.

To examine the effect of Google core updates on news outlets search traffic, we consider 606 news outlets from the following 15 European countries[19]: 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. We have selected the news outlets in our sample considering the national rankings published by Alexa (www.alexa.com) and SimilarWeb (www.similarweb.com) as well as several websites and sources specialized in the media market. We also picked top rated news outlets and webpages from TV and radio stations that offer news contents for every country. Our dataset is restricted to news sites with more than 5000 daily visits because SimilarWeb does not provide information from domains with lower traffic levels. The data 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 start 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 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. We do not use daily data on the mobile organic search visits because SimilarWeb only collects such information 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

[19] In parenthesis, the number of news outlets in the corresponding country.

variable of interest) and other sources such as direct visits, visits from social media, paid search, and display ads. The same applies for *Daily Total Mobile Visits*.

Our analysis also considers 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. 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 visits up or down. 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 core algorithm 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 desktop traffic, considering the percentage of direct, organic search and social networks traffic. Our analysis also considers the *Daily Desktop Paid Visits* and the *Daily Desktop Ads Visits*, which are respectively the traffic sent to the news outlets from paid search ads on search engines,[20] and the traffic sent from display and video ads via an ad-serving platform (i.e., GDN, Doubleclick). We use these two variables in section 5.3 to examine whether news outlets react to Google Search core updates by increasing their advertising effort in Google Search and in other websites.

When exploring outlet heterogeneity, we classify news outlets in our sample according to different criteria. First, we consider their specialization, which can be *National, Regional, Business, Sports* or *TV/Radio*. To make this classification we have searched for verbal descriptions in several sources such as Alexa, SimilarWeb and Wikipedia. Second, we divide news outlets according to their national rank. Specifically, we distinguish between *Top Rank* and *Bottom Rank* news sites, considering if their national rank is above or below the median in their own respective country. Third, we classify domains according to the percentage of visits they receive from other countries. *Top Domestic* and *Bottom Domestic* separate news outlets into two groups according to whether their share of domestic visits is above or below the median in their own respective country. Fourth, news sites are classified considering the percentage of the total search visits originated in Google Search. Thus, we distinguish between *Top Google* and *Bottom Google* news outlets, considering whether the search traffic from Google is above or below the median in their own respective country.

SimilarWeb also collects information on the *monthly traffic* generated by each keyword considered in Google Search's queries. We use this information in section 7.1 to measure the impact of Google core updates on the concentration of the search traffic at the keyword level. Specifically, using Spain as an example, we estimate the effect of the core updates on the concentration of the traffic generated by more than 100,000 keywords. These data show the number of months each keyword has been active, and the frequency

[20] 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).

news outlets have been ranked in top 10 organic search results by Google for each keyword.

Our dataset includes several measures of 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.[21] According to Ahrefs, this metric works in a similar way to the original PageRank calculation, [22] although it ranks websites and not web pages.[23] 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.

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 .[24] 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. It is important to mention 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. It is also important to note that Ahrefs organic keywords metric is a country-specific variable. 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 variable *Words Top 100,* which is the number of keywords (queries) that news outlets have in the top 100 organic search results in Google, and the variable *Words Top 11-100,* which reflect the number keywords that news outlets have in the top 11-100 organic search results in Google. According to industry reports, the results in the first page of Google Search capture around 71% of search traffic clicks, and the results in the second capture less than 5.5% of the clicks.[25] 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 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

[21] See https://ahrefs.com/blog/seo-metrics/#section7 and https://ahrefs.com/blog/domain-rating/ for more information.

[22] Google' PageRank (PR) is no longer a quality metric to assess websites. Although the firm has said that it still uses it internally in its web positioning algorithm, it stopped updating it in 2013. One problem with the PR was that it only considered its own metric, and it was relatively easy to increase the PR of a domain by buying sponsored articles, commenting on blogs, or getting links on high PR sites.

[23] See https://ahrefs.com/blog/google-pagerank/

[24] See https://ahrefs.com/blog/seo-metrics/#section6

[25] See https://moz.com/blog/google-organic-click-through-rates-in-2014

first search result page. The first search result for a news outlet is for CNBC, in the sixth position. Previous results are for Wikipedia and for several institutional sites. In this example, Google's first results page includes "zero-click searches", which are answers to queries that do not send consumers to third-party websites. Google uses its *Direct Answer Box* to offer answers to many consumers' queries, such as queries for celebrities, geography, or history. Search queries about the weather or stock market prices are also answered directly by Google. 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 explained, Google's algorithms index news outlets, thereby determining the position of these outlets in the search results pages when users perform queries. The higher news outlets rank in the queries the higher the probability that users will click-through their links and generate visits.[26] This means that we should observe an empirical relationship between the search visits of news outlets and the number of keywords these have in top 10 search results. Therefore, our baseline specification is as follows

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

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 site *i* in day *t*, and $\ln[Words\ Top\ 10_{it}]$ is the natural logarithm of the number of keywords that the news site *i* has in the top 10 organic search results in day *t*. Moreover, X_{it} is a set of variables varying across news sites and days, and α_i and δ_t are news site and day fixed effects respectively. The usual iid assumptions apply to the error term u_{it}.

To account for unobserved heterogeneity at the news site 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

[26] Varian (2007) shows there is a positive relationship between the position of an ad and the number of clicks on that ad in the context of position auctions. Decarolis et al. (2023) examine whether search engines that sell ad slots through search auctions can benefit from obfuscating the data available to the AI bidding algorithms.

variable with realizations of the variable four days before such that $\Delta\ln[y_{it}] = \ln[y_{it}] - \ln[y_{it-4}]$.[27] We assume that $\text{cov}(\Delta\ln[Words\ Top_{it}], \Delta u_{it}) = 0$ to grant identification of the coefficient of interest β.

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.[28]

To address this endogeneity problem, we pursue an instrumental variable 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 has no effect on the outlet's search visits or in the probability that readers will search for the outlet content, other than indirectly through the number of keywords in top positions. The instrument that we use for this objective is the occurrence of Google's core updates, which directly modifies the news outlets' indexation for each consumer´s query, and they are a source of exogenous variation for the news outlets' visits. We estimate an IV model where the second stage is specification (2) above, and where the first stage is 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,[29] 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. Any other day in our sample outside of those nine seven-day treatment periods are our control period. It is important to note that these are global core updates. This means that all outlets in our sample are impacted by the updates, albeit in different ways.

Our analysis considers the 9 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

[27] We have also examined this specification using 7-day gaps and we find qualitatively and statistically equivalent results. In the end, we chose to implement our empirical strategy of 4-day gaps to capture immediate results after Google Search core updates were announced.

[28] Baye, de los Santos and Wildenbeest (2016) show that there are returns to SEO efforts that allow search engines to determine a site's relevance for a particular product search. These types of efforts include making effective use of anchor texts, descriptive headings, and meta tags, robot.txt files, and using accurate and unique page titles. Hagar and Diakopoulus (2019) describe the use of A/B headline testing to increase traffic to stories. Leung and Strumpf (2023) show how the New York Time and the Wall Street Journal modify their headlines and the impact of these changes on the article's performance metrics.

[29] Our first stage results are qualitatively equivalent when using 4-day and 14-day windows for the Core Update dummy.

$\text{cov}(Core\_update\_plus7, \Delta u_{it}) = 0$. This means that Google core updates are orthogonal to changes in visits (search or total) to a news site i, and they are not launched following an 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 changes in the traffic to the specific websites we analyze.

It is important to acknowledge that some outlets might have SEO teams or hire external SEO consultants to help them react to the updates of the algorithms. These SEO services cannot predict the date in which a core update will be released or the exact nature of the update. SEO services can forecast what core updates may look like in the future,[30] but even so, adjusting to new algorithms take time, ranging between two weeks and six months.[31][32] For most accounts, "accurately forecasting SEO performance is almost impossible due to the ever-changing search engine algorithms, intense competition, and unpredictable user behavior."[33] Therefore, core updates come as a surprise when they are rolled out and outlets need time to react to them. In our analysis below, we show the heterogenous results of core algorithm updates across different types of outlets, which can reflect differences in the level of expertise of their SEO services.

5. Results

5.1 Main Results

This section analyses the effects of Google's indexation algorithms on the search visits of European news outlets. For the implementation of our empirical strategy in section 4, we consider the nine algorithm core updates confirmed by Google in the period from October 1st, 2017, to December 31st, 2020. The aim of these updates is to fix different aspects of the indexing algorithms and to introduce new features that improve search accuracy. Table 1A lists the updates used in our analysis. All the updates are unique in their purpose and impact on online traffic, but to simplify the exposition of our results we divide them in two groups, the 3 *biggest* core updates according to SEO specialists, and the remaining 6 *not big* core updates.[34]

Columns 1, 2 and 3 of Table 3 show results of running specification (2) without instrumenting. 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

[30] See https://capforge.com/seo-forecasting-why-its-important/.
[31] See https://www.forbes.com/sites/joshsteimle/2015/02/07/how-long-does-seo-take-to-start-working/?sh=17a7daad464c
[32] See https://twitter.com/jeremiahcsmith/status/1660359928106860544.
[33] See https://www.linkedin.com/pulse/can-you-guarantee-1-rankings-truth-forecasting-seo-results-1e/.
[34] 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.

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, an increase of 0.05% in desktop visits, and an increase of 0.13% in mobile visits.

As explained in the previous section, a potential limitation of the previous analysis is that news outlets can use keywords and phrases strategically to increase their visits. For example, they can repeat multiple times the use of specific keywords in the headlines and in their news stories to rank higher in the results for frequent queries. To deal with this endogeneity problem, we pursue the instrumental variable identification strategy in equation (3), using Google's core algorithm updates as an instrument. This strategy leverages plausibly exogenous variation in the number of keywords ranking in the top 10 search results during the 9 core updates confirmed by Google between October 2017 and December 2020.

Using this identification strategy, column 4 in Table 3 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 3 show the results for the two-stage least squares (2SLS) 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*.[35] The results of the second stage of the IV 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 the IV coefficients are significantly larger than the OLS coefficients reported in columns 1 to 3. One possible explanation for the attenuation bias in the OLS estimates would be the incidence of classical measurement error in search results. This is not likely to be the driving explanation as there is no clear source of such measurement error in the data. A second potential explanation would be the heterogeneity in the impact of the IV. Our analysis about the heterogenous effects of the updates across different types of outlets will document that such heterogeneity exists and it is likely to be a driving explanation for our OLS estimates being smaller than our IV estimates.

So far, our analysis has classified core updates as either *Big* or *Not Big*, overlooking the nuanced differences that might exist between updates in each group. Appendix Table A1 analyzes the impact of these updates in the number of top 100 keywords and the number

[35] 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 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%.

of top 11-100 keywords as well as the impact of each individual update on the number of keywords in top 100, top 10, and top 11-100, respectively. Columns 5, 6 and 7 show that each individual core update has its own idiosyncratic effect (positive or negative) on the number of keywords in top search positions. Notice that the update with a largest impact in the news market was rolled out in March 2019 (not considered a big update by industry specialists). The negative effect of this update was later compensated with the positive effect of the updates of June 2019 and September 2019. In 2020, the updates of January and May had a negative effect in the number of keywords, in contrast to the update of December 2020. To sum up, our analyses reveal that core updates had different effects on the number of keywords ranked in top positions for each news outlet, and that the individual effects were similar for the number keywords in top 10 and top 11-100 positions. Table A1 also presents the effects of other updates that have been confirmed by Google, which are classified as "non-core updates". Overall, the results show that non-core updates had a negative and significant effect on the number of keywords for which outlets ranked in top search positions.[36]

In summary, two main conclusions follow from the instrumental variable estimations. First, Google core algorithm updates have a significant effect in the number of keywords that news outlets have in top search results. The core updates rolled out in the 2018-2020 period 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. Second, the number of keywords that news outlets have in the top 10 search results pages have a positive effect in news outlets' search visits.

5.2. Heterogeneous Impact of Google Core Updates

We next investigate the heterogeneity of the effects of Google's core algorithm updates across national markets and different types of outlets. Table 4 repeats the IV estimations of Table 3 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 4 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 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

[36] Appendix Table A2 shows the results of the two-stage least squares (2SLS) instrumental variable estimation of the model in equation (3), using the variables *Big* and *Not Big Core Update Plus 7* as instruments for the variables *Words top* 100 and *Words top* 11-100. The results show that the number of top 100 and top 11-100 keywords have a positive and significant impact in the *Desktop Search Visits*. Results for *Desktop Total Visits* and *Mobile Total Visits* are not significant.

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 5 and 6 repeat the previous analysis with different sample splits. In Table 5, 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. 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 6 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. 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*).

5.3. Do News Outlets React to the Updates?

Once we have established that the overall impact of Google´s core updates has been a decrease in the number of search visits received by news outlets, primarily due to a reduction in the number of keywords for which news outlets have a top position in organic search results, next we examine whether news outlets react 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.[37] 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.[38] To the best of our knowledge, there is not a common understanding of how fast SEO may help outlets reestablish traffic with estimates ranging between two weeks and six months.

[37] 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/).
[38] SEO forecasting may speed up reaction times when SE updates occur, but it is still hard to predict their exact timing and nature beforehand (https://www.forbes.com/sites/joshsteimle/2015/02/07/how-long-does-seo-take-to-start-working/?sh=17a7daad464c).

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 7 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 7 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 8 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 3. 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. It is important to note this does not mean news outlets do not react at all, it only means their actions have no impact through these channels 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). Considering this, it is interesting to empirically 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 9 show that a high value of these variables reduces the change in the number of keywords that news outlets have in top 10 search results. That is, news outlets with a high domain authority are less affected by the loss of visibility in Google's search results pages. However, we do not find that core updates had a smaller effect in outlets with a high domain authority. Conversely, columns 3 and 4 show that core updates do not influence the *Ahrefs Domain Rating and Majestic Trust Flow*. Finally, when examining the role of heterogeneity, columns 5 and 6 suggest 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 Effects of Google Core Updates

The analysis of the previous section has shown that search engines such as Google have the capacity to modify the search traffic of media outlets through the adjustment of their indexation algorithms. Specifically, we have shown that Google's recent core updates have reduced the number of keywords that news outlets have in top positions in organic search results, and that this has reduced their organic search visits. The objective of this section is to investigate whether the effects established at the outlet level have had implications for the concentration at the market level. Although this is ultimately an empirical question, it is important to understand whether the changes introduced by the updates have increased the visibility of small and niche news outlets, or whether they have favored the consolidation of “superstar” news outlets. We believe that this analysis is useful 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 considering the market share of news outlets in their corresponding national markets, for each of the three outcome variables examined in our study. 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, the figure reveals 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 10 and 11 show the results of the estimation of equation (4) for different subsamples of outlets. Table 10 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.[39] If we now consider

[39] 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 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

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.

Table 11 analyzes the effect of Google's core updates considering the impact in different types of news outlets. 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. This suggest that the reduction in the number of top keywords, resulting from core updates, was more important for small than for large national news outlets, and that it was more important for large than for small sport news outlets.

7. Further Analyses Using Monthly-Aggregated Data

Up to this point, we have investigated the impact of Google algorithm's core updates on outlet-level traffic and market-level concentration using daily data. In this section, we complete our analysis exploiting rich keyword-level data and search engine-specific traffic data. This information is only available at the monthly level, but it is useful to examine the heterogeneous effects of core updates across different categories of keywords. Moreover, we can analyze whether Google's core updates have increased the concentration of Google Search in comparison to the concentration of other search engines (Bing, DuckDuckGo and Yahoo). 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 also affect traffic concentration at the keyword level. Notice that each keyword used during consumers' queries generates visits to different news outlets. Considering this, we want to analyze whether core updates have increased concentration at the individual keyword level. For this purpose, we use *monthly data* on all the keywords that generate search visits from Google to the news outlets. In a nutshell, our data details search visits at the keyword-outlet-month-year level. Similarly to equation (5) above, our regression specification is such that,

$$\ln[\mathrm{HHI}_{\mathrm{it}}] = \varphi_0 + \varphi_1 \mathrm{CoreUpdate}_{\mathrm{t}} + \varphi_2 \mathrm{CoreUpdatePlus1}_{\mathrm{t}} + \varphi_3 \mathrm{CoreUpdatePlus2}_{\mathrm{t}} + \gamma \mathrm{Age}_{\mathrm{it}} + \theta_{\mathrm{i}} + \mu_{\mathrm{t}} + \varepsilon_{\mathrm{it}}, \quad (5)$$

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* Google core 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 12 presents our findings. Column 1 estimates specification (5) above and shows overall keyword concentration goes down a 3.4% on the month of the core update, no change the month after, and a 3.1% increase in concentration two months after the core 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 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 HHI drops by 8% in the month of a *Big* update, is unaffected the month following the update, and spikes 8% two months after the big update. There is no apparent impact on keywords appearing in few outlets. Columns 5 and 6 show findings after classifying keywords into those appearing more than 8 months and those appearing 8 months or less. Our results show opposite effects of the updates. Column 5 shows that keywords with longer runs see their concentration increase after the updates with no additional impact of a *Big* core update. In contrast, keywords with shorter runs are not affected by general updates, but they have a 5% reduction of concentration after a *Big* update and 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), Bing (0.1%), DuckDuckGo (0.2%) and Yahoo (0.3%). We estimate the following model,

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

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 13 investigates the impact of core updates in search engine concentration. Column 1 estimates specification (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 13 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* core updates had an impact in search engine visit concentration at the outlet level. In particular, it increased concentration by 0.6% on the month of the updates, and by 0.4% one month after.

8. Conclusions

Search engines are crucial intermediaries to access the news contents available on the Internet. Consumers frequently look for the latest news in Google, Bing or Yahoo, rather than directly visiting online news outlets. They expect search engines to answer to their queries with links to the latest breaking news and information on the top stories, weather, business, entertainment, and on politics. This situation raises the question of how search engines can affect citizens' access to a variety and diversity of high-quality news, opinion-based editorials, and information analyses through different sources of information. The concern is not just about how news outlets adjust their headlines and news stories to rank higher in the search results, but also about the risk that search engines can reinforce the market and the media power of some publishers.

Our paper constitutes a first step to study these questions by examining the capacity of Google Search to influence the type of news outlets that access European readers. We have addressed two basic questions. First, we have analyzed the mechanisms that determine the number of visits that news outlets receive from Google. Every time a consumer makes a query for a specific keyword, Google ranks all the web pages considering how accurately they can answer the consumer's query and displays them on its search results page. The rank news outlets obtain for each keyword determines the search visits they receive from Google, and for this reason they invest important resources to increase the number of keywords they have in top positions. To isolate the effects that Google's indexation has on the search visits of news outlets, we have used an instrumental variable approach. Specifically, we have relied on Google's core algorithm updates to obtain an exogenous source of variation in news outlets' indexation. Our results show that the core updates rolled out by Google in the period 2018-2020 affected news outlets in different directions and magnitudes, and that overall had a negative effect in the number of keywords that news outlets have in top search results. We also obtain that the number of keywords that news outlets have in top 10 search results have a positive effect in their visits. Specifically, we have shown that a 1% increase in the number of keywords in top 10 positions generates around 6% increase in the number of search visits, and 4% increase in the total number of desktop and mobile visits. These results are confirmed when we classify news outlets according to different criteria (e.g., specialization, national rank), but are less clear-cut when we analyse national markets individually.

The second question addressed in our paper is whether Google core updates have increased the concentration in the European media markets. We have found that the three *Big* core updates released in this period implied a 1% reduction of market concentration. However, this effect was mostly compensated by the effect of the *Not Big* core updates, a 0.08% increase of market concentration. In addition, we have shown that *Not Big* updates increased the market concentration of search visits for National news outlets, and that they reduced the concentration for Sports news outlets. Finally, when we consider the effects of the updates at the country level, we find that results are quite heterogeneous. *Big* core updates reduced market concentration in Finland, Germany, and Greece, but increased it in Portugal. *Not Big* core updates increased concertation in Finland and Netherlands.

Interestingly, in a case study of Spanish keyword search data, we have also shown that *Big* core updates decreased concentration at the market-keyword level, but these results are heterogeneous depending on the importance of the keyword, its wide coverage across news sites as well as its run length in the search algorithm. Thus, for example, core updates importantly increased market concentration for those keywords more 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, although each specific update can affect national markets in different directions.

This research has important implications for policy makers interested in understanding competition in online markets. While our paper has focused on the media market, future investigations are needed to examine how search engines' indexation algorithms shape search traffic and competition in other markets. The European Union has recently implemented new regulations to improve the transparency in online intermediation activities. In July 2019, the EU approved a legislative initiative, known as the platform-to-business (P2B) regulation, that aims at creating a fair, transparent and predictable business environment for smaller businesses and traders participating on online platforms (European Commission, 2019).[40] In December 2020, the EU proposed more instruments to regulate online intermediaries, through the Digital Services Act (DSA) and the Digital Markets Act (DMA). These regulations were approved in 2022 and will be enacted at the beginning of 2024. Similar initiatives are taking place in other parts of the world. The DMA and DSA will bring profound changes in the governance of digital markets. Some of its most relevant obligations include rules that mandate digital platforms to share certain types of data and information with regulators and researchers. Other rules impose transparency obligations regarding digital platforms' algorithms and internal content moderation activities. In 2023, the EU designated Google Search and Bing as *Very Large Search Engines* that have to comply with the full set of new obligations under the DSA.[41]

An aspect not addressed in our paper is how human editorial decisions in newspapers is complemented (or even replaced) by algorithms that offer personalized recommendations to readers (Agrawal et al. 2018; Claussen et al. 2021). As explained 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 become more expert at catering consumers tastes, societies may lose their ability to receive neutral information and might confine consumers into echo chambers with algorithms trained on prior individual-level data reinforcing this phenomenon (Sunstein, 2001; Boxell, Gentzkow,

[40] This regulation, which entered into application on 12 July 2020, establishes that search engines shall set out the main parameters determining their rankings and the relative importance of these parameters. For example, intermediation platforms should disclose whether their ranking are influenced by direct or indirect remuneration from business users. They shall also show in their terms and conditions a description of any differentiated treatment they might give to goods or services offered by themselves or by businesses they control compared to third party businesses (e.g., related to access to data, ranking, fees).

[41] See https://ec.europa.eu/commission/presscorner/detail/en/IP_23_2413

and Shapiro, 2017; Gentzkow, 2018; Goldfarb and Tucker, 2019; Claussen et al., 2021).[42] Another relevant aspect not considered in our analysis is the fact that search engines and news outlets might compete to attract users and obtain proprietary information about their preferences that can then be sold in the advertising market (Prat and Valletti, 2021).

Finally, our paper is also relevant to understand the role that search engines and news aggregators have for the journalism and democratic institutions. Gentzkow and Shapiro (2010) explain that in the US government regulation of news media ownership is based on the proposition that news content has a powerful impact on politics, and that unregulated media markets will tend to produce too little ideological diversity. These beliefs have justified significant controls on cross-market consolidation in broadcast media ownership, on foreign ownership of media, and on cross-media ownership within markets. The emergence of digital platforms and social networks poses a new treat for the regulation of the media market. On the one hand, search engines and social network are easy and immediate intermediaries to access news contents. On the other hand, algorithmic indexation and recommendation systems can potentially limit the diversity of information sources that consumers receive.

[42] Claussen et al (2021) carry out a field experiment with a major news outlet in Germany and obtain that personalized recommendation reduces consumption diversity and that this effect is reinforced over time. They also find that users associated with lower levels of digital literacy and more extreme political views engage more with algorithmic recommendations.

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.

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.

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.

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.

Decarolis, F., G. Rovigatti, M. Rovigatti and K. Shakhgildyan (2023), AI & Data Obfuscation: Algorithmic Competition in Digital Ad Auctions, CEPR Discussion Paper 18009.

Drugovy, M. and D.-S. Jeon (2019), Vertical Integration and Algorithm Bias. 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.

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.

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.

Varian, H. R. (2007), Position auctions, *International Journal of Industrial Organization* 25 (6), 1163-1178.

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.

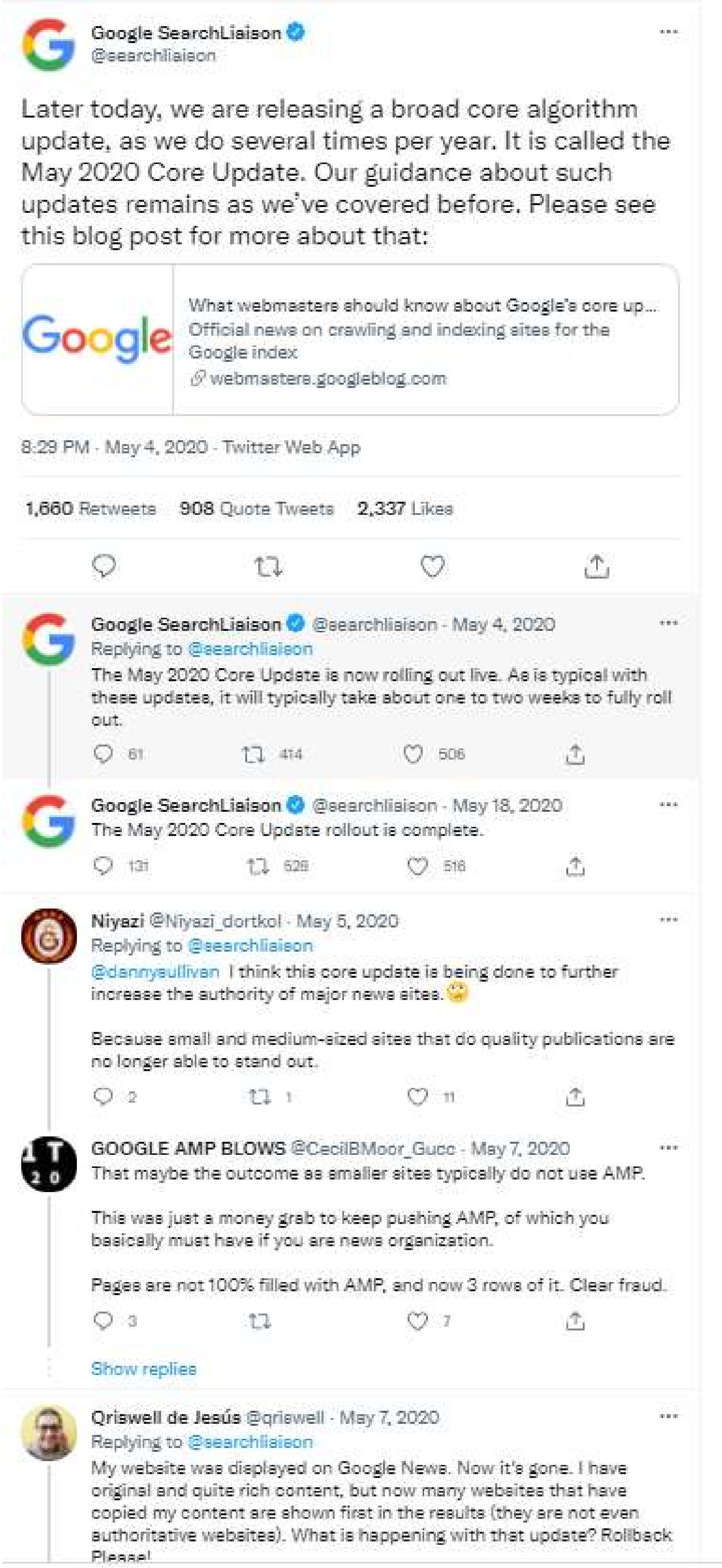
Xu, L., Chen, J., and A. Winston (2012), Effects of the presence of organic listing in search advertising. Information System Research 23, 1284-1302.

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.

Figure 1. Google's announcement of May 2020 Core



Google SearchLiaison
@searchliaison

Later today, we are releasing a broad core algorithm update, as we do several times per year. It is called the May 2020 Core Update. Our guidance about such updates remains as we've covered before. Please see this blog post for more about that:

Google

What webmasters should know about Google's core up...
Official news on crawling and indexing sites for the Google index
webmasters.googleblog.com

8:29 PM · May 4, 2020 · Twitter Web App

1,660 Retweets 908 Quote Tweets 2,337 Likes

Google SearchLiaison @searchliaison · May 4, 2020
Replying to @searchliaison
The May 2020 Core Update is now rolling out live. As is typical with these updates, it will typically take about one to two weeks to fully roll out.
61 414 506

Google SearchLiaison @searchliaison · May 18, 2020
The May 2020 Core Update rollout is complete.
131 529 516

Niyazi @Niyazi_dortkol · May 5, 2020
Replying to @searchliaison
@dannysullivan I think this core update is being done to further increase the authority of major news sites.

Because small and medium-sized sites that do quality publications are no longer able to stand out.
2 1 11

GOOGLE AMP BLOWS @CecilBMoor_Gucc · May 7, 2020
That maybe the outcome as smaller sites typically do not use AMP.

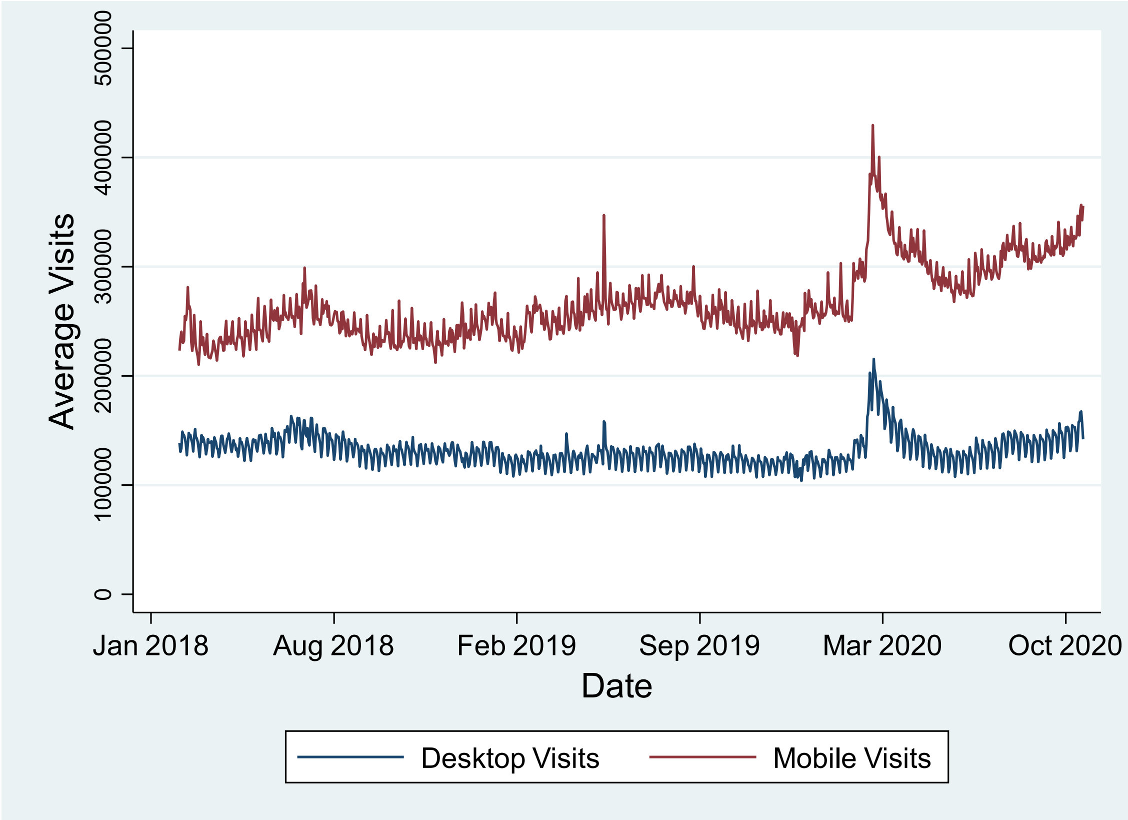
This was just a money grab to keep pushing AMP, of which you basically must have if you are news organization.

Pages are not 100% filled with AMP, and now 3 rows of it. Clear fraud.
3 7

Show replies

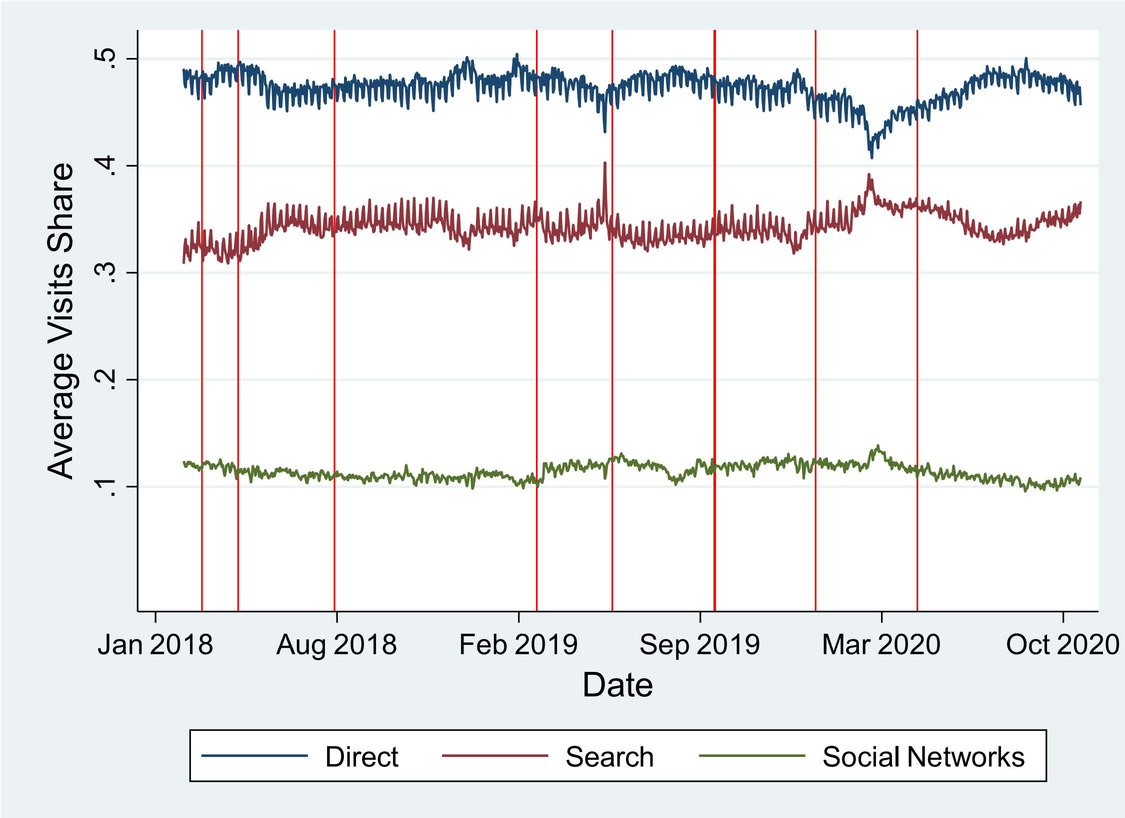
Qriswell de Jesús @qriswell · May 7, 2020
Replying to @searchliaison
My website was displayed on Google News. Now it's gone. I have original and quite rich content, but now many websites that have copied my content are shown first in the results (they are not even authoritative websites). What is happening with that update? Rollback Please!

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.

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

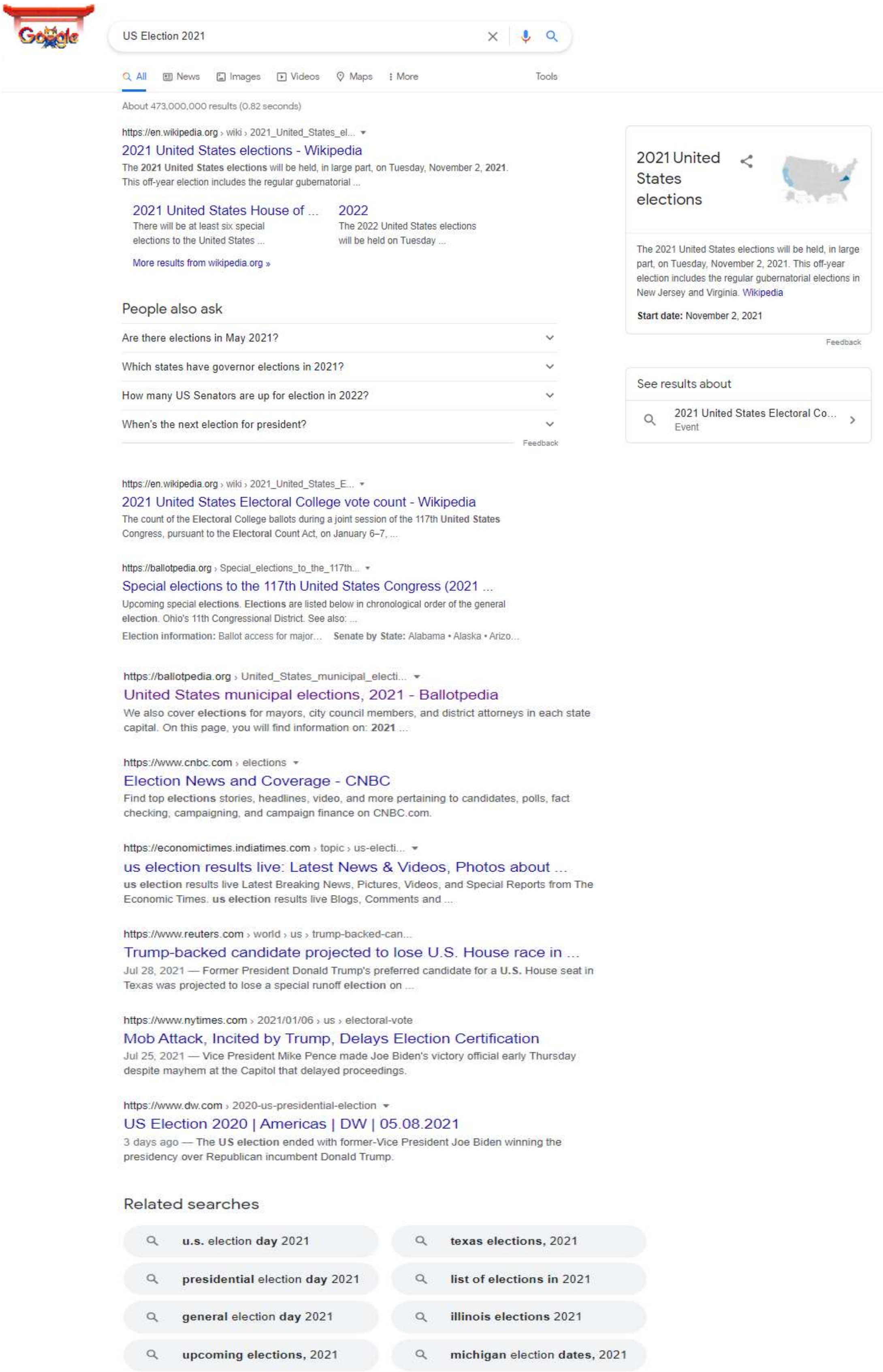
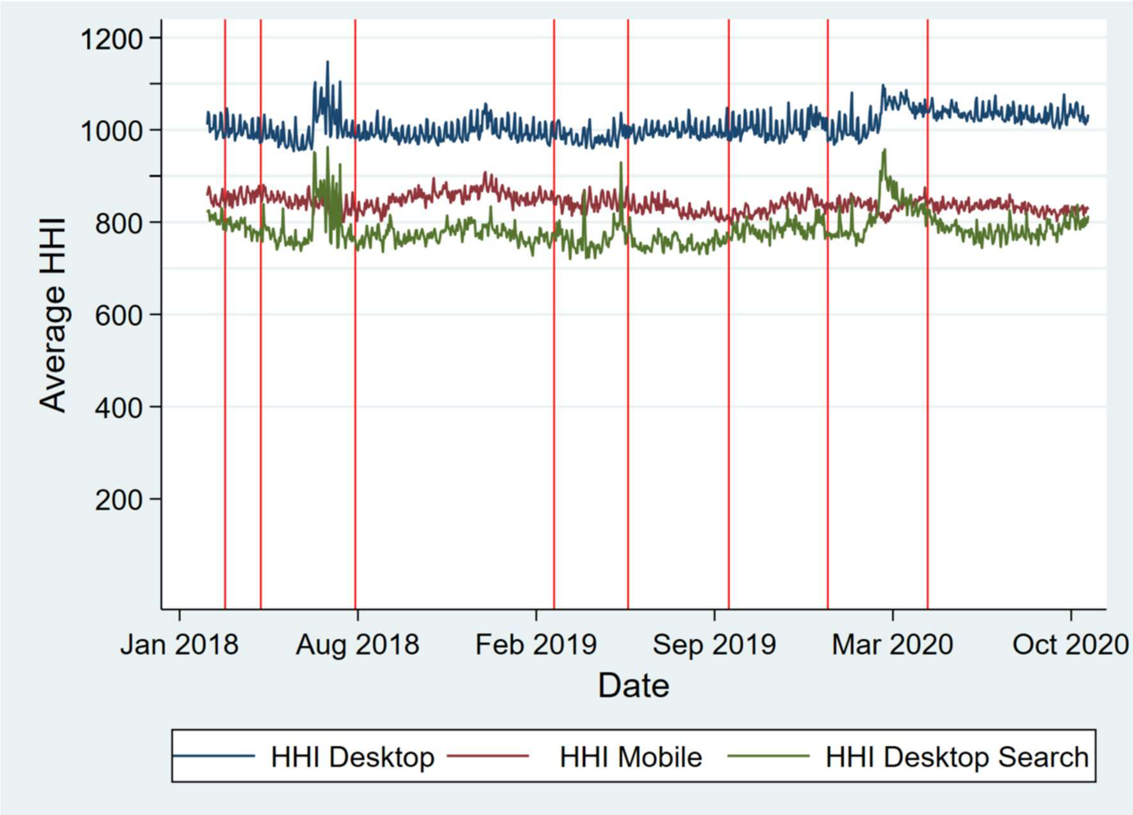


Figure 5. HHI for Desktop Search Visit, Desktop Visits and Mobile Visits (January 2018 - November 2020)



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.[43]
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.[44]
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.

Source: Own elaboration and Moz.com

43 See also: https://searchengineland.com/googles-may-2020-core-update-was-big-and-broad-search-data-tools-show-334393

44 In addition, Google said that this update eliminated duplicate results it order to avoid some site to be listed several times on top results (it increase site diversity) for most search queries.

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.cor	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 3. 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 4. 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 5. 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 6. Impact of Google Core Updates on Number of Search, Desktop and Mobile visits per News Outlets Type

Dependent Variable	**First Stage - IV** **Δln(Keywords Top10) t-4**		**Second Stage - IV** **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 7. 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 8. 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 9. 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 10. 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 11. 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.

Table 12. 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 13. 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.