Do Search Engines Increase Concentration in Media Markets?

Joan Calzada
Associate Professor
Universitat de Barcelona

Nestor Duch-Brown
Joint Research Centre (JRC)

Ricard Gil
Associate Professor
Smith School of Business, Queen's University

EODS
Economics of Digital Services
@Penn

CTIC
Center for Technology, Innovation and Competition

The Warren Center
for Network & Data Sciences

Penn
University of Pennsylvania

Economics of Digital Services (EODS) is an initiative of the University of Pennsylvania's Center for Technology, Innovation and Competition (CTIC) and The Warren Center for Network & Data Sciences. Its aim is to generate independent research on the economics of digital services and the role of data and algorithms in the business strategies of digital platforms. The initiative was funded by a major grant from the John S. and James L. Knight Foundation to support scholarly inquiry and novel approaches in the evolving digital age.

To learn more about the initiative, visit www.law.upenn.edu/digitaleconomics/.

Center for Technology, Innovation and Competition (CTIC)

CTIC is an interdisciplinary academic center at the University of Pennsylvania Carey Law School that bridges law and technology for academia and students. Focusing on intellectual property, antitrust, internet law, and privacy law and policy. CTIC delivers foundational research that shapes the way legislators, regulatory authorities, and scholars develop policy and regulatory frameworks. CTIC also produces programming that explores the full range of scholarly perspectives, engages with industry experts, and prepares the next generation of technology scholars, lawyers, and policymakers.

www.pennCTIC.org

The Warren Center for Network & Data Sciences

The University of Pennsylvania's Warren Center fosters research and innovation in interconnected social, economic, and technological systems. Collaborating with Penn affiliates, it focuses on the role of data and algorithms to understand networked systems and how they can improve lives. The center also produces events that connect researchers, students, and entrepreneurs across the spectrum of network science.

www.warrencneter.upenn.edu

Do Search Engines Increase Concentration in Media Markets?

Joan Calzada[1] Nestor Duch-Brown[2] Ricard Gil[3]

September 2021

Abstract

Search engines are one of the main channels to access news content of traditional newspapers. In the European Union, organic search traffic from Google accounts for 35% of news outlets' visits. Yet, the effects of Google Search on market competition and information diversity are ambiguous, as the firm indexes news outlets considering both domain authority and information accuracy. Using detailed daily data traffic for 606 news outlets from 15 European countries, we assess the effect of Google Search's indexation on search visits. Our identification strategy exploits nine core algorithm updates rolled out by Google between 2018 and 2020 in order to achieve exogenous variation in news outlets' indexation. Several conclusions follow from our estimations. First, Google core updates overall reduce the number of keywords that news outlets have in top positions in search results. Second, keywords ranked in top search position have a positive effect on news outlets' visits. Third, our results are robust when we focus the analysis on different types of news outlets, but are less conclusive when we consider national markets separately. Our paper also analyzes the effects of Google core updates on media market concentration. We find that the three "big" core updates identified in this period reduced market concentration by 1%, but this effect was mostly compensated by the rest of the updates.

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, Innovación y Universidades (MCIU), Agencia Estatal de Investigación (AEI) and the Fondo Europeo de Desarrollo Regional (FEDER) (Grant: RTI2018-096155-B-I00). The usual disclaimer applies.
[1] Universitat de Barcelona. Barcelona, Spain; calzada@ub.edu.
[2] Joint Research Centre (JRC), Seville, Spain; Nestor.DUCH-BROWN@ec.europa.eu.
[3] Smith School of Business, Queen's University. Kingston, Canada; ricard.gil@queensu.ca.

1. Introduction

A large and increasing fraction of consumers use algorithm-driven platforms to access the contents of traditional news outlets. In the European Union, around 45% of news outlets' visits comes from direct traffic by consumers that directly browse the news sites' address when looking for news contents, 35% from organic search traffic from search engines (mostly from Google), and around 12% from social network traffic (Facebook, Twitter).[4] Recent studies have analyzed the effects of digitalization on competition in the media market (Athey et al. 2017, Chiou and Tucker, 2017: Calzada and Gil, 2021), the quality of journalism (Cagé et al, 2020; Bandy and Diakopoulos, 2020), and the development of democratic institutions (Gentzkow and Shapiro, 2011; Boxell, Gentzkow, and Shapiro, 2017; Peterson, Goel and Iyengar, 2019). However, very little is known about the effects that search engines and social networks might have in the development and future prospects of media markets (Sismeiro and Mahmood, 2018; Cagé, Hervé and Mazoyer, 2020).

The empirical literature has shown that digital search increases the proportion of traffic going to sites that are relatively less visited, a situation known as the "the long tail" (Anderson, 2006; Fleder and Hosanagar, 2009; Brynjolfsson, Hu, and Simester, 2011; Zhang, 2018; Goldfarb and Tucker, 2019). The online channel facilitates the discovery of unknown products and increases the variety of products available and purchased from retailers. A relevant question for the media market is whether search engines increase the visits to large and well know news sites, or whether they can thicken the long tail by giving more visibility to less popular, niche, and local newspapers. Our paper aims to address this question by examining how recent changes in Google Search's indexation activity has modified the search traffic of European news sites. Specifically, we analyze the effects of Google's core algorithm updates on the concentration of the European media markets.

Google Search uses bots to crawl news outlets pages and collect information about their contents. Then, when a consumer has a query about a keyword or a phrase it 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). 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. 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).[5] Considering this, news outlets with a low static ranking (low domain authority) might

[4] Own calculations, based on SimilarWeb data.

[5] 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.

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 a news outlet in the search market depends on how well it ranks relatively to its closer competitors, and more generally on how Google's algorithms weight domain authority and content accuracy.

One important difficulty for studying how Google's indexation affects news outlets search traffic is that the visits to news outlets can be correlated with relevant but unobserved news sites characteristics, or with the contents of the news stories 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. To deal with this endogeneity problem our paper adopts an instrumental variable identification strategy. Specifically, because algorithm updates have a direct effect in news sites' indexation and are a source of exogenous variation for the sites' visit results, we use Google's core algorithm updates as an instrument for the number of keywords that news outlets have on top search positions.

Our paper examines 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 web sites characteristics. The updates are designed to improve the way Google's system assess content and to ensure that overall it offers relevant and authoritative content to searchers. We exploit these quasi-natural experiments to examine how changes in news outlets' indexation affect news outlets' search visits and the distribution of traffic across outlets.[6] Specifically, we analyze whether Google core updates are reinforcing the skewness of the distribution of search traffic across news outlets, or if they are making 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, search and social network visits, and can distinguish between desktop and mobile traffic. We complement these data with information on keywords ranking distribution from Ahrefs. These data show the daily number of keywords that news outlets have on positions 1-10 and 11-100 on Google's 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 search traffic received by European news outlets. In particular, we use an instrumental variable approach

[6] 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. https://t.co/ohdP8vDatr (Google Search Liaison (@searchliaison) October 11, 2018). In spite of this, there are economic incentives for manipulating 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)

that exploits the changes in Google's core algorithm rolled out between 2018 and 2020 to obtain exogenous variation in the news outlets' indexation. Our results show that the three "big" core algorithm updates identified in this period had a negative effect in the number of keywords that news outlets had in Google's top 1-10 search results and a positive effect in the number of keywords in top 11-100 positions. The rest of core updates had a negative effect in the number of keywords in the top 100 positions. Overall, these findings imply that core updates have reduced the visibility of news outlets in Google's results pages, as they have lost positions in search results. Our analysis also reveals that the number of keywords that news outlets have in top 100 search results is positively related to their search visits and to the total desktop and total mobile visits. These results are robust when we replicate the analysis for different types of news outlets (national, regional, business, sports, tv/radio), or when we group them according to different features (national rank, domestic traffic, traffic from Google). Results are less conclusive when we examine national markets separately. In this case, big and non-big core updates exhibit different results across countries.

The second contribution of the paper is to analyze the effect of Google' core algorithms updates in 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. Finally, it is interesting to note that Google's core updates have increased the market concentration among national generalist news outlets.

Our analysis and findings have important policy implications. In the last few years, policy concerns have emerged around the growing market power of digital platforms that are based on indexation or recommendations algorithms. It is unclear which are the biases that 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 has 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 Commission found that none of the alternative sources of traffic available to competitors could effectively replace the generic search traffic from Google.

Of particular importance is the role of search engines in media markets. The particular sources used by consumers to obtain news and information can affect their political attitudes and voting intentions, alter their perceptions and opinions, and reinforce stereotypes (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 the concentration of the media market. Finally, section 7 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). Prior theoretical work has shown that search engines can adjust their organic results to favor sponsored search from which they obtain larger profits (Xu, Chen, and Whinston, 2012; Taylor, 2013; and White, 2013). Search engines set the quality of their organic search taking into account that this service attracts consumers but cannibalizes sponsored search profits.

In a similar line, Cornière and Taylor (2014) and Burguet, Caminal, and Ellman (2015) analyze biases in search results when search engines are vertically integrated with a seller.[7] De Cornière and Taylor (2014) consider a market with two websites and a search engine that obtain their revenues from advertising. They 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, generating more ad revenues in the engine. As a result, vertical integration can increase or decrease the level of search bias, depending on the type of bias existing without integration. Burguet, Caminal, and Ellman (2015) consider a model in which a search engine interacts with two distinct but related markets. Its organic search results help consumers match with publishers that provide online content, and its sponsored search results help consumers to interact with merchants selling offline products. Moreover, publishers display ads on their contents and compete with the engine to provide ads in the product market. The

[7] Zhu and Liu (2018) study Amazon's entry in markets covered by its marketplace sellers. They find that Amazon targets successful product spaces and avoids products that require greater efforts to grow.

engine's organic search service attracts consumers who then can use the engine's sponsored search results. In this context, the engine can reduce publishers' ad-effectiveness by diverting content-searching consumers, although this reduces its reputation in the search market. The model show that the integration of the engine with a fraction of content providers internalizes these vertical externalities and improve organic and sponsored reliability, but also generates horizontal effects that can reduce social welfare.

Other papers have shown that search engines may degrade the quality of their search results in order to reduce competition among sellers and increase their fees. Chen and He (2011) and Eliaz and Spiegler (2011) show that search engines can lower the quality of their results to relax sellers' competition and extract higher profits. Hagiu and Jullien (2011) examine when an intermediary may degrade the quality of the search process through which consumers find sellers. First, since the intermediary derives revenues whenever consumers transact with stores, it can introduce some noise in the search process (i.e. to divert search) in order to increase the number of searches that consumer make. Second, the intermediary may distort search when it cannot price discriminate among stores and the participation of the marginal stores is binding, or when it extracts a higher fraction of revenues from less popular stores. Third and last, an intermediary may divert search to influence the strategic choices (i.e. pricing) of affiliated stores. Consumer surplus can increase when the intermediary alters the composition of the demand faced by each store.

Other recent papers study whether digital platforms bias results in "recommendation systems". Bourreau and Gaudin, (2018) examine a monopoly streaming platform that offers access to two differentiated content providers. They show that if consumers are sufficiently insensitive to bias, the platform uses the recommendation system to reduce the market power of content providers, and hence 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. They examine whether the platform is more biased under a prominence-for-money scheme or under a prominence-for-data scheme, showing that this depends on the marginal revenue from shared data. Drugov and Jeon (2017) study the incentives of a vertically-integrated platform to bias recommendations towards its own content when consumers' utility in the long-run is shaped by their short-run usage. In the static setting, the platform has no incentives to bias since the fee charge to content providers is fixed ex-ante. In the dynamic setting, however, past consumers' experience affects their willingness to pay for contents and this affects the bargaining between the platform and the content providers for the fee.

There is also an empirical literature addressing the existence of platform biases. 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 vertical integration 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 more visibility to them. Hunold, Kesler, and Laitenberger (2017) investigate the default hotels' rankings offered by Booking and Expedia to their consumers, which differ from the rankings they would obtain when asking for hotels prices or reviewer ratings. Using data on hotels for 250 European cities, they find that ranking position of hotels in these platforms are lower when they are also announced in a rival platform, at a lower price. Aguiar, Waldfogel and Waldfogel (2021) analyze potential biases in Spotify. Using data on Spotify curators' rank of songs on New Music Friday playlists in 2017, they find that Spotify's New Music Friday rankings favor independent-label music as well as music by female artists. Songs with higher New Music ranks obtain more ex post streaming success. Moreover, independent music, and music by female artists, receive higher ranking positions than their eventual performance seems to warrant.

Our paper also contributes 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, Hu and Simester, 2011; Oestreicher-Singer and Sundararajan, 2012; Datta, Knox and Bronnenberg, 2018; and Aguiar and Waldfogel, 2020). There is ambiguous evidence that recommendation systems favor products in the long tail and encourage sellers' participation because these products become more attractive for niche consumers. Oestreicher-Singer and Sundararajan (2012) analyze more than 200 book categories in Amazon.com. They collect information on the co-purchase links shown to consumers when these look at a particular book (links on titles that other consumers bought together with each book). They explain that when the co-purchase links are shown to consumers there is a three-fold increase in the influence that complementary books have on each other's demand. They obtain that book categories with a higher popularity rank are associated with a significantly lower demand diversity. In addition, consistent with the theory of the long tail, they show that niche books perform better and popular books perform relatively worse in book categories where recommendations are more important. Hosanagar et al. (2014) examine whether recommender systems fragment users. Using data from an online music service, 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 2-week randomized field experiment in a top online retailer in North America, they demonstrate across a wide range of product categories that collaborative filters are associated with less sales diversity relative to a world without product recommendations. Absolute sales and views for niche items increase, but their gains are smaller than for popular items.

Another stream of the literature that we contribute to investigates whether algorithms that automate decision-making may produce discriminatory outcomes. Lambrecht and Tucker (2019) show the difficulties of regulating algorithms to prevent instances of apparent discrimination, such as gender biases in ad targeting.[8] They analyze a field experiment

[8] 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.

investigating the impact of an algorithm that delivered ads promoting job opportunities in the Science, Technology, Engineering and Math (STEM) fields. The advertisement campaign was intended to be gender-neutral in its delivery, but the ad was shown to over 20% more men than women. The reason is that younger women are a prized demographic and are more expensive to show ads to. This suggest that algorithms that optimize cost-effectiveness in ad delivery might generate discriminatory outcomes.

Finally, this paper 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 number of 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 these papers in that we show how search engines, which are an important channel to access news and policy information, can affect news outlets' visits. In this sense, we contribute to the literature that examines how the media markets may affect political polarization (Gentzkow and Shapiro, 2011; Boxell et al., 2017; Bakshy et al., 2015) by adding a potential channel connecting search algorithms and concentration in online media markets.

3. Google search algorithm

Search engines such as Google, Bing and Yahoo use bots to crawl pages on the web, going from site to site, collecting information about these pages and indexing them. When consumers have a specific query, search engines use algorithms to analyze the pages they have indexed and rank them according to multiple factors that determine the order in which the links to the pages appear in the consumers' search results. The indexation of webpages can respond to several 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.

Googlebot is the robot of Google that crawls accessible webpages, sees and classifies their content, and indexes each website. Google ranks web pages according to the EAT criteria, which consider their Expertise, Authoritativeness and Trustworthiness. Specifically, pages are evaluated considering three dimensions:[9] the quality of the website; the quality of the main content on the page; and the quality of the author(s) of the main content.[10] Google explained the relevance of these aspects in 2011, after rolling out the "Panda update" of its

[9] A detailed definition of the concepts "Expertise", "Authoritativeness", and "Trustworthiness" can be find in the Google's guidelines for its reviewers: https://guidelines.raterhub.com/searchqualityevaluatorguidelines.pdf See also https://www.pi-datametrics.com/blog/google-e-a-t-ultimate-guide/

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

algorithm.[11] Furthermore, in 2015 Google published its EAT guidelines (updated on July 20, 2018, and May 16, 2019) to explain its human search evaluators[12] how they have to evaluate web pages, and how this is used as a reference to rate the performance of Google's algorithms.

These guidelines show how Google determines the quality of web contents. According to them, websites and pages that aim at helping users are considered of a high quality. Specifically, they establish that high quality pages should fulfill its intended purpose, but also their purpose should be user-centered. 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. In spite of 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 firms. Interestingly enough, Google does not share any scoring or indexing criteria externally. However, SEO software companies have applied reverse engineering to identify the factors used by Google to index websites and have created several metrics that try to approximate the ranking or "domain authority" of websites.[13] Site owners can take several actions to improve the rankings of their websites, but according to industry experts these actions only work after Google updates its algorithms.

3.1 Google's Core Updates

Google introduces many changes in its algorithm and systems every year. However, only a few times per year it makes large "core updates" that generate significant modifications in the way it ranks and indexes search results. According to Google, these changes "*are*

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

[12] 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/

[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.

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 web sites characteristics. However, the updates might affect different types of websites in different ways. The updates generate fluctuations in search rankings throughout the next days and weeks after their adoption. Google notifies the launch of its core updates because "*they typically produce some widely notable effects. Some sites may note drops or gains during them. We know those with sites that experience drops will be looking for a fix, and we want to ensure they don't try to fix the wrong things. Moreover, there might not be anything to fix at all.*" Site owners are aware that traffic recovery can be extremely challenging after a core update. According to Google, there is nothing site owners can do to recover their search traffic after core updates.[15] In spite of this, Google offers advice and guidelines to webmasters on how to orientate their pages to improve search results.[16] Figure 1 shows as an example the announcement on twitter of Google's May 4, 2020 core update, and some of the immediate reactions of small newspapers. See in Table 1A the list of the nine core updates confirmed by Google during the period 2018-2020 and that we consider in our empirical analysis.[17]

3.2 General updates of search algorithms

In addition to the core updates, Google regularly introduces changes in its algorithms. 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. On March 26, 2018, Google announced that the Mobile-First Index was finally rolling out. On March 2020, the firm reported that over 70% of crawled sites were on Mobile-first indexing and that they planned to use if for the whole web on September 2020, although finally they decided to delay it to the end of March 2021. Considering that Google was testing the index for many months, and that they were migrating

[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) October 11, 2018). See https://blog.searchmetrics.com/us/google-update-november-2019/

[16] Google Webmaster Blog (https://webmasters.googleblog.com/2019/08/core-updates.html) suggests different actions after being affected by Core Updates. Google also publishes their "Webmaster Guidelines", showing how they index and rank web site. These guidelines also outline some of the illicit practices that may lead to a site being removed entirely from the Google index or otherwise affected by an algorithmic or manual spam action. See https://developers.google.com/search/docs/advanced/guidelines/webmaster-guidelines.

[17] A complete list of Google's core updates can be found here: https://moz.com/google-algorithm-change

sites gradually, it is unclear how this specific roll-out affected the overall index and desktop and mobile search traffic.

In addition to these changes, every day Google releases one or more changes to its algorithm in order to improve the search results for consumers and to correct different types of bugs. Many of these changes are unnoticeable. 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 Google algorithm updates 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 when Google introduced "passage indexing" in February 10, 2021 to index specific passages, not just the overall page. Google considers that passage-based indexing can affect 7% of search queries across all languages. In other occasions, rank tracking tools and webmaster chatter suggest the existence of unconfirmed updates by Google, although these can be temporary and disappear after a while.[18]

Another 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. Google considers that BERT can affect 10% of searches. 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.

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

[18] See the previous footnote for more details.
[19] In parenthesis, the number of news outlets in the corresponding country.

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) and reviewing 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 report traffic information for sites 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 starts on January 1, 2018. Overall, we aimed to have a well-balanced sample of news outlets. We classify the sites in our sample in different categories such as their specialization (national, regional, business, sports, tv/radio), their rank at the national level, their internationalization level (percentage of domestic visitors), and the percentage in the search traffic coming from Google Search (as opposed to other search engines).

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 search traffic is originated in Google Search. We do not have daily data on the mobile search visits because SimilarWeb does not collect such information. 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. We also consider as a control variable the *Daily Desktop Direct Visits*, which shows the daily traffic to a news outlet from a different web domain or from the beginning of an empty browsing session. This variable helps us 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 November 2020. The red lines in the figure show the dates of Google's core algorithm updates. We observe that in this period the mobile traffic has grown at a higher rate than the desktop traffic. The figure also shows that the COVID-19 dramatically increased the desktop and mobile visits 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, search and social networks traffic.

We classify news outlets 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.

Our dataset includes several measures of website performance from Ahrefs,[20] one of the more important SEO software firms. As explained above, in the last years Google has modified its algorithm to reflect its EAT criteria, and SEO companies have developed their own software to monitor websites' SEO health over time.[21] We have collected information on two daily metrics from Ahrefs.com. *Ahrefs Domain Rating* (ADR) measures the strength of a website's backlink profile compared to the others in their database on a 100-point scale.[22] This metric looks at the quality and quantity of domains linking to an entire website. Therefore, ADR is a measure of the "relative link popularity" of websites. According to Ahrefs, this metric works in a similar way to the original PageRank calculation (although it ranks websites and not web pages).[23] *Ahrefs Organic Keywords*, shows the number of keywords that a news outlet has in the top 100 organic search results.[24] Specifically, it analyzes if a news outlet ranks in the top 100 search results for any of the ~605 million keywords Ahrefs have in their database. The number of organic keywords news outlets have in top positions can change over time simply because Ahrefs 'database is growing, and not because the outlet ranks higher in search queries. It is also important to mention that Ahrefs organic keywords metric is country-specific. Ahrefs collect information on *Keywords 1-3*, *Keywords 4-10*, *Keywords 11-100*, to measure the number of keywords that a site has in each of these intervals. To simplify our analysis, we use these measures to create three variables. *Words Top 100* shows the sum of all keywords that a news outlet has in the top 100 organic search results. In addition, *Words Top 10* and *Words Top 11-100* reflect the number of words that news outlets have in the top 10 and in the top 11-100 organic search results, respectively. According to Moz, the results in 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] This implies that obtaining keywords in top search results is crucial for news outlets to obtain search traffic, although they might have hundreds of keywords in top 11-100 positions that complement their visits. Also, note that users can redefine their search keywords and phrases after a first search to obtain more accurate information. Figure 4 shows an example of the 10 first search results for "US Election 2021", which are in the 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

[20] Other important SEO companies are Majestic and Moz.

[21] 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. As a result of its misuse and SPAM, PR is no longer a quality metric to assess websites. Google stopped updating it since 2013, although the firm has said that it still uses it internally in its web positioning algorithm.

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

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

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

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

institutional sites. Notice that Google's first results page includes "zero-click searches", which are answers to queries that do not send consumers to a third-party websites. Google uses its Direct Answer Box to offer answers to many consumers' queries, such as for celebrities, geography or history. Search queries about the weather or stock market prices are also answered directly by Google. It is considered that around 50 percent of searches currently end without a click on an organic search result. Table 2 shows summary statistics for all the variables obtained from SimilarWeb and AhRefs.

4.2. Empirical Strategy

Our empirical model examines how Google Search affects the visits received by European news outlets. We consider that Google's algorithms index news outlets and that this indexation determines the rank of news outlets in the search results pages when consumers make a query. The higher news outlets rank in the queries the higher the probability that users will click-through their links and generate visits. 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 100 search results. Our baseline specification is as follows,

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

where $\ln[visits_{it}]$ is the natural logarithm of the number of visits (desktop search visits, total desktop visits and total mobile visits), to news site *i* in day *t*, and $\ln[\ Words\ Top_{it}]$ is the natural logarithm of the number of keywords that the news site *i* has in the top search results (*Words Top 100*, *Words Top 10*, and *Words Top 11-100*) 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 assumption applies to the error term u_{it}.

To account for potential unobserved heterogeneity at the news site level, we use first differences of equation (1) such that

$$\Delta\ln[visits_{it}] = \alpha + \beta\, \Delta\ln[Words\ Top_{it}] + \gamma\Delta X_{it} + \Delta\delta_t + \Delta u_{it}, \quad (2)$$

where we difference out the term α_i and we take care of potential autocorrelation in the error term. All other variables are the result of differences between the contemporaneous variable with realizations of the variable four days before such that $\Delta\ln[y_{it}] = \ln[y_{it}] - \ln[y_{it-4}]$. 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 sites can gather data on keyword volume and trends, keywords targeted by competitors, and can search for combinations of keywords and phrases that generate more visits. They then invest in keywords that can maximize their audience and ad revenues.

In order to deal with this endogeneity problem, we pursue an instrumental variable identification strategy. We need some variable (instrument) that is correlated with the number of keywords that news sites have ranked in Google's top search position but that has no effect on the outlets' search visits other than indirectly through the keywords. The instrument that we use for this objective are the Google's core updates, which can directly modify the news outlets' indexation for each consumer query, but are a source of exogenous variation for the news outlets' visits. We estimate an IV model where the second stage is as (2),

$$\Delta\ln[visits_{it}] = \alpha + \beta\,\Delta\ln[Words\,Top_{it}] + \gamma\Delta X_{it} + \Delta\delta_t + \Delta u_{it}$$

and where the first stage is such that,

$$\Delta\ln[Words\,Top_{it}] = \theta_0 + \theta_1 CoreUpdatePlus7_{it} + \theta_2\Delta X_{it} + \Delta\omega_{it}, \qquad (3)$$

The instrument *Core Update Plus* 7 is a dummy variable that takes value 1 on the day that Google rolls up a core update and in the seven days after that. Our analysis considers the 9 core algorithm updates launched by Google in the period we analyze. This dummy variable is an instrument for the independent variable $\Delta\ln[Words\,Top_{it}]$ under the assumption that $\text{cov}(Core\_update\_plus7, \Delta u_{it}) = 0$. This means that Google core updates are orthogonal to changes in visits (search or total) to a news site *i*. That is, Google does not choose to "roll out" an update because there is a surge in visits to news outlets.

5. Results

5.1 Main Results

This section analyses the effects of Google's algorithm on the search visits of European news outlets. Table 3A uses specification (2) to examine how the number of keywords that news outlets place in Google Search affect their visits. We present two specifications for each of the three outcome variables: *Desktop Search Visits*; *Desktop Total Visits*; and *Mobile Total Visits*. 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 a same outlet.

Columns 1, 3 and 5 consider as independent variable *Words Top 100*, which reflects the number of keywords that news sites place in the first 100 search results in Google. The OLS analysis shows the existence of a positive and significant effect of this variable in the number of visits. Specifically, the results indicate that a 1% increase in the number of keywords generates a 0.05% increase in the number of search visits, and that the increase can rise to 0.079% when we consider the increase in mobile visits. Columns 2, 4 and 6 repeat the previous analysis, but considering as independent variables *Words Top 10* and *Words Top 11-100*. These variables reflect the number of keywords that news sites have in the 1-10 and 11-100 top positions in Google Search results, respectively. In this case, we find that an increase in the number of keywords in the top 10 positions has a positive and larger effect in the outcome variables. However, an increase of keywords in the top 11-100 search positions is associated with a reduction of search visits, and does not have a significant impact in the total desktop visits and total mobile visits.

As explained above, one potential limitation of the previous analysis is that news outlets can use keywords and phrases in order to maximize the visits they receive. For example, they can repeat several times some specific keywords in the headlines and in the contents of their news stories to rank higher in the results for some specific queries. In order 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. Our analysis considers the 9 updates confirmed by Google in the period October, 1 2017 – December 31, 2020 (See Table 1A). Columns 7, 8 and 9 in Table 3A examine the effect of the updates in the number of keywords that news outlets have in the top search results. Columns 7 and 9 show that the dummy variable *Core Update Plus 7* had not a significant effect in the number of keywords in the top 100 search results and in the top 11-100 search results. However, Column 8 reveals that they had an overall negative and significant effect in the number of keywords in the top 10 positions. This implies that the net effect of all the updates was a reduction in the number of keywords that European news outlets had in the top 10 search results.

Table 3B shows 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 variable *Core Update Plus 7* as an instrument for the variables *Word Top* 100, 10, and 11-100. Columns 1-3 show the results for desktop *Search Visits*, columns 4-6 for total *Desktop Visits*, and columns 7-9 for total *Mobile Visits*. The first-stage regressions for the IV estimations are in columns 7, 8 and 9 in Table 3A. We focus here in column 8 as that is the first stage we use in columns 2, 5 and 8 in Table 3B. Our instrument *Core Update Plus 7* only explains variation on *Words Top 10*. The coefficient of our instrument is negative and highly significant.[26] By contrast, in the case of the variables *Words Top 100* and *Words Top 11-100* the instrument is not significant, which implies that Google core updates do not have an impact on this category of keywords. Focusing on the variable *Words Top 10*, the results of the second stage of the IV estimation shows that it 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 6.3% increase in the number of search visits, and a 3.8% increase in the total number of desktop and mobile visits.

So far, our analysis has considered that all Google's core updates are equally relevant. However, as explained in Section 3, each update aims at fixing different aspects of the indexing algorithms, or introduce different features to improve search accuracy. See again Table 1A for a list of the Google's core updates implemented between 2018 and 2020 and that are used in our paper. Considering these, Table 4 examines the effect of the updates grouping them in different ways. Columns 1-3 divide the updates in two groups, the 3 biggest Google core updates according to SEO specialists, and the remaining 6 non-big core updates.[27] In contrast to the results of Table 3A, we find that "big core updates" had a positive and significant impact on *Words Top 100* and "non-big core updates" had a negative effect. Moreover, if we split keywords between those ranked in top 10 and in top 11-100 positions, we obtain that big core updates had a negative impact on *Words Top 10* and a positive effect on *Words Top 11-100*. These results suggest that big updates moved news outlets' links from the top 10 to 11-100 positions, and that non-big updates generated a general reduction of keywords in top search results for news outlets.

The table also considers the effects of other updates that have been confirmed by Google, but that the firm does not consider as core updates (hereafter "non-core updates"). Columns 4-6 repeat the previous analysis but including as a control variable the Google's "non-core updates". We find that the main insights from the previous analysis are confirmed, and we also obtain that non-core updates had a negative and significant effect on the number of keywords ranked in top positions.

[26] Likewise, the F-test of excluded instruments is 57 and highly significant. We are also able to reject the null hypothesis that the model is underidentified (Chi-sq=52) and reject the null of weak instruments (Cragg-Donald Wald F statistic = 126.75 and Kleibergen-Paap Wald rk F statistic = 57.12).

[27] 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.

Finally, columns 7-9 examine the individual impact of each core algorithm update. This analysis reveals the heterogeneous effects of the updates, regarding both their direction and magnitude. If anything, we find that each individual core updates has a homogenous effect in the number of keywords in top 10 and top 11-100 positions. Notice that the update that had a higher impact was rolled out in March 2019 (not considered a big update by industry specialists). This effect was later compensated with the update of June 2019 and more importantly with the update of September 2019. In 2020, the updates of January and May had a negative effect in the number of keywords that was partly compensated by the update of December. To sum up, our analysis reveals that core updates might have different effects in the number of keywords ranked in top positions for each news outlet, and that the effects of each individual update are similar for the number keywords in top 10 and top 11-100 positions.

We complete our analysis with Table 5, which shows the results of the IV estimation of the model in equation (3), when we consider as instruments the "big" and "non-big" core updates. The IV regressions use as a first stage the results in columns 1-3 in Table 4.[28] The results confirm our previous finding. First, the variable *Words Top 10* has a positive and significant effect in the number of search visits, total desktop visits and mobile visits. Second, the variables *Words Top 100* and *Words Top 11-100* have a positive and significant in effect in the number of *Search Visits*.

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 search results. Second, the number of keywords that news outlets have in the top 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 6 repeats the IV estimations of Table 5 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 core updates" and "non-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*,

[28] Here again the first stage regression is sound. 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.

total *Desktop Visits* and total *Mobile Visits*. Columns 1 and 2 in Table 6 show the results of the first stage estimation. Although results vary across national markets, in most countries we find evidences that "big" and "non-big" core updates had a negative effect in the number of keywords that news outlets had in the top 10 search results. Columns 3, 4 and 5 present the results of the second stage regressions for the three outcome variables. Results for search visits are ambiguous. We find a positive relationship between *Word Top 10* and the number of visits in Demark, Poland and Spain, and a negative relationship in Greece and the UK.

Tables 7 and 8 repeat the previous analysis but classifying news outlets in different ways. In Table 7, 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 non-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 *Search Visits*, and in the number of total *Desktop Visits* and *Mobile Visits*.

Table 8 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 "non-big" core updates and *Word 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*).

6. Market Concentration Effects of Google Core Updates

The objective of this section is to analyze the effects of Google core updates on the concentration of European media markets. The analysis of the previous section has shown that one consequence of Google's recent core updates has been the reduction of news outlets' keywords in top positions, and the subsequent reduction in search visits. Now we want to examine whether this reduction has been more important for large news outlets than for small ones, and if the result of this situation has been a reduction in market concentration. 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 taking into account 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 core updates and non-big core update dummies. All specifications include month, year, day of the week FE and changes in the number of direct visits as controls.

Figure 4 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 (international football competitions, covid pandemic).

Tables 9 and 10 show the results of the estimation of equation (4) to examine if core updates are reinforcing market concentration. Table 9 shows the effects of "big" and "non-big" core updates for the whole sample of news outlets and for each individual country. 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 non-big core updates.[29] If we now consider the effects of core updates at the individual country level, we find that results are quite heterogeneous. Big core updates had a negative effect in Finland, Germany and Greece, but a positive effect in Portugal. Non-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 10 analyzes the effect of Google's core updates considering the impact in different types of news outlets. The results reveal that "big" updates did no generate any effect in the concentration of national markets. In contrast, "non-big" updates increased market concentration of search visits for *National* news outlets, and they reduced the concentration for *Sports* news outlets. This suggest that the reduction in the number of keywords ranked in top positions as a result of 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.

[29] 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

7. Conclusions

Search engines are crucial intermediaries to access the news contents available in the Internet. Consumers frequently look for the latest news in Google, Bing or Yahoo, rather than directly visiting on line 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 news stories to rank higher in the search results on more keywords, but also about the risk that some publishers can become too large and therefore too influential.

Our paper constitutes a first step to study these questions by examining how Google Search affects the concentration of the European media markets. 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 some news contents, Google identifies all the web pages that can offer a precise answer to it and indexes them in its search results page. Considering this, news outlets invest in the keywords that can generate more visits and that allow them to rank higher in Google's indexation. In order to isolate the effects that Google's indexation has on the 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. This reduction in the visibility of news outlets could have been compensated by the growth in the number of queries formulated by consumers. We also obtain that the number of keywords that news outlets have in top search results pages have a positive effect in their visits. Specifically, we obtain 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 a 0.08% increase of market concentration due to the effect of the "non-big" core updates. In addition, we have explained that non-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. Non-big core updates increased concertation in Finland and Netherlands. 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.

These results have important implications for policy makers interested in understanding the effects of search engines in the competition of online markets. We have seen that search engines' indexation algorithms have a crucial effect in the commercial success of retailers and content providers. For this reason, it is important to be aware of the effects that algorithm updates can have on competition. 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).[30] In addition, in December 2020, the EU proposed more instruments to regulate online intermediaries, through the Digital Services Act (DSA) and the Digital Markets Act (DMA). Similar initiatives are taking place in other parts of the world.

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, and Shapiro, 2017; Gentzkow, 2018; Goldfard and Tucker, 2019; Claussen et al., 2021).[31] 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).

[30] 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).

[31] 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.

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.

7. 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).

Bakshy, E., S. Messing and L.A. Adamic (2015), Exposure to ideologically diverse news and opinion on facebook. Science, 348 (6239), 1130–1132.

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

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 (2018), Streaming platform and strategic recommendation bias.

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 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.

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.

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.

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

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.

Edelman, B. (2011), Bias in search results: Diagnosis and response, Indian JL & Tech., 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.

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.

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.

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).

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.

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. 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.

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

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

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.

Zhu, F. and Q. Liu (2018), Competing with complementors: An empirical look at Amazon. com, Strategic management journal, 39, 2618-2642.

Figure 1. Google's announcement of May 2020 Core Update

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.
131 528 516

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)

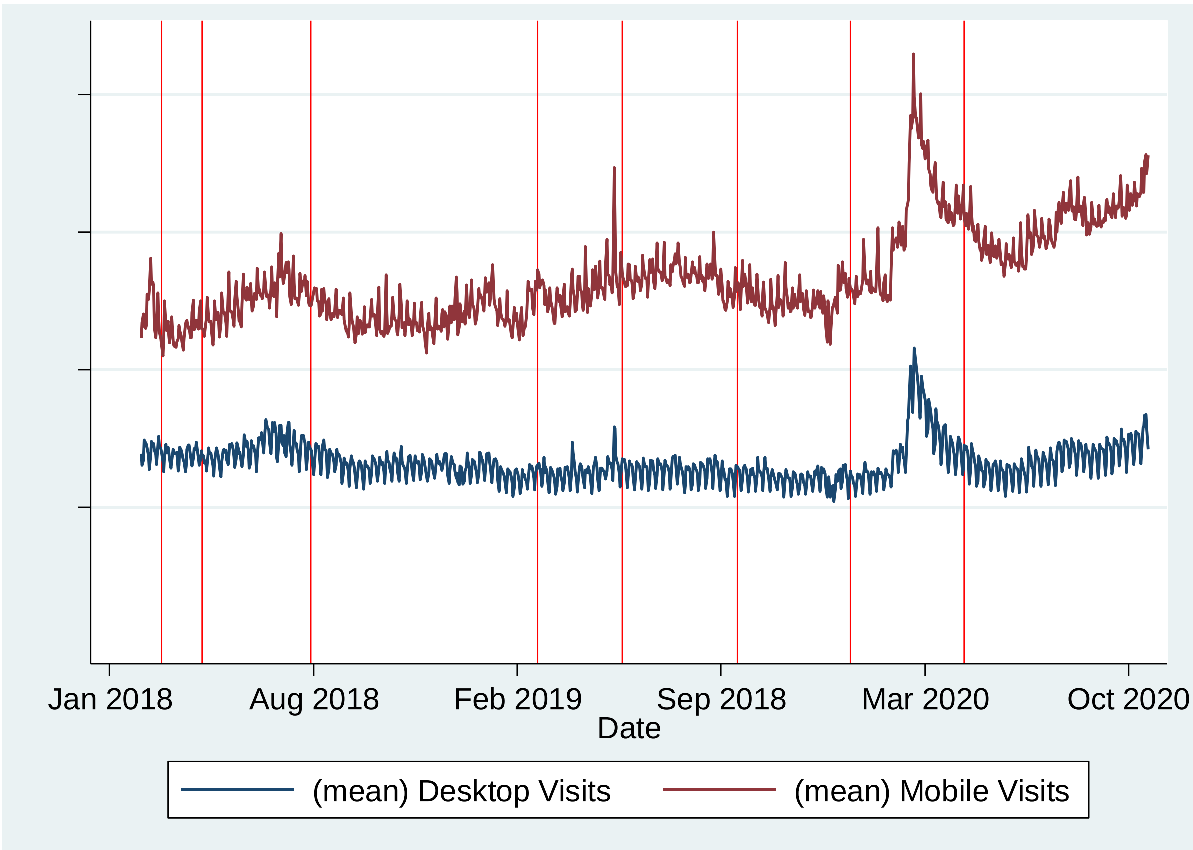

Figure 3. Share of Desktop Direct, Search and Social Networks Daily Visits (October 2017-December 2020)

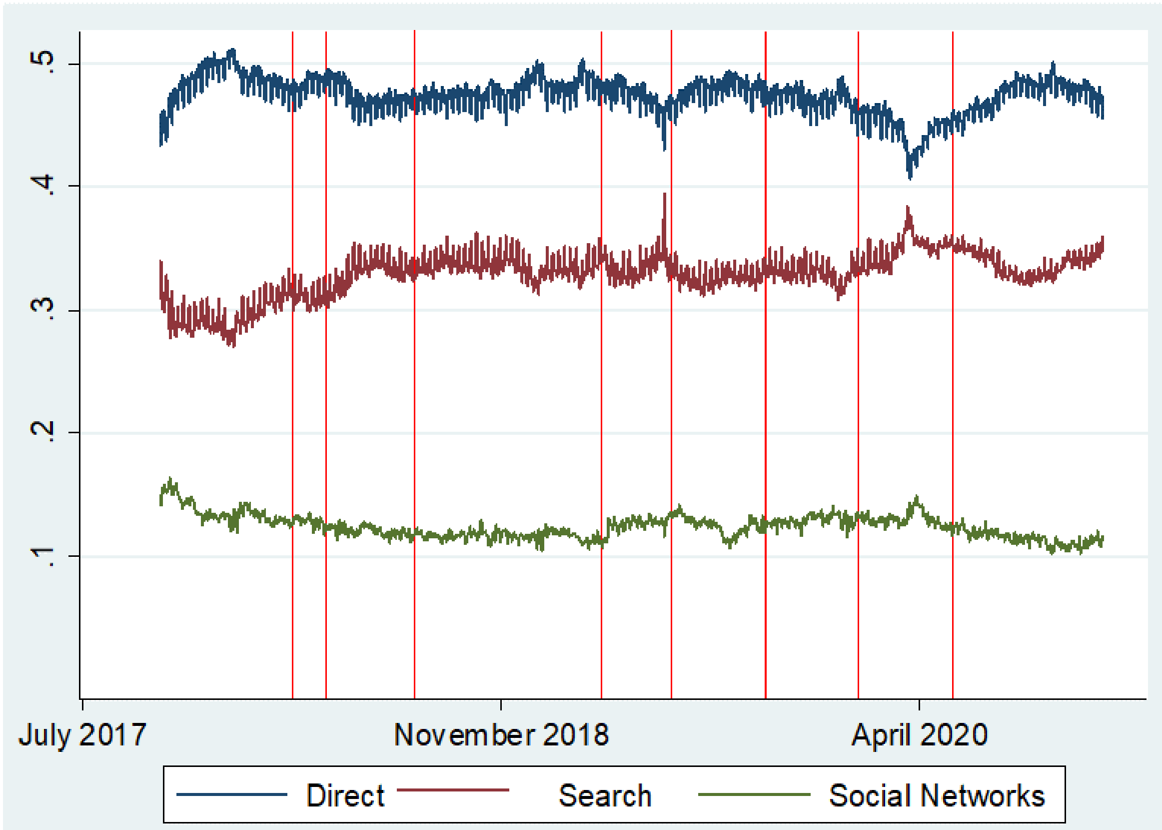

Figure 4: Example of Google Search's page results

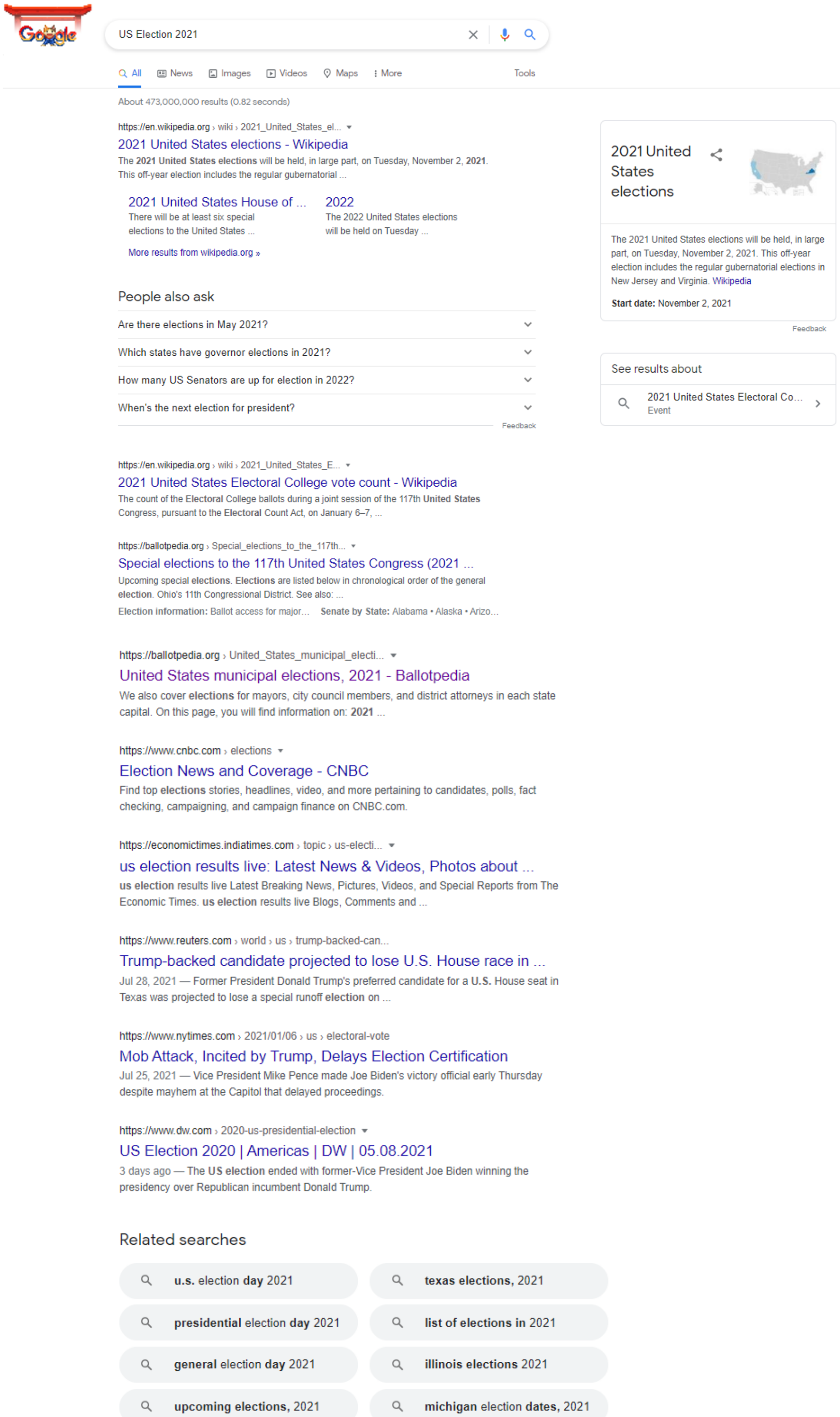

Figure 5. HHI for Desktop Search Visit, Total Desktop Visits and Total Mobile Visits (January 2018 - November 2020)

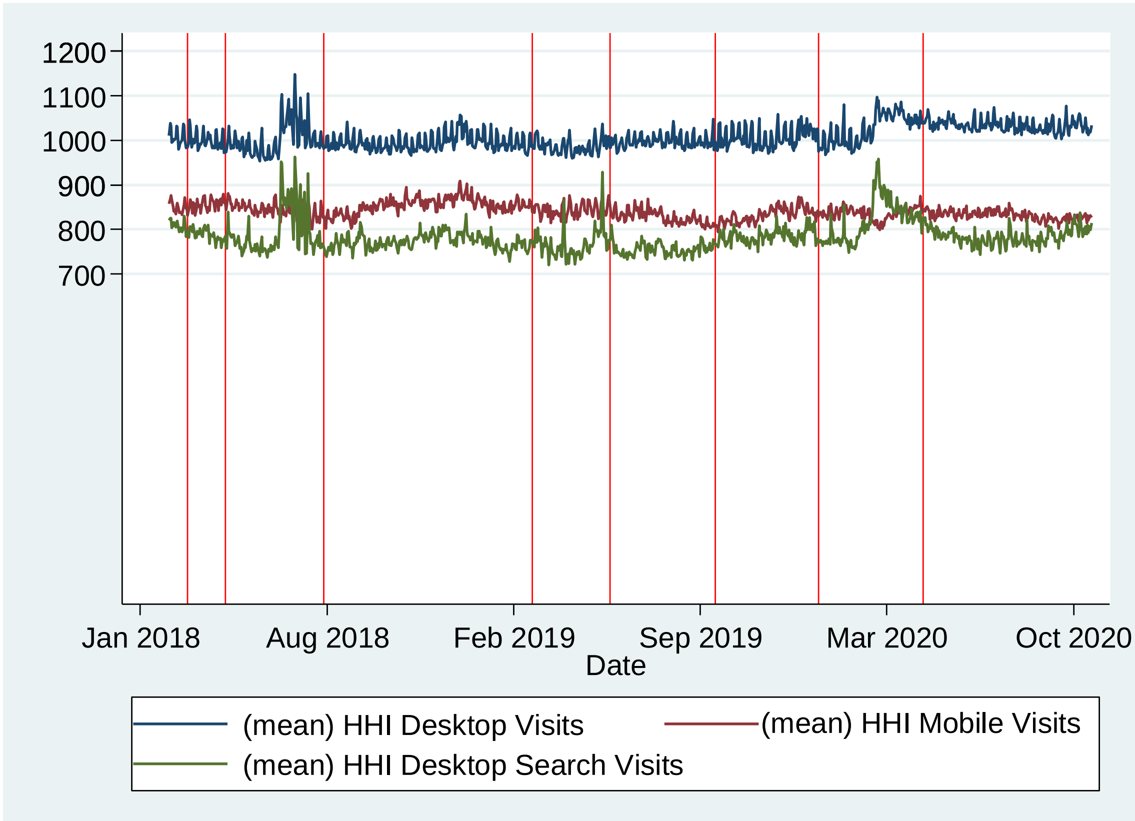

Table 1A. Google's confirmed core updates

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.[32]
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.[33]
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

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

[33] 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	676,070	141479.5	257851.8
Mobile Visits	630,212	288258.7	511417.2
Desktop Search Visits	674,609	43466.61	79207.39
Desktop Direct Visits	675,619	77498.82	164839.1
Keywords Top 1-100	653,315	777894	1231113
Keywords Top 1-10	653,315	88148.8	166258.9
Keywords Top 11-100	653,315	689745.2	1081249
National	680,641	0.298	0.457
Regional	680,641	0.313	0.464
Sports	680,641	0.109	0.312
Business	680,641	0.116	0.320
Radio/TV	680,641	0.131	0.337
Google Updates			
Core Update +7	680,641	0.050	0.217
Big Core Update +7	680,641	0.019	0.135
Non-Big Core Update +7	680,641	0.031	0.174
Non Core Update +7	680,641	0.105	0.306
Concentration Measures			
HHI Mobile Visits	17,117	916.992	1127.473
HHI Desktop Visits	17,117	1128.977	1077.180
HHI Search Visits	17,117	831.399	756.295
HHI Mobile Visits per segment	96,007	3490.983	2836.643
HHI Desktop Visits per segment	96,007	3955.063	2720.061
HHI Search Visits per segment	96,007	3543.197	2647.790

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

Table 3A. First Differences OLS Regressions of Search Visits, Total Visits and Mobile Visits on the Number of Key Words and Google Core Updates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent Variable	**Δln(Search Visits) t-4**	**Δln(Search Visits) t-4**	**Δln(Desktop Visits) t-4**	**Δln(Desktop Visits) t-4**	**Δln(Mobile Visits) t-4**	**Δln(mobile visits) t-4**	**Δln(Words top 100) t-4**	**Δln(Words top 10) t-4**	**Δln(Words top 11-100) t-4**
Δln(Words top 100) t-4	0.05037035***		0.03806090***		0.07934895***				
	(0.0155)		(0.0121)		(0.0225)				
Δln(Words top 10) t-4		0.11698244***		0.04269657*		0.13310628***			
		(0.0217)		(0.0249)		(0.0338)			
Δln(Words top 11-100) t-4		-0.04627326**		0.00408977		-0.00939223			
		(0.0201)		(0.0312)		(0.0476)			
Google Core Update t to t+7							0.00012228	-0.00179259***	0.00020906
							(0.0002)	(0.0002)	(0.0002)
Δln(Desktop Direct Visits) t-4	0.34438175***	0.34437259***	0.60205561***	0.60205180***	0.39134531***	0.39135336***	0.00011409	0.00010184	0.00010638
	(0.0257)	(0.0258)	(0.0242)	(0.0242)	(0.0263)	(0.0263)	(0.0001)	(0.0001)	(0.0001)
Constant	0.03372459***	0.03303768***	0.00011051	-0.00013574	0.10042736***	0.09963768***	0.00229709***	0.00745421***	0.00161839**
	(0.0072)	(0.0072)	(0.0050)	(0.0050)	(0.0082)	(0.0082)	(0.0008)	(0.0008)	(0.0008)
Week FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Day of Week FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	644 469	644 463	645 597	645 589	597 968	597 962	645 597	645 589	645 597
R-squared	0.26	0.26	0.64	0.64	0.16	0.16	0.02	0.04	0.02

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

Table 3B. First Differences Regressions Using Google Core Updates as Instrumental Variables for Changes in the Number of Keywords

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent Variable	**Δln(Search Visits) t-4**	**Δln(Search Visits) t-4**	**Δln(Search Visits) t-4**	**Δln(Desktop Visits) t-4**	**Δln(Desktop Visits) t-4**	**Δln(Desktop Visits) t-4**	**Δln(Mobile Visits) t-4**	**Δln(Mobile Visits) t-4**	**Δln(Mobile Visits) t-4**
Δln(Words top 100) t-4	-98.69234104			-56.64347852			-20.37670929		
	(201.9679)			(110.5324)			(18.0511)		
Δln(Words top 10) t-4		6.34907243***			3.86399755***			3.89321761***	
		(1.2008)			(0.8585)			(1.4560)	
Δln(Words top 11-100) t-4			-56.28066492			-33.1315013			-16.17003479
			(69.3983)			(40.0621)			(12.2789)
Δln(Desktop Direct Visits) t-4	0.35533840***	0.34367783***	0.35016865***	0.60849797***	0.60163826***	0.60555991***	0.39388178***	0.39100981***	0.39320979***
	(0.0376)	(0.0258)	(0.0282)	(0.0281)	(0.0242)	(0.0249)	(0.0262)	(0.0264)	(0.0262)
Constant	0.26056172	-0.01347421	0.12488683	0.13034227	-0.02858516***	0.05384689	0.12843672***	0.07392431***	0.11066460***
	(0.4779)	(0.0129)	(0.1248)	(0.2620)	(0.0084)	(0.0723)	(0.0343)	(0.0137)	(0.0203)
Week FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Day of Week FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	644,469	644,463	644,469	645,597	645,589	645,597	597,968	597,962	597,968
R-squared					0.55			0.09	

Note columns 7, 8 and 9 in Table 2A are first stages for columns here 1, 4 and 7; 2, 5 and 8; and 3, 6 and 9, respectively.
Robust standard errors clustered at the domain level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 4. Differences Across Core Updates and Non-Core Updates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent Variables	**Δln(Words top 100) t-4**	**Δln(Words top 10) t-4**	**Δln(Words top 11-100) t-4**	**Δln(Words top 100) t-4**	**Δln(Words top 10) t-4**	**Δln(Words top 11-100) t-4**	**Δln(Words top 100) t-4**	**Δln(Words top 10) t-4**	**Δln(Words top 11-100) t-4**
"Big" Google Core Update t+7	0.00165930***	-0.00075434*	0.00170706***	0.00190337***	-0.00051195	0.00196076***			
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)			
"Not Big" Google Core Update t+7	-0.00093630***	-0.00250766***	-0.00082265***	-0.00110596***	-0.00267616***	-0.00099900***			
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)			
Core Update December 2020 t+7							0.00382522***	0.01222836	0.00291446***
							(0.0015)	(0.0083)	(0.0010)
Core Update May 2020 t+7 #							-0.00104945	-0.00800046***	-0.00037177
							(0.0008)	(0.0008)	(0.0008)
Core Update January 2020 t+7							-0.00392604***	-0.00810493***	-0.00365368***
							(0.0005)	(0.0006)	(0.0005)
Core Update September 2019 t+7							0.01359838***	0.02446748***	0.01260113***
							(0.0007)	(0.0008)	(0.0007)
Core Update June 2019 t+7 #							0.00707605***	0.00403478***	0.00743705***
							(0.0005)	(0.0006)	(0.0006)
Core Update March 2019 t+7							-0.01034261***	-0.01610592***	-0.01007042***
							(0.0006)	(0.0007)	(0.0007)
Core Update August 2018 t+7 #							0.00011309	0.00283611***	-0.0007327
							(0.0008)	(0.0008)	(0.0008)
Core Update April 2018 t+7							-0.0012382	-0.00186781*	-0.00105773
							(0.0011)	(0.0011)	(0.0011)
Core Update March 2018 t+7							-0.00482907***	-0.01307236***	-0.00407381***
							(0.0004)	(0.0005)	(0.0005)
Non-Core Google Update				-0.00221744***	-0.00220230***	-0.00230498***	-0.00248912***	-0.00244460***	-0.00259259***
				(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Δln(Desktop Direct Visits) t-4	0.00011361	0.00010151	0.00010591	0.00011609	0.00010398	0.00010848	0.00009817	0.00007355	0.00009124
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Constant	0.00230211***	0.00745760***	0.00162328**	0.00242542***	0.00758008***	0.00175146**	0.00239818***	0.00749925***	0.00173081**
	(0.0008)	(0.0008)	(0.0008)	(0.0008)	(0.0008)	(0.0008)	(0.0008)	(0.0008)	(0.0008)
Week FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Day of Week FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	645,597	645,589	645,597	645,597	645,589	645,597	645,597	645,589	645,597
R-squared	0.02	0.04	0.02	0.02	0.04	0.02	0.03	0.05	0.03

We denote with # 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 5. First Differences Regressions Using Google Core Update Heterogeneity s as Instrumental Variables for Changes in the Number of Keywords

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent Variable	**Δln(Search Visits) t-4**	**Δln(Desktop Visits) t-4**	**Δln(Mobile Visits) t-4**	**Δln(Search Visits) t-4**	**Δln(Desltop Visits) t-4**	**Δln(Mobile Visits) t-4**	**Δln(Search Visits) t-4**	**Δln(Desktop Visits) t-4**	**Δln(Mobile Visits) t-4**
Δln(Words top 100) t-4	4.5169***	1.6302	1.7529						
	(1.4600)	(0.9994)	(1.4963)						
Δln(Words top 10) t-4				6.6376***	3.7405***	4.1166***			
				(1.1211)	(0.7467)	(1.4006)			
Δln(Words top 11-100) t-4							3.8913***	1.2498	1.4390
							(1.4180)	(0.9965)	(1.4756)
Δln(Desktop Direct Visits) t-4	0.3439***	0.6019***	0.3911***	0.3436***	0.6017***	0.3910***	0.3440***	0.6019***	0.3912***
	(0.0257)	(0.0242)	(0.0263)	(0.0258)	(0.0242)	(0.0264)	(0.0257)	(0.0242)	(0.0263)
Constant	0.0235***	-0.0035	0.0981***	-0.0156	-0.0277***	0.0724***	0.0275***	-0.0018	0.0996***
	(0.0088)	(0.0056)	(0.0088)	(0.0126)	(0.0079)	(0.0138)	(0.0082)	(0.0053)	(0.0084)
Week FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Day of Week FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	644,469	645,597	597,968	644,463	645,589	597,962	644,469	645,597	597,968
R-squared	0.13	0.63	0.15		0.55	0.08	0.16	0.63	0.15

First Stage of columns 1, 4 and 7 is column 1 in Table 3. First Stage of columns 2, 5 and 8 is column 2 in Table 3. First Stage of columns 3, 6 and 9 is column 2 in Table 3.
Robust standard errors clustered at the domain 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 country

	FIRST 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 "Small" Core Update**	**β Δln(KeyWords TOP10) t-4**	**β Δln(KeyWords TOP10) t-4**	**β Δln(KeyWords TOP10) t-4**
Austria	0.0004 (0.0012)	-0.0006 (0.0013)	18.5679 (25.6457)	-0.0515 (10.2516)	1.1297 (23.5141)
Belgium	-0.0006 (0.0021)	-0.0046*** (0.0011)	2.3189 (1.6237)	2.7453** (1.0280)	5.1430* (2.5972)
Denmark	-0.0030* (0.0016)	-0.0042*** (0.0009)	7.0638** (2.7906)	1.4451 (1.8946)	
Finland	0.0014 (0.0012)	-0.0027** (0.0011)	-3.2330 (1.9556)	-4.1269** (2.0029)	-0.3457 (5.0669)
France	-0.0034*** (0.0012)	0.0019 (0.0032)	-7.8110 (5.6929)	-4.0528 (3.0022)	-5.6614 (4.8236)
Germany	-0.0005 (0.0012)	-0.0014* (0.0008)	6.8452 (6.0689)	5.7789 (3.8744)	-1.4815 (6.1417)
Greece	0.0085*** (0.0020)	-0.0030*** (0.0011)	-3.0628* (1.8023)	-2.3780** (1.0637)	-2.1888* (1.1690)
Ireland	-0.0074*** (0.0021)	-0.0066*** (0.0016)	-0.0312 (0.5429)	-0.3740 (0.8609)	1.1909 (1.2613)
Italy	0.0002 (0.0012)	-0.0001 (0.0009)	46.9613 (204.4139)	16.7190 (74.5345)	-66.1106 (293.4514)
Netherlands	-0.0034*** (0.0012)	-0.0032*** (0.0008)	0.7125 (1.3368)	1.9045** (0.8993)	-3.0322 (2.5607)
Poland	-0.0032*** (0.0010)	-0.0034*** (0.0009)	10.3908*** (2.1928)	6.0095*** (1.5756)	7.8332*** (2.2501)
Portugal	0.0019 (0.0018)	-0.0041*** (0.0013)	2.6860 (2.4840)	1.8328 (1.5616)	6.1433 (4.2944)
Spain	0.0013 (0.0009)	-0.0047*** (0.0006)	6.9571*** (1.0232)	3.2958*** (0.7012)	2.3239 (1.4320)
Sweden	-0.0045*** (0.0015)	-0.0008 (0.0008)	0.5642 (1.3304)	0.7596 (1.5677)	-1.9122 (2.3194)
United Kingdom	-0.0030** (0.0014)	-0.0024** (0.0009)	-4.4395* (2.3095)	-3.4895* (1.8522)	1.3677 (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 non-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 domain level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 7. Impact of Google Core Updates on number of search, desktop and mobile visits per domain type

	FIRST 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 "Small" 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 DOM%	-0.0013**	-0.0028***	5.7484***	3.4065***	5.9952***
	(0.0005)	(0.0004)	(1.2311)	(0.8461)	(1.7329)
BOT DOM %	-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 domain (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 small big core update dummies. Then for each type of domain, 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 domain level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 8. Impact of Google Core Updates on number of search, desktop and mobile visits per domain

	FIRST 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 "Small" 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 domain (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 small big core update dummies. Then for each type of domain, 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 domain level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 9. 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	
Coefficient	β "Big" Core Update	β "Small" Core Update	β "Big" Core Update	β "Small" Core Update	β "Big" Core Update	β "Small" 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 determine 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" core update and "small" core update. 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 10. Impact of Google Core Updates on HHI per news outlet segment and countr9y

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 "Small" Core Update**	**β "Big" Core Update**	**β "Small" Core Update**	**β "Big" Core Update**	**β "Small" Core Update**
NATIONAL	-0.0081 (0.0047)	0.0095*** (0.0021)	0.0005 (0.0024)	0.0039** (0.0017)	-0.0036 (0.0040)	0.0017 (0.0032)
REGIONAL	-0.0055 (0.0041)	0.0046 (0.0028)	-0.0024 (0.0038)	-0.0018 (0.0027)	-0.0052 (0.0076)	0.0085 (0.0082)
BUSINESS	0.0028 (0.0065)	-0.0036 (0.0050)	0.0062 (0.0046)	0.0015 (0.0045)	0.0061 (0.0076)	-0.0069 (0.0062)
SPORTS	0.0008 (0.0045)	-0.0067** (0.0026)	0.0077 (0.0058)	0.0003 (0.0028)	0.0027 (0.0049)	-0.0008 (0.0048)
TV/RADIO	0.0056 (0.0064)	0.0064 (0.0041)	0.0015 (0.0048)	0.0007 (0.0029)	0.0010 (0.0061)	-0.0034 (0.0067)

This table contains results of 15 different regressions. For each type of domain (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 non-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 domain level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.