Journal Pre-proofs

Full length article

The impact of the General Data Protection Regulation (GDPR) on online tracking

Klaus M. Miller, Karlo Lukic, Bernd Skiera

PII: S0167-8116(25)00022-9
DOI: https://doi.org/10.1016/j.ijresmar.2025.03.002
Reference: IJRM 1649

To appear in: *International Journal of Research in Marketing*

Received Date: 24 October 2023

Please cite this article as: K.M. Miller, K. Lukic, B. Skiera, The impact of the General Data Protection Regulation (GDPR) on online tracking, *International Journal of Research in Marketing* (2025), doi: https://doi.org/10.1016/j.ijresmar.2025.03.002

The Impact of the General Data Protection Regulation (GDPR) on Online Tracking

Klaus M. Miller[1], Karlo Lukic, Bernd Skiera

========

ARTICLE INFO

Article history

First received on 24 October 2023 and was under review for 6 months.

Area Editor: Peter C. Verhoef

Accepting Editor: Koen Pauwels

=======

Klaus M. Miller, Assistant Professor of Marketing, HEC Paris, Rue de la Libération 1, 78350 Jouy-en-Josas, France, Hi!PARIS Chairholder, Center in Data Analytics & AI for Science, Technology, Business & Society, Phone +33-1-39-67-70-88, millerk@hec.fr.

Karlo Lukic, Post-Doctoral Researcher, Department of Marketing, Faculty of Economics and Business, Goethe University Frankfurt, Theodor-W.-Adorno-Platz 4, 60323 Frankfurt, Germany, Phone +49-17643234998, lukic@wiwi.uni-frankfurt.de.

Bernd Skiera, Professor of Electronic Commerce, Department of Marketing, Faculty of Economics and Business, Goethe University Frankfurt, Theodor-W.-Adorno-Platz 4, 60323 Frankfurt, Germany, Phone +49-69-798-34649, skiera@wiwi.uni-frankfurt.de); also Professorial Research Fellow at Deakin Business School, 221 Burwood Highway, Burwood, VIC 3125, Australia (bernd.skiera@deakin.edu.au).

Acknowledgments:

The authors thank the WhoTracks.me team for providing the data for this research project; participants of the 43rd Annual INFORMS Society for Marketing Science (ISMS) Conference 2021, Munich Summer Institute (MSI) 2021, 19th ZEW Conference on the Economics of Information and Communication Technologies 2021, 17th Symposium on Statistical Challenges in Electronic Commerce Research (SCECR) 2021, Interactive Marketing Research Conference (IMRC) 2021, NBER Economics of Privacy Conference, Spring 2022, European Marketing Academy (EMAC) Conference 2022, 44th Annual INFORMS Society for Marketing Science

[1] Corresponding author.

(ISMS) Conference 2022, 20th ZEW Conference on the Economics of Information and Communication Technologies 2022, VfS Annual Conference 2022, and 2nd DPSN International Data Protection Day 2023; and seminar participants at Goethe University Frankfurt for helpful comments.

Funding:

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 833714). Miller gratefully acknowledges support from the Hi! PARIS Center on Data Analytics and Artificial Intelligence for Science, Business, and Society.

Declarations of interest: none.

The Impact of the General Data Protection Regulation (GDPR) on Online Tracking

Abstract

This study explores the impact of the General Data Protection Regulation (GDPR) on online trackers—vital elements in the online advertising ecosystem. Using a difference-in-differences approach with a balanced panel of 294 publishers, it compares publishers subject to the GDPR with those unaffected (the control group). Drawing on data from WhoTracks.me, which spans 32 months from May 2017 to December 2019, it analyzes how the number of trackers used by publishers changed before and after the GDPR. The findings reveal that although online tracking increased for both groups, the rise was less significant for EU-based publishers subject to the GDPR. Specifically, the GDPR reduced about four trackers per publisher, equating to a 14.79% decrease compared to the control group. The GDPR was particularly effective in curbing privacy-invasive trackers that collect and share personal data, thereby strengthening user privacy. However, it had a limited impact on advertising trackers and only slightly reduced the presence of analytics trackers.

Keywords: Economics of Privacy; Online Privacy; Online Tracking; Privacy Law; Quasi-Experiment; Difference-in-Differences; Synthetic Control

1. Introduction

Online advertising is a cornerstone of the modern digital economy, enabling publishers to monetize their content by delivering targeted advertisements to users. Central to online advertising are online trackers (hereafter, "trackers"), which raise substantial privacy concerns because they process personal data—often without the user's explicit knowledge or consent (see, e.g., Beke et al. 2018; Bleier et al. 2020; Lobschat et al. 2021; Verhoef et al. 2022). In response, the European Union (EU) enacted the General Data Protection Regulation (GDPR) in May 2018 to grant individuals more control over their personal data.

Although the GDPR aims to protect users' privacy by requiring explicit consent and limiting data collection to specific purposes, it remains uncertain whether it has curbed the most privacy-invasive trackers. Existing literature has extensively explored user privacy concerns (e.g., Eggers et al. 2023; Beke et al. 2022; Wieringa et al. 2021), the functioning and evolution of the market for trackers (e.g., Mayer and Mitchell 2012; Lerner et al. 2016; Karaj et al. 2018a), and the short-term impacts (up to six months) of privacy regulations like the GDPR on online advertising and tracker usage (e.g., Peukert et al. 2022; Johnson et al. 2023). However, questions persist about whether the GDPR has sustainably reduced publishers' reliance on trackers from various categories, such as trackers that pose higher privacy risks over a more extended period (up to 20 months) and how these shifts affect key actors—users, publishers, advertisers, and tracker providers.

Therefore, we examine the period before and after the GDPR's enactment to study whether the regulation achieves its intended aim of mitigating highly privacy-invasive tracking and to explore any unintended consequences for online advertising. More specifically, we address the following three research questions:

- **RQ1:** What is the effect of the GDPR on the number of trackers?
- **RQ2:** Does the effect of the GDPR differ across categories of trackers (e.g., advertising vs. analytics)? In particular, does the GDPR impact categories with higher privacy risks more strongly than those with lower privacy risks?
- **RQ3:** Does the effect of GDPR differ across publisher types (news vs. non-news) and tracker providers (large vs. small)?

To answer these questions, we draw on balanced panel data from WhoTracks.me covering 294 publishers over 32 months and, in the Web Appendix, an unbalanced panel of 29,735 publishers. We employ a difference-in-differences (DiD) analysis, comparing EU publishers subject to the GDPR with a control group of non-EU publishers to derive the effect of the GDPR. We also explore the heterogeneity of the GDPR's effects across different tracker categories, publisher types and tracker providers.

Understanding these consequences is crucial, as they can inform various actors about the effectiveness of the GDPR and guide future privacy policies. For instance, identifying unintended consequences such as increased market concentration or disproportionate impacts on certain publishers can help regulators refine the regulation to achieve its objectives.

2. Description of Online Tracking and the Impact of GDPR

2.1. Definition, Purpose, and Tracking Functionality of Online Trackers

A tracker is software developed by a tracker provider (e.g., Google) that collects information about a user's activities online. The installing actor—a user, publisher, or advertiser—uses the tracker for a particular purpose. Publishers, for example, use trackers to learn about what their users do ("analytics"), display advertising, integrate social media, and enhance the user's experience.

The tracking functionality of a tracker refers to the specific activities it performs when fulfilling its purpose. These activities involve collecting, retaining, using, or sharing data—particularly user data—often across multiple contexts. For example, an analytics tracker may collect user data, such as browsing history and interaction patterns, to provide insights into how users engage with a publisher's website.

Trackers often bundle their purpose with tracking functionality. For example, a user might install the Google Translate browser extension in their browser to view translations instantly as they browse the web. This extension allows the user to access convenient translation services; however, it also includes tracking functionality that enables the tracker provider and sometimes also the publisher or advertiser to collect, retain, or share data on how the user interacts with the extension, sometimes even without the user's explicit awareness.

Similarly, a publisher might install Google AdSense on its website to display targeted ads to users via the Google ad network. Because the publisher and its advertisers want to know which users clicked on ads, ad serving includes tracking functionality that enables such ad measurement on the publisher's website. Thus, tracking increases the value of online advertising for firms but also raises privacy concerns for users. Addressing this conflict is essential for creating a balanced and privacy-conscious online environment.

2.2. Actors Involved in Online Tracking

Trackers sit at the center among four main actors in online advertising: users, publishers, advertisers, and tracker providers. Figure 1 illustrates how these actors interact.

<Add "Figure 1: Main Actors Involved in Online Tracking" about here>

Users access content provided by publishers and interact with trackers embedded in websites. Publishers use trackers to monetize content better and enhance user experience. Advertisers rely on trackers to collect data for targeted advertising and ad performance measurement. Tracker providers supply the technology that enables data collection and targeted advertising, benefiting all parties but also raising privacy concerns for users.

2.3. GDPR as an Attempt to Increase Online Privacy

The GDPR, enacted on May 25th, 2018, is an online privacy law applicable to all EU member states (European Commission 2016). The GDPR aims to increase users' online privacy by strengthening their control over personal data (e.g., Skiera et al. 2022), which the European

Commission (2016) defines as "any information relating to an identified or identifiable natural person" (Article 4). Since online tracking involves creating unique identifiers, such as IP addresses and cookie identifiers, representing individuals, any information gathered is considered personal data under the GDPR and is subject to its stipulations.

The GDPR imposes obligations on publishers to conduct Data Protection Impact Assessments (DPIAs), which audit their data processing practices. These assessments require publishers to examine how personal data is collected, used, and shared and to determine whether these practices meet the GDPR standards. Publishers can legally track user data only (i) if the user has explicitly provided consent to be tracked, (ii) if tracking is necessary for providing the requested service, or (iii) if a legitimate interest exists that justifies tracking (Article 6, 7). As a result, publishers may choose to remove or adjust specific trackers to reduce the risk of non-compliance.

Previous EU privacy laws only affected firms based in the EU. In contrast, GDPR applies to all firms that process EU users' personal data. It only treats EU and non-EU firms differently in processing non-EU users' personal data; in that case, the GDPR applies to EU firms but does not apply to non-EU firms. Thus, a non-EU publisher must comply with the GDPR when processing data of EU users but is not obligated to apply the same standards for non-EU users (European Data Protection Board 2018).

2.4. Categorization of Online Trackers

We categorize trackers across five dimensions: purpose, necessity, tracking functionality, type of publisher, and size of tracker provider. These dimensions capture critical aspects relevant to publishers and regulators, particularly under the GDPR, and allow for a compelling description of online advertising and an understanding of how the GDPR has impacted it.

2.4.1. Online Trackers by Purpose and Necessity

Trackers serve various purposes, such as analytics, advertising, social media integration, and consent management. Categorizing trackers by their purpose helps us understand their value for publishers.

Beyond purpose, we also classify trackers based on necessity, distinguishing between essential and non-essential trackers. Essential trackers support a website's core functionality, ensuring it operates as intended. In contrast, non-essential trackers enhance user experience or provide additional insights but are not critical for the website's functionality. Under the GDPR, only non-essential trackers require user consent, whereas essential trackers do not (European Data Protection Board 2019).

This categorization by necessity helps us assess whether the GDPR achieves its intended consequence: reducing the use of non-essential trackers. From a societal perspective, essential trackers are acceptable since they serve to provide the services users request. Reducing non-essential trackers reduces privacy risks. Table 1 categorizes trackers by purpose and necessity (Karaj et al. 2018a).

<Add "Table 1: Categorization of Online Trackers by Purpose and Necessity" about here>

2.4.2. Online Trackers by Tracking Functionality

We categorize trackers based on whether they collect personal data, share personal data, or both. This categorization aligns closely with the GDPR's emphasis on collecting, processing, and protecting users' personal data. Trackers that collect personal data might gather information such as IP addresses, browsing history, or other identifiers that tracker providers can use to track user behavior across different publishers. Trackers that share personal data can transmit user information to other entities, such as advertising networks or data brokers, which use these data for, among others, targeted advertising or user profiling. By focusing on tracking functionality, we can better understand the level of privacy intrusion associated with different trackers and assess their compliance with the GDPR. This assessment allows us to evaluate whether the GDPR effectively reduced high-risk trackers, such as trackers that collect and share personal data. We thus answer whether the GDPR better protects user privacy, which is an intended consequence of the regulation.

2.4.3. Online Trackers by Type of Publisher

Different types of publishers have distinct business models and goals, influencing their use of trackers. For example, news publishers may rely heavily on advertising revenue and thus use more advertising trackers to monetize their content (Libert and Nielsen 2018). In contrast, non-news publishers like e-commerce websites or blogs might focus more on analytics trackers to improve user experience and increase sales.

The motivation for this categorization lies in the GDPR's uniform application across different types of publishers, but with the recognition that different publishers may have varying capabilities to adapt to the regulation. From a societal perspective, it is essential to ensure that vital services, like news, remain accessible to users. By examining publisher types, we can identify if the GDPR disproportionately affects specific sectors, potentially leading to unintended consequences such as reduced content available to users or increased financial pressure on certain publishers.

2.4.4. Online Trackers by Size of Tracker Provider

We categorize trackers by the size of their tracker providers. High market share providers are often associated with well-known firms familiar to users, such as Google and Facebook, offering multiple trackers and dominating the market. These providers typically own multiple trackers. In contrast, low market share providers are often less well-known firms. Examples include providers like [24]7 or Accord Group AdMicro.

The motivation for this categorization stems from the GDPR's aim to promote fair competition and prevent dominant players because the resulting market concentration could reduce user choices and innovations. By analyzing the impact on tracker providers of different sizes, we can assess whether the GDPR inadvertently benefits larger firms that can absorb compliance costs, leading to unintended consequences like increased market concentration.

3. Related Literature

Our work contributes to three separate streams of literature: (i) the literature on user privacy concerns, (ii) the literature describing online trackers, and (iii) the literature on the impact of privacy regulation on online trackers. In Table 2, we summarize the key findings of these three literature streams and outline how our study extends them.

3.1. Literature on User Privacy Concerns

Extensive research on user privacy concerns highlights how individuals weigh the benefits of data sharing (e.g., personalization) against perceived risks (e.g., data misuse) in a "privacy calculus" (Beke et al. 2022; Eggers et al. 2023). This trade-off often leads to a "privacy paradox," whereby users express serious concerns but still share data (Bleier et al. 2020). Studies also emphasize the role of contextual integrity, which stresses that privacy concerns spike when data use diverges from user expectations (Lobschat et al. 2021; Martin et al. 2017). In turn, transparency and user control are widely cited as mechanisms that mitigate user anxieties (Wieringa et al. 2021; Gopal et al. 2023). Furthermore, personalized advertising can exacerbate privacy worries if users feel uninformed about data usage, indicating that companies should balance personalization with corporate digital responsibility (Tucker 2012; Rocher et al. 2019). As summarized in Table 2, these insights show that firms can better address privacy concerns by combining clear communication with privacy-enhancing features and responsible data handling.

3.2. Literature Describing Online Trackers

Research on online trackers indicates widespread adoption by publishers, who use them to monetize content and obtain analytics (Mayer and Mitchell 2012). A small group of dominant providers (e.g., Google, Facebook) collectively account for a large share of the tracking ecosystem, raising both efficiency benefits (funding free content) and privacy concerns tied to concentrated data collection (Lerner et al. 2016; Karaj et al. 2018b). Table 2 details that these studies document an increasing market concentration over time and discuss the tension between economic gains and user privacy in online advertising.

3.3. Literature on the Impact of Privacy Regulation on Online Trackers

Studies of privacy regulation, especially the GDPR, reveal its potential to reduce the number of trackers and limit data collection (Peukert et al. 2022; Johnson et al. 2023). However, findings are mixed, with some research suggesting short-term declines in tracking and a possible rebound thereafter (Johnson et al. 2023). Others highlight that compliance costs may disproportionately favor large tracker providers, increasing market concentration (Peukert et al. 2022; Johnson et al. 2023). As outlined in Table 2, some also find limited negative impacts on user engagement or publisher revenue, suggesting publishers adapt through alternative approaches like contextual ads (Lefrere et al. 2024; Wang et al. 2024).

3.4. Contribution of Our Study to the Related Literature

In the context of these literature streams, our study makes three main contributions: First, we introduce a categorization of trackers that reflects the GDPR's aims, allowing us to analyze the regulation's impact with nuance. This categorization focuses on dimensions such as purpose and necessity, tracking functionality, type of publisher, and size of tracker provider, which are critical for understanding how the GDPR affects different trackers.

Second, we conceptually describe the importance of trackers in online advertising, detailing the roles of users, publishers, advertisers, and tracker providers, illustrating the conflicting aspects of trackers—providing value (e.g., monetization, personalization) for the various actors while raising privacy concerns for users.

Third, we empirically analyze the development of trackers before and after the GDPR, using balanced panel data from WhoTracks.me covering 294 publishers over 32 months and, in the

Web Appendix, an unbalanced panel data with 29,735 publishers. We also assess the GDPR's impact on the use of trackers through a difference-in-differences (DiD) analysis, comparing EU publishers subject to the GDPR with non-EU publishers as a control group. By exploring the heterogeneity of the GDPR's effects across different tracker categories, publisher types, and tracker providers, we derive both intended (e.g., reduced privacy-invasive tracking) and unintended consequences (e.g., potential shifts in market concentration or impact on certain types of publishers) of the regulation on trackers.

Making these contributions helps clarify whether the GDPR achieves its intended aim of mitigating privacy-invasive tracking and highlights how these regulatory changes impact online advertising.

<Add "Table 2: Summary of Key Findings of Related Literature and Our Contribution" about here>

4. Setup of Empirical Study

4.1. Description of the Data Sets

Table 3 provides an overview of the data sets used in our study, highlighting the type of information they contain, the periods they cover, and their purpose in reaching the aim of our study.

WhoTracks.me is our primary data set, offering detailed information on 294 publishers' use of trackers over 32 months from May 2017 to December 2019. (i.e., 12 months pre-GDPR and 20 months post-GDPR; Karaj et al. 2018b). We ended our observation period in December 2019 because the California Consumer Privacy Law (CCPA) was enacted in January 2020, and its effects might have interacted with those of the GDPR, confounding our observations from that month onward. Additionally, we chose to use a balanced panel of publishers for our main analysis to avoid panel attrition. In Web Appendix 9.5.2, we show the robustness of our results using an unbalanced panel with 29,735 publishers.

The WhoTracks.me data allows us to empirically describe trackers and assess the impact of the GDPR on them. We use various information from WhoTracks.me, including the number of trackers and categorizations by purpose, necessity, and size of tracker provider. Additionally, the data includes information that enables us to designate publishers as EU vs. non-EU based on their top-level domain. We detail the raw data from WhoTracks.me in Web Appendix 9.1.1 and how WhoTracks.me collects its data in Web Appendix 9.1.2.

We augment the WhoTracks.me data with data from SimilarWeb. This data provides traffic shares from the top five countries (EU and non-EU) for each of 294 publishers as of August 2021, which we use to refine our EU vs. non-EU publisher designations further.

Lastly, we augment the WhoTracks.me data with data from Evidon, which provides additional information on trackers from their privacy policies. This data enables us to categorize trackers by their tracking functionality, particularly regarding personally identifiable information (PII) data collection and sharing practices. Evidon data includes information on 724 trackers (76%) of the 949 trackers from the WhoTracks.me data, offering detailed insights into their PII data collection and sharing practices.

<Add "Table 3: Description of the Data Sets" About Here>

4.2. Description of the Sample Construction Process

This section outlines the process of constructing our sample of 294 publishers for the main analysis. To provide context, WhoTracks.me offers two types of data sets: the "global" sample and the "EU/US" sample.

The global sample contains data from May 2017 until December 2019, covering 32 months. It includes detailed tracking information for websites (referred to as publishers), but it does not provide any data about the location of the users visiting these publishers. In the raw global data, WhoTracks.me covers, on average, 8,334 publishers per month across this period, with the number of publishers increasing from 3,645 in May 2017 to 15,004 by May 2018. The panel of

publishers in the raw data is unbalanced, as the number of publishers released each month varies, with a maximum of 17,987 publishers in September 2018. After filtering the data to focus only on consistently tracked publishers over the entire 32-month period, we arrived at a balanced panel of 962 publishers (-88% change from the average of 8,334 publishers per month).

In contrast, the EU/US sample contains user location information, distinguishing between EU and US users. Still, it covers a shorter period, from April 2018 to December 2019 (21 months), and includes only a single pre-treatment period (April 2018). The unbalanced version of this sample includes an average of 7,264 publishers per month, with a maximum of 11,596 publishers in September 2018 and a minimum of 2,869 publishers in January 2019. We created a balanced panel of 717 publishers (-90% change from the average of 7,264 publishers per month) by selecting those consistently tracked throughout the 21 months in the EU/US sample. While this EU/US sample is not the focus of our main analysis, we use it in a robustness test, as outlined in Web Appendix 9.2.2.

For the main analysis, we focused on the global sample because it offers a longer pre-treatment period, which is essential for robustly estimating the impact of the GDPR. We ensured consistency in our analysis by selecting the publishers that appeared in both the global and EU/US samples. In both data sets, this step yielded 354 publishers (a reduction of 63% from the balanced global sample of 962 publishers).

Finally, we examined the assumption of parallel trends in the global sample. Specifically, we observed the pre-treatment trend of each publisher in the control group (in terms of the number of trackers). We removed publishers from the control group with pre-treatment trends that deviated significantly from the treatment group. This process ensured that the treatment and control groups were well-matched and that the assumption of parallel trends remained intact. After removing these outliers, we arrived at a final balanced sample of 294 publishers (a 17% reduction from the intersection of global and EU/US samples), comprising 67 publishers in the treatment group and 227 in the control group. Table 4 summarizes the steps to refine the sample for our analysis.

< Add "Table 4: Steps Taken to Prepare the Sample of 294 Publishers" About Here>

As shown in Table 4, we started with 962 publishers from the balanced global sample and 717 from the balanced EU/US sample. We filtered these publishers down to 354 (-63% change from the balanced global sample) by focusing on the ones in both samples. Finally, after removing outliers to ensure the parallel trends assumption held, we arrived at 294 publishers (-17% change from the intersection of global and EU/US samples). Because of the longer pre-treatment period, this sample forms the basis for our main analysis. We use the EU/US sample as a robustness test, described in Web Appendix 9.2.2.

4.3. Description of the Number of Trackers

For each publisher in our WhoTracks.me data, we measure the number of trackers that the publisher installed in a particular month and use it as our dependent variable in the subsequent analysis. Tracing the number of trackers allows us to document the publisher's privacy

practices—particularly the large-scale information collection outlined by Beke et al. (2018)—and ultimately serves as a measure of a user's exposure to privacy risk.

To obtain a reliable measurement of the number of trackers initiated by our focal publishers, we adjusted the raw number of trackers provided by WhoTracks.me, as follows:

First, WhoTracks.me counts specific browser extensions (e.g., Kaspersky Labs, Adguard) that users voluntarily install as "trackers". WhoTracks.me identifies these trackers as "extensions". As users install these trackers rather than the publishers themselves, we excluded them from each publisher's overall tracker count.

Second, we counted only third-party trackers because the European Data Protection Authorities (DPAs) consider them a greater privacy risk than first-party trackers (Article 29 Data Protection Working Party 2012). We defined these as trackers whose tracker providers differed from the publisher. Our definition of third-party trackers closely follows that of the European DPAs (Article 29 Data Protection Working Party 2012).

We acknowledge that the set of trackers on a publisher's website that WhoTracks.me reports may not be fully comprehensive. Specifically, WhoTracks.me does not capture certain types of trackers, so we cannot include them in our dataset. As Ghostery (2017) noted, these trackers are typically found on fewer than ten publishers or do not rely on cookies or fingerprinting technologies to track user identifiers. Consequently, we do not expect these trackers' omissions to substantially impact our results' reliability.

To explore the effects of the GDPR on trackers, we also count, for each publisher and each month, the numbers of trackers corresponding to the tracker categorizations as elaborated in Section 2.4.

4.4. Construction of the Treatment and Control Groups

We broadly define our treatment group as publishers subject to the GDPR; as discussed above, these publishers corresponded to EU firms. We broadly define our control group as publishers not subject to the GDPR—i.e., non-EU firms.

Determining whether a publisher represents an "EU firm" or a "non-EU firm" is challenging. The GDPR defines an "EU firm" as any firm established within the EU, including firms based in the EU or that process EU citizens' personal data—regardless of where the firm is based. In contrast, the GDPR defines a "non-EU firm" as any firm that is not established within the EU and does not process the personal data of EU citizens (European Commission 2016; European Data Protection Board 2018).

We suggest several proxies to identify whether a publisher is an "EU" or a "non-EU firm": the publisher's (1) target audience (i.e., the main set of users it caters to), (2) (country-code) top-level domain (TLD), (3) cookie banner display, (4) server location, and others. Our main analysis combines two proxies: the publisher's target audience and its TLD.

Specifically, we identify a publisher as an "EU publisher" if it fulfills one of the following criteria: (1) the publisher's TLD includes an EU country code (e.g., .de); *or* (2) the publisher receives more traffic from EU users than from non-EU users in at least one month (e.g., August 2021), according to our SimilarWeb data. Correspondingly, a publisher is designated as a "non-EU publisher" if it fulfills *both* of the following criteria: (1) the publisher uses a non-EU TLD (e.g., .com), and (2) the publisher receives more traffic from non-EU users than from EU users.

Our sample corresponds to 294 EU and non-EU publishers over 32 months (N = 9,408 observations). Table 5 shows how we assign the different publishers and their corresponding monthly observations to the treatment and control groups. In total, 22.79% (N observations = 2,144) of all observations belong to the treatment group, and 77.21% (N observations = 7,264) belong to the control group.

< Add "Table 5: Distribution of Observations (Monthly Publishers) Across Publisher Designation" About Here>

4.5. Identification Assumptions for the Difference-in-Differences Analysis

We use a DiD analysis to estimate the GDPR's effect on trackers. DiD is suitable because it allows for comparing changes in tracker usage over time between EU publishers (affected by GDPR) and non-EU publishers (unaffected), effectively isolating the GDPR's causal impact. DiD analysis relies on the following two assumptions, the violation of which could bias our results: (1) stable unit treatment value assumption (SUTVA) and (2) parallel trends assumption (Huntington-Klein 2022). In what follows, and following Goldfarb et al. (2022), we discuss these two assumptions in the context of our study.

SUTVA comprises two parts. The first part states that there is no hidden variation of treatment, i.e., that all treated publishers receive the same level of treatment. This assumption is fulfilled, as the treatment we observe is the GDPR's enactment—which imposes consistent requirements across all firms subject to the GDPR.

The second part of SUTVA states that no spillover effects exist between the treatment and control groups. We stringently define our control group to avoid such spillover (see description above and Table 5). We acknowledge, however, that, despite not being required to do so, some publishers in our control group might have altered their tracking practices following the GDPR to avoid inadvertently incurring the GDPR fines, thus "contaminating" our control group. This concern is present in most, if not all, studies on the impact of the GDPR (e.g., Johnson 2023), and we address it in Section 5.3.

The parallel trends assumption states that in the absence of the GDPR, the outcome difference between the treatment and control group would have remained the same after the GDPR as before the GDPR. DiD studies commonly evaluate this assumption by providing visual evidence of group outcomes in a period before the intervention.

5. Results of Empirical Study

In what follows, we empirically describe the distribution of trackers and determine the impact of the GDPR on them.

5.1. Distribution of Number of Online Trackers

5.1.1. Distribution of Number of Online Trackers Across Publishers

Figure 2 shows the average number of trackers per publisher, averaged across all months in our observation period. It shows that there are, on average, 16.689 trackers per publisher (SD

= 13.496). Their distribution is unimodal and right-skewed. Thus, most publishers typically use between 1 and 10 trackers; only a few use many. The largest number of trackers for a publisher in a particular month is 111 trackers. Because the data only includes publishers with at least 1 tracker, the minimum number of trackers is 1.

5.1.2. Distribution of Number of Trackers By Categorizations of Online Trackers

Table 6 provides the distribution of the average number of trackers per publisher across five dimensions of tracker categorizations, providing a baseline understanding of publishers' use of trackers before examining how GDPR impacts them.

By necessity, publishers tend to use three times more non-essential trackers than essential ones, with an average of about 12 non-essential trackers compared to around four essential trackers per publisher.

<Add "Figure 2: Distribution of the Average Number of Trackers per Publisher" About Here>

By purpose, the distribution indicates that publishers use specific types of essential and non-essential trackers in varying amounts. Publishers commonly use essential trackers such as tag managers and content delivery network (CDN) trackers, with an average of 0.774 tag manager trackers and 2.931 CDN trackers per publisher. For non-essential trackers, publishers mostly use trackers for advertising and analytics purposes, with averages of 7.257 and 2.864 trackers per publisher, respectively. Notably, publishers rarely use privacy-friendly analytics trackers, averaging 0.032 per publisher, suggesting publishers' limited use of these trackers across our entire observation period. Similarly, publishers' use of consent trackers is minimal (0.032 trackers) across the observation period.

<Add "Table 6: Distribution of the Average Number of Trackers per Publisher By Categorizations of Trackers" About Here>

By tracking functionality, the distribution reveals that publishers commonly use trackers that collect and share PII, with an average of 8.424 trackers per publisher. Publishers use trackers that collect but do not share PII less frequently, averaging 3.258 per publisher. In contrast, they use trackers that neither collect nor share PII, least commonly, with an average of 1.872 trackers per publisher.

By type of publisher, the distribution shows that news publishers use almost two times more trackers than non-news publishers, averaging about 29 trackers compared to 15 trackers for non-news publishers. Notably, within the non-news publisher industry, e-commerce publishers stand out with an average of 24.971 trackers per publisher, while government publishers tend to use the fewest, averaging just 7.281 trackers. The distribution reveals that even entertainment publishers who fall under the non-news publisher industry can employ many trackers, with one entertainment publisher using up to 103 trackers in a particular month.

By size of tracker provider, publishers tend to use a similar number of trackers from both high and low market share tracker providers, with averages of 8.351 and 8.338 trackers, respectively, from each type.

5.1.3. Distribution of Number of Trackers By Categorizations of Online Trackers in the Treatment and Control Groups

Table 7 presents the distribution of the average number of trackers per publisher across the treatment and control groups, categorized by tracker purpose, necessity, functionality, publisher type, and size of the tracker provider.

The treatment group consistently uses more trackers than the control group across most tracker categorizations, with a 31.32% higher overall average (20.457 vs. 15.577 trackers). By purpose and necessity, the treatment group uses more non-essential trackers (35.65% higher), particularly in advertising (58.97% higher). By tracking functionality, the treatment group uses more trackers that collect and share personally identifiable information (19.24% higher). News publishers in the treatment group also use substantially more trackers (54.49% higher) than those in the control group.

In addition to the differences in tracker use between the treatment and control groups, Table 7 also summarizes publisher characteristics.

<Add "Table 7: Distribution of the Average Number of Trackers per Publisher By Categorizations of Trackers in the Treatment and Control Groups" About Here>

The treatment group has a higher share of traffic from EU users (48.37%) compared to the control group (10.24%), while the control group has a larger share of traffic from non-EU users (43.28% vs. 15.63%). The five most TLDs also differ between the groups, with the treatment group featuring more EU-specific TLDs like ".co.uk" and ".de". The control group, in contrast, commonly uses more global or non-region-specific TLDs like ".com" and ".ru".

5.2. Impact of the GDPR on Number of Online Trackers

5.2.1. Change in Number of Online Trackers Before and After the GDPR

To investigate the effect of the GDPR on trackers, we first show how the average number of trackers differs between the treatment and control groups before (May 2017-April 2018) and after the GDPR's enactment (May 2018-December 2019). We run an independent-samples t-test to test whether the group averages differ significantly in the two periods and present the results in Figure 3.

<Add "Figure 3: Comparison of the Average Number of Trackers in the Treatment and Control Groups Before and After the GDPR's Enactment" About Here>

The average number of trackers per publisher was significantly higher in the treatment group than in the control group before the GDPR ($M_{treatment}$ = 16.609 trackers; $M_{control}$ = 9.262 trackers; t(992.7283) = 14.018; p < 0.001). After the GDPR, the number of trackers increased in both groups ($M_{treatment}$ = 22.765; $M_{control}$ = 19.367), with the treatment group still having a significantly higher average number of trackers compared to the control group (t(1836.5045) = 6.920; p < 0.001).

5.2.2. Change in the Number of Online Trackers

Figure 4 shows how each group's average number of trackers changed over time.

<Add Figure 4: Change in the Number of Online Trackers in the Treatment and Control Group" About Here>

Figure 4 suggests an increasing trend in the number of trackers in the treatment and control groups. The number of trackers in the treatment group was consistently higher than that of the control group throughout the observation period. Around the time of the GDPR's enactment (May 2018), the number of trackers in the treatment group slightly decreased, while the control group increased slightly. However, by the end of 2018, the number of trackers in both groups rises, with the treatment group maintaining a higher level than the control group.

Starting from December 2018, both groups continue to increase the number of trackers, with the treatment group peaking in the last observed month (December 2019) at about 25 trackers, while the control group also increases but remains lower than the treatment group at about 22 trackers.

5.2.3. Difference-in-Differences (DiD) Analysis for the Number of Online Trackers

Having obtained a preliminary indication that the GDPR reduced the number of trackers among publishers subject to the regulation (compared with the control group), we examine this effect with a DiD analysis that accounts for unobserved influences.

First, we manually calculate the DiD (average treatment effect on the treated, ATT)–the difference between the average differences in the number of trackers of both groups (Table 8).

<Insert Table 8: Average (Monthly) Number of Trackers in the Treatment and Control Groups Before and After the GDPR's Enactment" About Here>

After the GDPR, the average number of trackers increased by 6.155 (37.06%) in the treatment group and 10.104 (109.09%) in the control group. The control group increased the average number of trackers more than the treatment group after the GDPR. The difference between those two numbers captures the effect of the GDPR. It equals -3.949 trackers, suggesting that the GDPR lowered the average number of trackers by about four trackers per publisher.

In other words, these results suggest that if the GDPR had not been enacted, the average publisher in the treatment group would have used about 27 trackers in the post-GDPR period rather than about 23 trackers. This difference corresponds to a 14.79% decrease.

After calculating the DiD, we use an ordinary least squares (OLS) regression to control for other factors that might influence the DiD estimate (e.g., differences between publishers):

$$Y_{i,t} = \alpha + \gamma_i + \delta_t + \beta(Treatment_i \times PostGDPR_t) + \epsilon_{i,t} \qquad (1)$$

In Equation (1), our dependent variable for a publisher i in month t is $Y_{i,t}$. γ_i captures the publisher-fixed effects (group-specific changes in the outcome variable unrelated to GDPR), and δ_t captures the month-fixed effects (time-specific changes in the outcome variable unrelated to GDPR).

$Treatment_i$ is an indicator variable describing whether the publisher i is in the treatment group (i.e., subject to the GDPR; $Treatment_i = 1$) or not ($Treatment_i = 0$). $PostGDPR_t$ indicates the period before the GDPR ($PostGDPR_t = 0$) and the entire period after the GDPR (i.e., after and including May 2018; $PostGDPR_t = 1$). Because $Treatment_i$ does not vary within a publisher and $PostGDPR_t$ does not vary across publishers, these main effects are absorbed by γ_i and δ_t, leaving the β coefficient on $Treatment_i \times PostGDPR_t$ as our DiD estimate.

$\epsilon_{i,t}$ represents the error term, which includes unobserved factors affecting $Y_{i,t}$. We cluster standard errors at the publisher and month levels to account for autocorrelation within publishers over time and across months. Our coefficient of interest β measures the average difference in the number of trackers between both groups over time (i.e., represents the DiD estimate). So, our baseline estimator is a standard two-way fixed effects (TWFE) estimator with the publisher as the observation unit. Table 9 presents the results.

The DiD coefficient (β = -3.949, $p < 0.05$, 95% CI [-7.082; -0.816]) is significantly negative in our DiD model presented in column (1). The size of this DiD coefficient is the same as the calculated difference-in-differences in Table 8. These results confirm that the GDPR lowered the number of trackers by about four per publisher (14.79%). We report the estimated publisher and month-fixed effects in Web Appendix 9.4.2.

< Add "Table 9: Result of Difference-in-Differences (DiD) Analysis for the Number of Trackers" About Here>

5.2.4. Differences in the Impact of the GDPR By Categorizations of Online Trackers

Lastly, we show how the GDPR impacted various tracker categorizations in our sample. We estimate the following model to measure the effect of the GDPR across different tracker categorizations:

$$Y_{i,t}^k = \alpha + \gamma_i + \delta_t + \beta^k(Treatment_i \times PostGDPR_t) + \epsilon_{i,t}^k \qquad (2)$$

In Equation (2), $Y_{i,t}^k$ represents the number of trackers in category k for publisher i in month t. We include publisher-fixed effects (γ_i) and month-fixed effects (δ_t), with $Treatment_i \times PostGDPR_t$ as our key DiD interaction. The coefficient of interest β^k measures the average difference in the number of trackers in category k between the treatment and control groups

over time, representing the DiD coefficient for tracker category k. We apply this model separately to three tracker categorizations: by purpose (e.g., advertising, analytics) and necessity (i.e., essential vs. non-essential trackers), by tracking functionality (e.g., trackers that collect and share PII), and by size of tracker provider (i.e., tracker providers with high vs. low market share).

To examine how the GDPR affected publisher industries, we use the following model:

$$Y_{i,t}^{p} = \alpha + \gamma_i + \delta_t + \beta^{p}(Treatment_i \times PostGDPR_t) + \epsilon_{i,t}^{p} \qquad (3)$$

In Equation (3), $Y_{i,t}^{p}$ represents the number of trackers for publisher i in publisher industry p in month t. We estimate this model separately for each publisher industry p (i.e., News vs. Non-News). The coefficient of interest β^{p} captures the impact of the GDPR on the number of trackers within each publisher industry.

Lastly, to examine how the GDPR affected specific types of publishers within these industries, we use:

$$Y_{i,t}^{c} = \alpha + \gamma_i + \delta_t + \beta^{c}(Treatment_i \times PostGDPR_t) + \epsilon_{i,t}^{c} \qquad (4)$$

In Equation (4), $Y_{i,t}^{c}$ represents the number of trackers for publisher i in publisher type c during month t. We estimate this model separately for each specific publisher type c (e.g., News & Portals, E-Commerce, Recreation). The coefficient β^{c}captures the impact of the GDPR on the number of trackers within each publisher type (see Web Appendix 9.1.3. for more details on publisher industries and types).

All three models include γ_i publisher-fixed effects and δ_t month-fixed effects, with standard errors clustered at the publisher and month levels. Figure 5 reveals the distribution of the effect of the GDPR on the number of trackers across these tracker categorizations.

By necessity, essential trackers experienced a significant decrease, with the DiD coefficient ($\beta = -1.192$, $p < 0.01$, 95% CI [−1.961;−0.422]) suggesting that the GDPR reduced the number of essential trackers by about one per publisher. Non-essential trackers also saw a significant reduction, with a DiD coefficient of $\beta = -2.757$ ($p < 0.05$, 95% CI [−5.097;−0.417]), indicating a decrease of about three per publisher.

By purpose, the results show that hosting trackers decreased significantly after the GDPR, as reflected by the DiD coefficient ($\beta = -0.305$, $p < 0.01$, 95% CI [−0.485;−0.125]). CDN trackers also experienced a significant reduction, with a DiD coefficient of $\beta = -0.792$ ($p < 0.01$, 95% CI [−1.268;−0.316]). Likewise, analytics trackers saw a substantial decrease, with a DiD coefficient of $\beta = -0.873$ ($p < 0.001$, 95% CI [−1.271;−0.476]), indicating a reduction of about 0.9 per publisher. Social media trackers were also significantly affected, with a DiD coefficient of $\beta = -0.217$ ($p < 0.05$, 95% CI [−0.382;−0.052]).

By tracking functionality, the results show that trackers that collect PII experienced a significant decrease, as indicated by the DiD coefficient ($\beta = -0.874$, $p < 0.05$, 95% CI [−1.664;−0.084]). Moreover, trackers that collect and share PII also saw a significant reduction after the GDPR, with a DiD coefficient of $\beta = -2.201$ ($p < 0.01$, 95% CI [−3.633;−0.770]).

By type of publisher, non-news publishers saw a significant decrease in the number of trackers after the GDPR, as indicated by a DiD coefficient of $\beta = -6.328$ ($p < 0.001$, 95%

CI [−8.963;−3.694]). Within this broader publisher industry, the recreation types of publishers experienced a particularly significant reduction, with a DiD coefficient of $\beta = -13.551$ ($p < 0.05$, 95% CI [−20.843;−6.260]).

Lastly, by the size of the tracker provider, the number of trackers from tracker providers with a high market share significantly decreased after the GDPR, with a DiD coefficient of $\beta = -2.360$ ($p < 0.01$, 95% CI [−3.783;−0.937]).

<Insert "Figure 5: Distribution of the GDPR's Impact Across Categorizations of Trackers" About Here>

5.3. Insights from the Robustness Tests

We conducted a series of robustness tests to ensure the reliability of our main analysis, addressing potential concerns about our methodology and our data limitations. Table 10 summarizes these robustness tests, with further details in the Web Appendices.

First, we performed alternative treatment assignments based on server location and user location to account for possible misclassifications of publishers into treatment and control groups. The results show that the GDPR reduced the number of trackers by 3.867 per publisher when using server location as the treatment assignment criterion and by 1.692 per publisher instance when using publisher designation and user location as the treatment assignment criteria. We estimated the monthly DiD coefficients and conducted placebo tests, confirming that the assumption of parallel trends likely holds.

Regarding spillover effects, we performed the "cleanest" comparison between EU-located users visiting EU publishers and US-located users visiting non-EU publishers. This robustness test demonstrated that the GDPR reduced the number of trackers by 2.922 per publisher instance, with minimal spillovers affecting the control group. Additionally, we tested whether the GDPR affected user behavior rather than publishers' use of trackers by examining the number of Ghostery users before and after the GDPR. The results indicated no significant change in Ghostery users, confirming that the reduction in trackers was due to publishers' user of trackers.

To address concerns about early compliance with the GDPR or shocks unrelated to the GDPR, such as the Cambridge Analytica scandal (Cadwalladr and Graham-Harrison 2018), we removed the months of March, April, May, and June 2018 from our main analysis. Even after doing so, the GDPR reduced the number of trackers by 4.523 per publisher, suggesting that anticipation or external shocks do not bias our results.

We also addressed the skewness in the distribution of the number of trackers by applying a log transformation to our dependent variable. The results showed that the GDPR reduced the logged number of trackers by 0.490 per publisher, confirming the robustness of our findings.

Another concern we addressed was the potential instability of publishers' website traffic shares over time, given that we relied on a single point-in-time SimilarWeb data set to designate publishers into treatment and control groups. By comparing SimilarWeb data (collected in September 2021) with SimilarWeb data (collected from January 2018 to December 2019), we

found that the average difference in EU traffic shares was 11.08 percentage points, indicating stable website traffic distributions over time.

<Insert "Table 10: Summary of Robustness Tests" About Here>

We applied the generalized synthetic control (GSC) method to account for potential model misspecifications in the DiD analysis. This method constructs a counterfactual (i.e., control group) by matching treated and control units more accurately, relying on pre-treatment trends. Unlike the DiD approach, GSC allows the algorithm to select the optimal control group. The GSC method's analysis showed that the GDPR reduced the number of trackers by 5.303 per publisher, further confirming the robustness of our findings.

Finally, we performed a robustness test using an unbalanced panel to ensure our results are representative for a larger number of publishers. While we use a balanced panel of 294 publishers in the main analysis to, among others, avoid panel attrition, the unbalanced panel includes 29,735 publishers, representing a 9,994% increase. It allows us to maximize the number of observations. We assigned treatment to publishers based on TLD and server location, and the results show that the GDPR reduced the number of trackers by 1.081 per publisher (treatment assignment based on TLD) and by 0.825 per publisher (treatment assignment based on server location). This robustness test also confirms the robustness of our findings.

6. Summary, Conclusions, and Implications

6.1. Summary of Results and Conclusions

Trackers are software that often combines a specific purpose with tracking functionality. Publishers embed them into their websites to monitor user behavior, personalize content, or deliver targeted ads. However, trackers often collect and share user data across multiple publishers and advertisers. While they generate value for publishers by enhancing content and attracting users—monetized through targeted advertising—they raise substantial privacy concerns by processing users' personal data. Consequently, regulators have enacted laws like the EU's GDPR to enhance online privacy.

In this paper, we examined the impact of the GDPR on online tracking. By categorizing trackers in a manner aligned with the GDPR's objectives, we assessed both the intended and unintended consequences of the regulation. Our main findings and conclusions are as follows (see Table 11 and Table 12 for a summary):

First, publishers rely heavily on trackers for various purposes, with advertising, analytics, and content delivery being the most commonly used. Most trackers are highly privacy-invasive, as they collect and share personal data. Although the number of trackers increased for EU and non-EU publishers from before to after the GDPR, the increase was significantly smaller for EU publishers. Thus, the GDPR achieved its intended consequence by decreasing the number of trackers by 14.79% compared to expectations without the regulation.

Despite the GDPR's enactment, many trackers remain on publishers' websites, and thus, the infrastructure for collecting, retaining, and sharing data remains largely intact. Notably, the GDPR did not significantly reduce the number of advertising trackers and only marginally reduced the number of analytics trackers. This outcome favors publishers seeking to enhance

their content, attract new users, and monetize them through ads. It also benefits advertisers aiming to reach their target audiences effectively and tracker providers monetizing collected data by offering enhanced tracking services or selling it to third parties.

Second, privacy concerns persist despite the continued presence of many trackers on publishers' websites—which provide value to online advertising firms and potentially to users. However, the introduction of the GDPR has allowed users to consent to or decline the processing of personal data, providing them with greater control over their personal information. It remains unclear whether all users will choose to consent to being tracked. Nonetheless, users now have the option to deny consent, which was not readily available before the GDPR. This option represents a significant privacy benefit for users.

<Add "Table 11: Summary of Empirical Findings on Description of Online Trackers and Their Conclusions" About Here>

<Add "Table 12: Summary of our Empirical Findings of Impact of GDPR on the Number of Online Trackers and Their Conclusions" About Here>

Third, advertisers face publishers that collect varying amounts of user data. While the GDPR did not achieve its intended consequence of decreasing advertising trackers, it reduced the use of highly privacy-invasive trackers. This reduction potentially leaves advertisers with less data from some publishers. There is substantial heterogeneity in tracker usage across publishers; most have between 1 and 10 trackers, but the distribution is heavily right-skewed, with some publishers using up to 111 trackers.

6.2. Implications

From the summary of our main findings and conclusions, we derive the following implications:

First, advertisers face considerable heterogeneity among publishers regarding the amount of user data collected. This heterogeneity implies that behavioral targeting may no longer function uniformly across all publishers. Advertisers can respond by reallocating their ad budgets toward publishers with more extensive user data and away from those with less data. Alternatively, they may invest in more privacy-preserving forms of advertising on publishers with limited data, such as contextual targeting or privacy-enhancing technologies like Google's Privacy Sandbox (Johnson 2024; Jerath and Miller 2024), to effectively reach their target audiences.

Second, it is unclear whether publishers that use less tracking will gain a competitive advantage through increased user engagement or will face disadvantages due to advertisers' reduced ability to employ behavioral targeting on their platforms. Advertisers might have to rely on alternative forms of advertising (e.g., contextual targeting) with these publishers, which could be less profitable than behavioral targeting. This uncertainty raises questions about the balance between enhancing user privacy and maintaining revenue streams for publishers.

Third, our study demonstrated that the GDPR achieved some of its intended consequences by reducing overall tracking by 14.79% and decreasing the use of highly privacy-invasive trackers that collect and share personal data. Nevertheless, even after the GDPR, publishers continue to

use trackers heavily and are increasingly doing so. If regulators aim to reduce online tracking further, further activities are necessary, which might include even stricter enforcement of the existing rules.

Fourth, the GDPR may have improved privacy for users, especially for those who exercised their right to decline consent for tracking. However, publishers have not significantly increased their use of privacy-friendly analytics trackers, which suggests that users do not have additional opportunities to opt-in to less privacy-invasive tracking. Our results also indicate reduced content delivery trackers, which may have diminished user experience on publishers' websites. Users might receive less relevant content recommendations and advertisements because of tracking less personal data. The unintended consequence of the GDPR in reducing essential trackers that do not collect or share personal data but deliver content (e.g., videos) might negatively impact users.

A potential way to preserve the advantages of trackers while addressing privacy concerns is to decouple the tracking purpose from the tracking functionality. This decoupling would let users access enriched content without being subject to tracking. For example, users could then access YouTube videos on a publisher's website without being tracked. However, this approach would require someone—the publishers, advertisers, tracker providers or even the users—to cover the cost of providing the tracker technology without its data-gathering capabilities. Essentially, such unbundling places the financial burden on one party to maintain the content-enhancing features while ensuring user privacy. Achieving a balance between privacy and tracking often involves trade-offs; in this case, between more privacy and covering the cost for the functionality that was bundled into the tracker.

To summarize, GDPR reduced the number of trackers in the EU compared to other areas, such as the US. However, the number of trackers in the EU and the US increased over time. Evaluating whether these varying effects represent good news for EU users' privacy is challenging. Similarly, the continued use of only a slightly reduced number of trackers may indicate that the advertising industry's ability to target consumers after the GDPR has remained largely unaffected.

7. Limitations and Future Research

We acknowledge several limitations of our study and suggest avenues for future research.

First, while the GDPR likely contributed to the reduction in trackers observed after its implementation, understanding the exact mechanisms behind this reduction is challenging. Other factors may have played a role, such as (i) innovations by tracker providers that allowed publishers to comply with the GDPR while continuing to use trackers; (ii) slow and inconsistent enforcement of the GDPR, which may have encouraged publishers to maintain or re-establish their use of trackers once they perceived lower risk of penalties; (iii) introduction of consent mechanisms over time (e.g., consent trackers), helping publishers manage compliance without significantly affecting their business goals (Johnson et al. 2023; Lefrere et al. 2024). Future research could explore these mechanisms by engaging with publishers and tracker providers to understand how they adapted to the GDPR over time.

Second, we stringently define our control group to avoid spillover effects between the treatment and control groups. However, despite not being required to do so, some publishers in our control group might have altered their tracking practices following the GDPR to avoid inadvertently incurring the GDPR fines, thus "contaminating" our control group. In such a case,

our measured effect of the GDPR would underestimate, but not overestimate, the "true" effect of the GDPR.

Third, we assume that changes in the number of trackers directly correlate with user privacy concerns. However, perceived privacy violations might depend on the trade-off between benefits (e.g., personalized services) and the costs of being tracked (see Jerath and Miller 2024; Lin 2022). We do not test how these economic trade-offs affect perceived privacy violations, and we leave it to future research to explore this relationship.

Fourth, the role of user consent in determining the number of trackers is complex. Studies like Demir et al. (2024) confirm that when users deny consent via GDPR-compliant cookie banners, publishers reduce the number of trackers deployed, contributing to lower tracking practices. However, other research indicates that publishers do not always honor user consent. Sanchez-Rola et al. (2019) and Bouhoula et al. (2024) show that some publishers continue to track users even after denying consent through cookie banners. Non-compliance can be intentional, as publishers may ignore users' choices to maintain monetization benefits despite risking GDPR fines, or unintentional if publishers are unaware that specific trackers remain active (Ghostery 2017; Müller-Tribbensee 2024). This ambiguity complicates the interpretation of tracker reductions as solely driven by user consent. Future research should investigate the extent of publisher compliance with consent mechanisms and how this affects tracking practices.

Fifth, we adopt a legal perspective in distinguishing personal from non-personal data, guided by the GDPR's provisions—our primary research context. We acknowledge, however, that other perspectives question whether a clear boundary exists between personally identifiable and non-personally identifiable information (see Dinur and Nissim 2003; Ponte et al. 2024).

Sixth, the WhoTracks.me dataset provides comprehensive metrics on publishers' tracking practices, but it may not fully represent the general internet user population. The data comes from users who have installed privacy tools like Ghostery, who are typically more privacy-conscious and technologically savvy than average users. This self-selection could limit the generalizability of our findings. Privacy-conscious users may engage in different browsing behaviors and are more likely to have additional privacy-focused extensions installed, which might block trackers before Ghostery (Yan et al. 2022) can detect them, leading to an underestimation of the number of trackers recorded. We mitigate this concern by using the unique number of trackers users encounter per publisher as our dependent variable. Since WhoTracks.me aggregates data from millions of users, as long as at least one user does not have additional privacy tools installed, the complete set of unique trackers a publisher uses can be captured.

Nevertheless, the potential lack of representativeness remains a limitation. Future research could incorporate more representative datasets that include a broader cross-section of users and compare results across different user groups to assess how varying levels of privacy awareness or cultural factors affect exposure to trackers. Such research could enhance the generalizability of findings related to publishers' tracking practices and the impact of privacy regulations like the GDPR.

8. References

Ahmadi, Iman, Nadia Abou Nabout, Bernd Skiera, Elham Maleki, and Johannes Fladenhofer (2024), "Overwhelming Targeting Options: Selecting Audience Segments for Online Advertising," International Journal of Research in Marketing, 41 (1), 24–40.

Article 29 Data Protection Working Party (2012), "Opinion 04/2012 on Cookie Consent Exemption," European Commission.

Beke, Frank T., Felix Eggers, and Peter C. Verhoef (2018), "Consumer Informational Privacy: Current Knowledge and Research Directions," Foundations and Trends in Marketing, 11 (1), 1–71.

Beke, Frank T., Felix Eggers, Peter C. Verhoef, and Jaap E. Wieringa (2022), "Consumers' Privacy Calculus: The PRICAL Index Development and Validation," International Journal of Research in Marketing, 39 (1), 20–41.

Bleier, Alexander, Avi Goldfarb, and Catherine Tucker (2020), "Consumer Privacy and the Future of Data-Based Innovation and Marketing," International Journal of Research in Marketing, 37 (3), 466–80.

Bouhoula, Ahmed, Karel Kubicek, Amit Zac, Carlos Cotrini, and David Basin (2024), "Automated {Large-Scale} Analysis of Cookie Notice Compliance," in 33rd USENIX Security Symposium (USENIX Security 24), 1723–39.

Cadwalladr, Carole and Emma Graham-Harrison (2018), "Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach," The Guardian, [https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election].

Commission Nationale de l'Informatique et des Libertés (CNIL) (2023), "Cookies: Solutions Pour Les Outils De Mesure D'audience," Site Web, Cookies, Et Autres Traceurs, , [https://www.cnil.fr/fr/cookies-et-autres-traceurs/regles/cookies-solutions-pour-les-outils-de-mesure-daudience].

Demir, Nurullah, Tobias Urban, Norbert Pohlmann, and Christian Wressnegger (2024), "A Large-Scale Study of Cookie Banner Interaction Tools and their Impact on Users' Privacy," Proceedings on Privacy Enhancing Technologies, 2024 (1), 5–20.

Dinur, Irit, and Kobbi Nissim (2003), "Reveaing Information while Preserving Privacy, " PODS '03: Proceedings of the Twenty-Second ACM SIGMOD-SIGACT-SIGART Symposium on Principles in Database Systems, 202-210.

Eggers, Felix, Frank T. Beke, Peter C. Verhoef, and Jaap E. Wieringa (2023), "The Market for Privacy: Understanding How Consumers Trade Off Privacy Practices," Journal of Interactive Marketing, 58 (4), 341–60.

European Commission (2016), "Regulation (EU) 2016/679 of the European Parliament and of the Council," Official Journal of the European Union, [https://eur-lex.europa.eu/eli/reg/2016/679/2016-05-04].

European Data Protection Board (2018), "Guidelines 3/2018 on the Territorial Scope of the GDPR (Article 3)," 1–23.

European Data Protection Board (2019), "Guidelines 2/2019 on the Processing of Personal Data Under Article 6(1)(B) GDPR in the Context of the Provision of Online Services to Data Subjects."

Ghostery (2017), "What Are Trackers?," Ghostery Blog, [https://www.ghostery.com/blog/what-are-trackers].

Godinho de Matos, Miguel and Idris Adjerid (2022), "Consumer Consent and Firm Targeting After GDPR: The Case of a Large Telecom Provider," Management Science, 68 (5), 3330–78.

Goldberg, Samuel G., Garrett A. Johnson, and Scott K. Shriver (2024), "Regulating Privacy Online: An Economic Evaluation of the GDPR," American Economic Journal: Economic Policy, 16 (1), 325–58.

Goldfarb, Avi and Catherine E. Tucker (2011), "Privacy Regulation and Online Advertising," Management Science, 57 (1), 57–71.

Goldfarb, Avi, Catherine Tucker, and Yanwen Wang (2022), "Conducting Research in Marketing with Quasi-Experiments," Journal of Marketing, 86 (3), 1–20.

Gopal, Ram D., Hooman Hidaji, Raymond A. Patterson, Erik Rolland, and Dmitry Zhdanov (2018), "How Much to Share with Third Parties? User Privacy Concerns and Website Dilemmas," MIS Quarterly, 42 (1), 143-A25.

Gopal, Ram D., Hooman Hidaji, Sule Nur Kutlu, Raymond A. Patterson, and Niam Yaraghi (2023), "Law, Economics, and Privacy: Implications of Government Policies on Website and Third-Party Information Sharing," Information Systems Research, 34 (4), 1375–97.

Greif, Björn (2017), "WhoTracks.me: Find Out Where You're Being Tracked on the Web," Cliqz Blog, [https://cliqz.com/en/magazine/whotracks-me-find-out-where-youre-being-tracked-on-the-web].

Huntington-Klein, Nick (2022), The Effect: An Introduction to Research Design and Causality, Chapman and Hall/CRC.

Jerath, Kinshuk, and Klaus M. Miller (2024), "Consumers' Perceived Privacy Violations in Online Advertising," Working Paper.

Johnson, Garrett (2023), "Economic Research on Privacy Regulation: Lessons from the GDPR and Beyond," National Bureau of Economic Research (NBER).

Johnson, Garrett, Scott Shriver, and Samuel Goldberg (2023), "Privacy & Market Concentration: Intended & Unintended Consequences of the GDPR,c Management Science, 69 (10), 1–27.

Johnson, Garrett (2024), "Unearthing Privacy-Enhanced Ad Technologies (PEAT): The Adoption of Google's Privacy Sandbox," Working Paper.

Kannan, P.K. and Hongshuang "Alice" Li (2017), "Digital Marketing: A Framework, Review and Research Agenda," International Journal of Research in Marketing, 34 (1), 22–45.

Karaj, Arjaldo, Sam Macbeth, Rémi Berson, and Josep M. Pujol (2018a), "WhoTracks.me: Shedding Light on the Opaque World of Online Tracking," ArXiv.

Karaj, Arjaldo, Sam Macbeth, Rémi Berson, and Josep M. Pujol (2018b), "WhoTracks.me: Shedding Light on the Opaque World of Online Tracking," ArXiv.

Kosinski, Michal, David Stillwell, and Thore Graepel (2013), "Private Traits and Attributes Are Predictable From Digital Records of Human Behavior," Proceedings of the National Academy of Sciences, 110 (15), 5802–5.

Laub, Rene, Klaus M. Miller, and Bernd Skiera (2024), "The Economic Value of User Tracking for Publishers," Working Paper.

Lefrere, Vincent, Logan Warberg, Cristobal Cheyre, Veronica Marotta, and Alessandro Acquisti (2024), "Does Privacy Regulation Harm Content Providers? A Longitudinal Analysis of the Impact of the GDPR," Working Paper.

Lerner, Adam, Anna Kornfeld Simpson, Tadayoshi Kohno, and Franziska Roesner (2016), "Internet Jones and the Raiders of the Lost Trackers: An Archaeological Study of Web Tracking from 1996 to 2016," Proceedings of the 25th USENIX Security Symposium (USENIX Security 16), 997–1013.

Libert, Timothy and Rasmus Kleis Nielsen (2018), "Third-Party Web Content on EU News Sites: Potential Challenges and Paths to Privacy Improvement," Reuters Institute for the Study of Journalism, 1–11.

Lin, Tesary (2022), "Valuing Intrinsic and Instrumental Preferences for Privacy," Marketing Science, 41 (4), 235–53.

Lobschat, Lara, Benjamin Mueller, Felix Eggers, Laura Brandimarte, Sarah Diefenbach, Mirja Kroschke, and Jochen Wirtz (2021), "Corporate Digital Responsibility," Journal of Business Research, 122, 875–88.

Martin, Kelly D., Abhishek Borah, and Robert W. Palmatier (2017), "Data Privacy: Effects on Customer and Firm Performance," Journal of Marketing, 81 (1), 36–58.

Mayer, Jonathan R. and John C. Mitchell (2012), "Third-Party Web Tracking: Policy and Technology," in 2012 IEEE Symposium on Security and Privacy, IEEE Computer Society, 413–27.

Miller, Klaus M. and Bernd Skiera (2024), "Economic Consequences of Online Tracking Restrictions: Evidence From Cookies," International Journal of Research in Marketing, 42 (2), 241–64.

Miller, Klaus M., Julia Schmitt, and Bernd Skiera (2024), "The Impact of the General Data Protection Regulation (GDPR) on Online Usage Behavior," Working Paper.

Müller-Tribbensee, Timo (2024), "Privacy Promise vs. Tracking Reality in Pay-or-Tracking Walls," in Privacy Technologies and Policy, M. Jensen, C. Lauradoux, and K. Rannenberg, eds., Cham: Springer Nature Switzerland, 168–88.

Müller-Tribbensee, Timo, Klaus M. Miller, and Bernd Skiera (2024), "Paying for Privacy: Pay-or-Tracking Walls," Working Paper.

Peukert, Christian, Stefan Bechtold, Michail Batikas, and Tobias Kretschmer (2022), "Regulatory Spillovers and Data Governance: Evidence from the GDPR," Marketing Science, 41 (4), 318–40.

Ponte, Gilian R., Jaap E. Wieringa, Tom Boot, and Peter C. Verhoef (2024), "Where's Waldo? A framework for quantifying the privacy-utility trade-off in marketing applications," International Journal of Research in Marketing, 41 (3), 529–46.

Rocher, Luc, Julien M. Hendrickx, and Yves-Alexandre de Montjoye (2019), "Estimating the Success of Re-Identifications in Incomplete Datasets Using Generative Models," Nature Communications, 10 (1), 3069.

Sanchez-Rola, Iskander, Matteo Dell'Amico, Platon Kotzias, Davide Balzarotti, Leyla Bilge, Pierre-Antoine Vervier, and Igor Santos (2019), "Can I Opt Out Yet?," in Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security, New York, NY, USA: Association for Computing Machinery (ACM), 340–51.

Schumacher, Christopher, Felix Eggers, Peter C. Verhoef, and Peter Maas (2023), "The Effects of Cultural Differences on Consumers' Willingness to Share Personal Information," Journal of Interactive Marketing, 58 (1), 72–89.

Schumann, Jan H., Florian von Wangenheim, and Nicole Groene (2014), "Targeted Online Advertising: Using Reciprocity Appeals to Increase Acceptance among Users of Free Web Services," Journal of Marketing, 78 (1), 59–75.

Skiera, Bernd, Klaus Miller, Yuxi Jin, Lennart Kraft, René Laub, and Julia Schmitt (2022), The Impact of the General Data Protection Regulation (GDPR) on the Online Advertising Market, Frankfurt: Self-Published.

Sweeny, Latanya (2002), "Simple Demographics Often Identify People Uniquely, "Working Paper.

Tucker, Catherine E. (2012), "The Economics of Advertising and Privacy," International Journal of Industrial Organization, 30 (3), 326–29.

Verhoef, Peter C., Koert van Ittersum, P. K. Kannan, and Jeff Inman (2022), "Omnichannel Retailing: A Consumer Perspective.," in APA Handbook of Consumer Psychology, Washington, DC, US: American Psychological Association, 649–72.

Wang, Pengyuan, Li Jiang, and Jian Yang (2024), "The Early Impact of GDPR Compliance on Display Advertising: The Case of an Ad Publisher," Journal of Marketing Research, 61 (1), 70–91.

WhoTracks.me (n.d.), "WhoTracks.me: Bringing Transparency to Online Tracking," GitHub, [https://github.com/whotracksme/whotracks.me].

Wieringa, Jaap, P.K. Kannan, Xiao Ma, Thomas Reutterer, Hans Risselada, and Bernd Skiera (2021), "Data Analytics in a Privacy-Concerned World," Journal of Business Research, 122, 915–25.

Yan, Shunyao, Klaus M. Miller, and Bernd Skiera (2022), "How Does the Adoption of Ad Blockers Affect News Consumption?," *Journal of Marketing Research*, 59 (5), 1002–18.

Figure 1: Main Actors Involved in Online Tracking

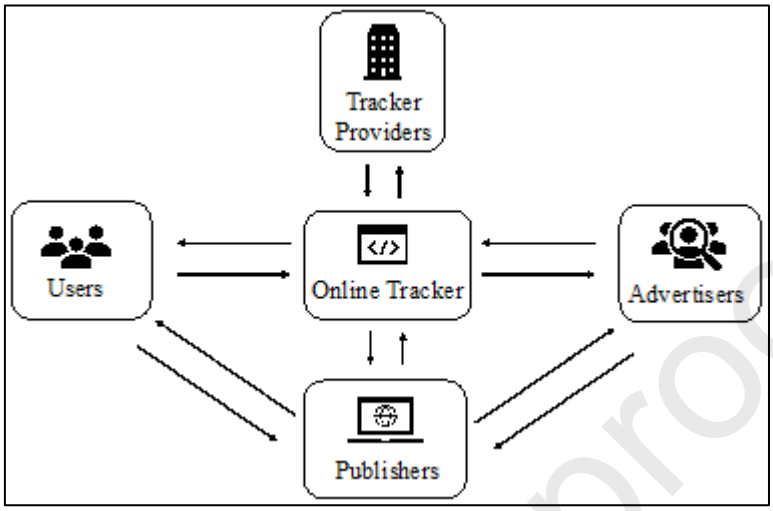
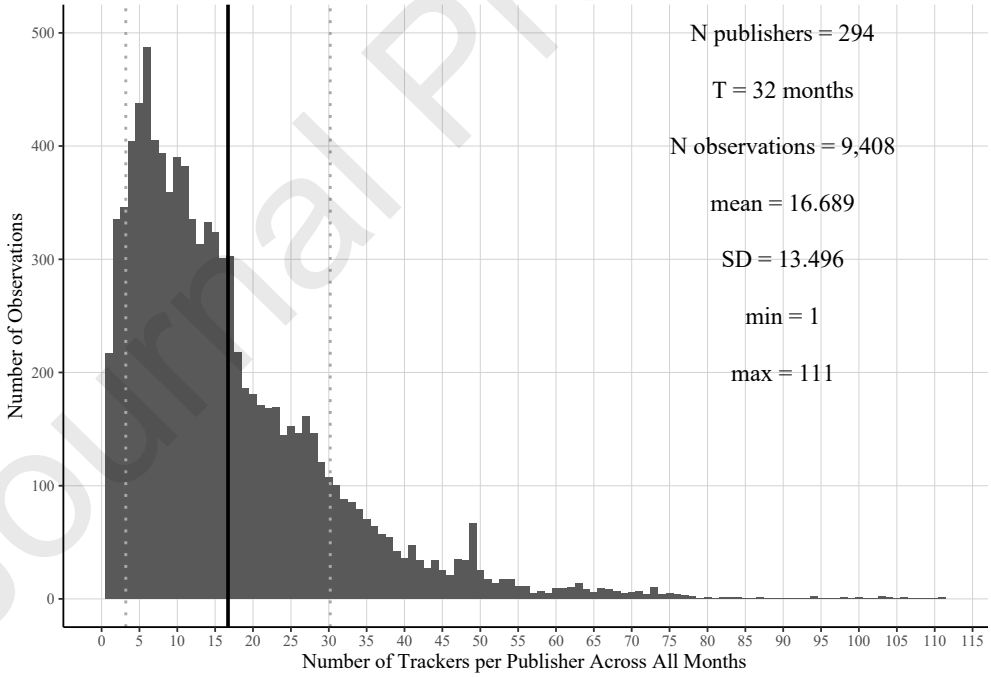
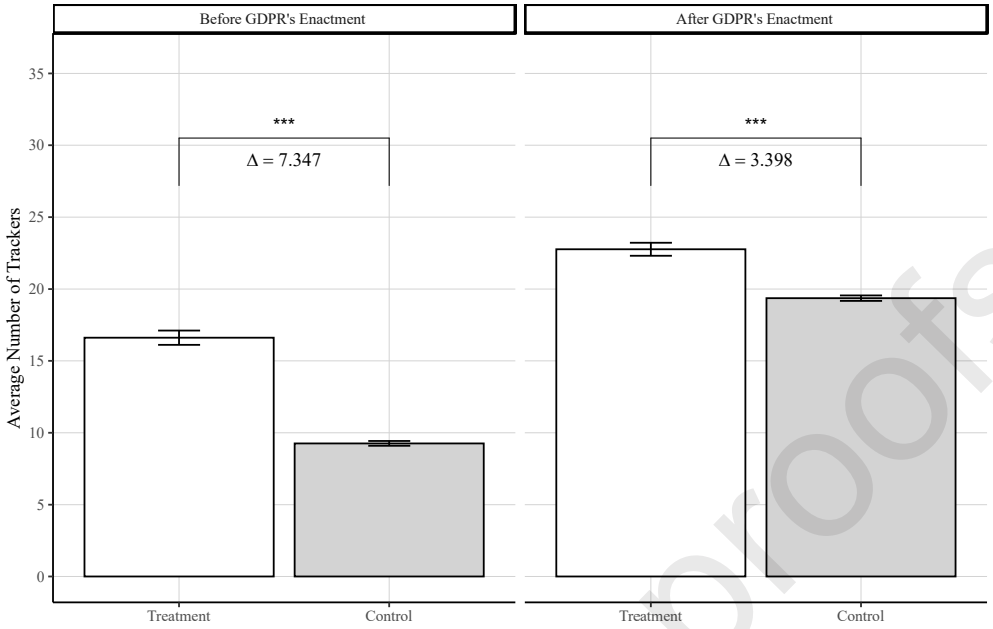


Figure 2: Distribution of the Average Number of Trackers per Publisher



Notes: Multiplying the number of publishers (N publishers = 294) and the number of months (T = 32 months) yields the number of observations (N observations = 9,408). The black vertical line indicates the mean number of trackers per publisher, while the gray lines represent ± one standard deviation from the mean.

Figure 3: Comparison of the Average Number of Trackers in the Treatment and Control Groups Before and After the GDPR's Enactment



Significance levels: * p < 0.05, ** p < 0.01, *** p < 0.001.
Notes: Error bars = +/− 1 SEs. This figure shows independent t−test comparisons between group averages in periods before (May 2017−April 2018) and after (May 2018−December 2019) the GDPR' s enactment using the number of trackers as a dependent variable.

Figure 4: Change in the Number of Online Trackers in the Treatment and Control Group

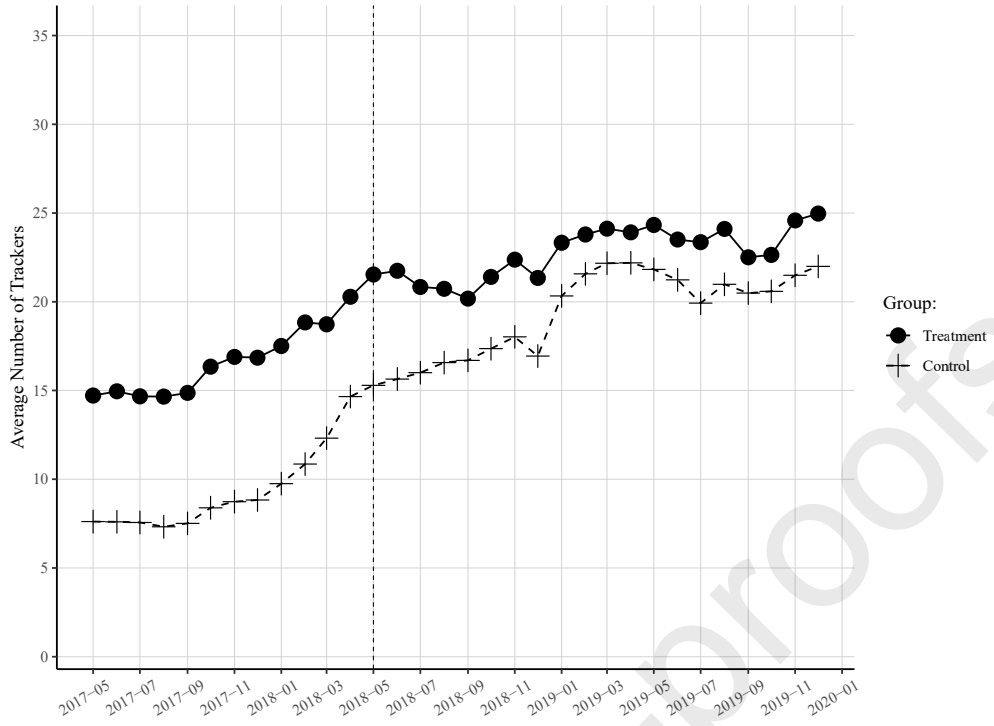
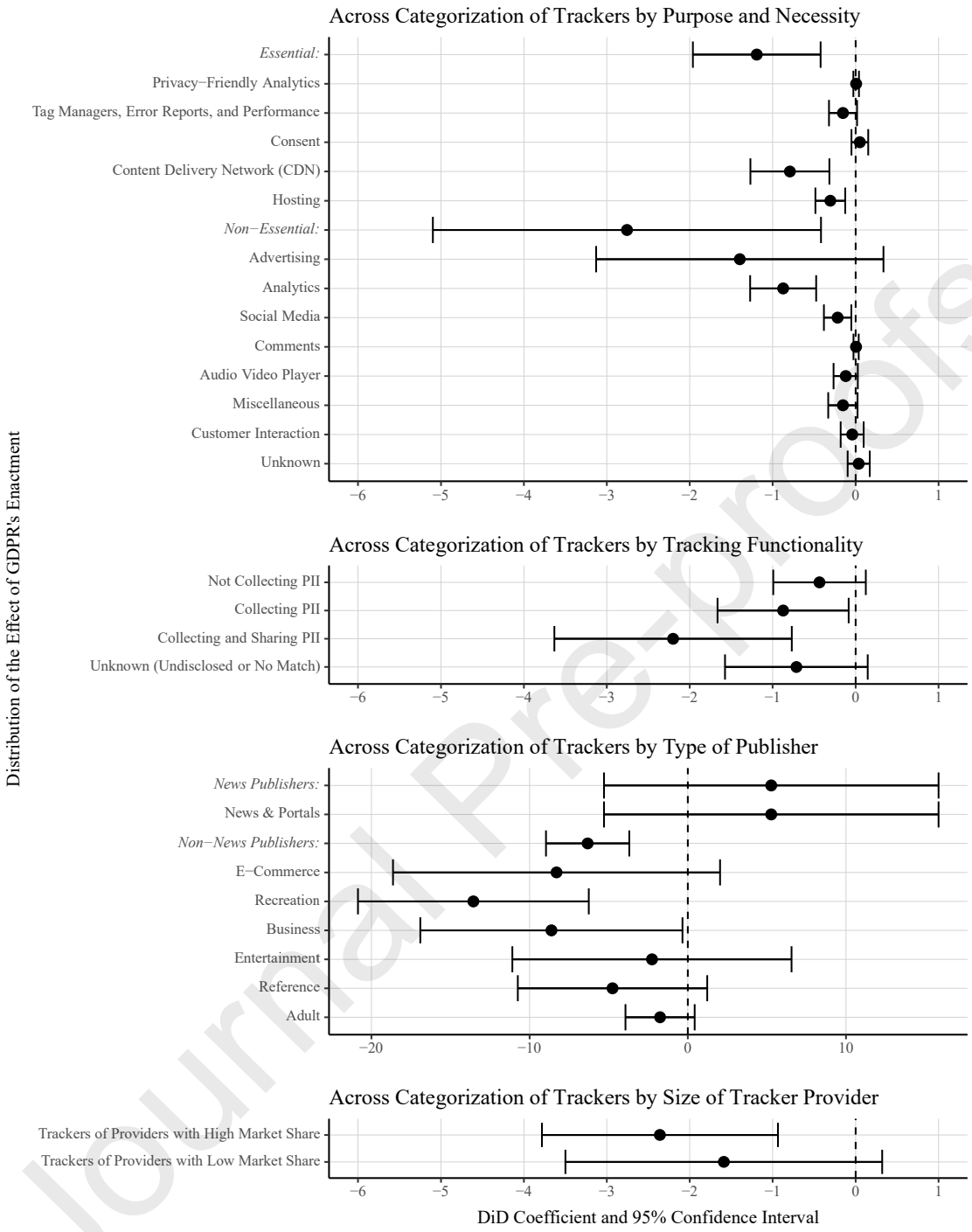


Figure 5: Distribution of the GDPR's Impact Across Categorizations of Trackers



Notes: This figure shows the difference−in−differences coefficients (Treatment x PostGDPR) from OLS regressions and 95% confidence intervals, with the dependent variables being the number of trackers across different categorizations of trackers. Italicized labels represent grouped variables, where a broad category estimation (e.g., "Essential:") is followed by estimations for subcategories within that group (e.g., "Privacy−Friendly Analytics"). Except for the categorization by Type of Publisher, all models have N = 9,408 observations (294 publishers * 32 months). For the categorization by Type of Publisher, the number of observations varies: News Publishers have N = 928 (29 publishers * 32 months), Non−News Publishers have N = 6,328 (198 publishers * 32 months), and specific types of publishers are as follows: News & Portals (N = 928), E−Commerce (N = 480), Recreation (N = 224), Business (N = 2,048), Entertainment (N = 2,432), Reference (N = 736), and Adult (N = 2,528). The Governmental publisher type is excluded due to having only one publisher. All models include website instance and month fixed effects. Two−way standard errors are clustered at the publisher and month levels.

Table 1: Categorization of Online Trackers by Purpose and Necessity

Purpose	Description of Purpose	Examples of Trackers	Defined By	Necessity	Description of Necessity
Privacy-Friendly Site Analytics	Collects and analyses data related to website usage and performance.	Piwik Pro, eTracker, eStat	CNIL	Essential	Strictly necessary for the basic functionality of the website. Exempt from user consent requirement under GDPR.
Tag Managers, Error Reports and Performance	Site requests that may be critical to website functionality, such as tag manager, privacy notices, error reports and performance.	Google Tag Manager, Google Recaptcha, Adobe Typekit	WhoTracks.me		
Consent	Cookie consent managers allow websites to track different levels of user activity.	OneTrust, Cookiebot, IAB Consent	WhoTracks.me		
Content Delivery Network (CDN)	Delivers resources for different site utilities and usually for many other customers.	Amazon CDN, CloudFlare, jQuery	WhoTracks.me		
Hosting	Service used by the content provider or site owner.	GitHub Pages, FastPic, Amazon CloudFront	WhoTracks.me		
Advertising	Provides advertising or advertising-related services such as data collection, behavioral analysis, or re-targeting.	DoubleClick, ShareThis, Experian Marketing Services	WhoTracks.me	Non-Essential	Not strictly necessary for the basic functionality of the website. Not exempt from user consent requirement under GDPR.
Site Analytics	Collects and analyzes data related to website usage and performance.	Google Analytics, Adobe Analytics, Hotjar	WhoTracks.me		
Social Media	Integrates features related to social media sites.	Facebook Social Plugins, Giphy, Twitter	WhoTracks.me		
Comments	Enables comments sections for articles and product reviews.	Disqus, eKomi, Livefyre	WhoTracks.me		
Audio Video Player	Enables websites to publish, distribute, and optimize video and audio content.	YouTube, Twitch, Spotify	WhoTracks.me		
Miscellaneous	This tracker does not fit into other categories.	Autoscout24, Oracle RightNow, Vinted	WhoTracks.me		
Customer Interaction	Includes chat, email messaging, customer support, and other interaction tools.	PayPal, Google Translate, LiveChat	WhoTracks.me		
Unknown	This tracker has either not been labeled yet or does not have enough information to label it.	boudja.com, xen-media.com, statsy.net	WhoTracks.me		

Table 2: Summary of Key Findings of Related Literature and Our Contribution

Literature Stream	Main Studies	Key Findings	Our Contribution
1. User Privacy Concerns	Eggers et al. (2023), Beke et al. (2022), Bleier et al. (2020), Schumacher et al. (2023), Lobschat et al. (2021), Martin et al. (2017), Wieringa et al. (2021), Gopal et al. (2023), Kannan and Li (2017), Beke et al. (2018), Schumann et al. (2014), Tucker (2012), Ahamdi et al. (2024), Verhoef et al. (2022), Rocher et al. (2019), Sweeney (2002), Dinur and Nissim (2003), Lin (2022), Jerath and Miller (2024)	• **Privacy calculus & paradox:** Users weigh costs (data misuse) against benefits (personalization, convenience), often sharing data despite stating serious concerns. • **Contextual integrity:** Comfort with sharing data depends on the perceived alignment of data use with user expectations; violations heighten privacy concerns. • **Transparency & Control:** Clear explanation of data practices and user control features consistently mitigate privacy concerns and build trust. • **Personalized ads:** While valuable, they can exacerbate privacy concerns if users feel uninformed about how firms use or share their data. • **Corporate digital responsibility:** Proactive privacy measures and accountability in data handling can reduce user unease; however, re-identification risks persist as data analytics advances.	• Examine publishers' privacy practices, particularly large-scale information collection (see Beke et al. (2018), by analyzing online advertising and the GDPR's impact thereon, specifically for high-risk trackers. • Determine effectiveness of privacy regulation in reducing number of trackers, which may help mitigate privacy concerns (Martin et al. 2017, Gopal et al. 2023). • Determine how different types of publishers respond to GDPR (Beke et al. 2022; Lobschat et al. 2021).
2. Describing Online Trackers	Mayer and Mitchell (2012), Lerner et al. (2016), Karaj et al. (2018b)	• **High pervasiveness of trackers pre-GDPR:** Multiple trackers per publisher have become the norm, funding free content and enabling data-driven services. • **Market concentration:** A small group of dominant tracker providers (e.g., Google, Facebook) is embedded on most websites, reflecting an increasing	• Examine online tracking from 2017 to 2019, extending earlier research on older periods. • Assess regulatory interventions on trackers and how GDPR impacts tracker usage across publisher types.

Literature Stream	Main Studies	Key Findings	Our Contribution
		consolidation and raising privacy concerns. • **Power imbalance:** Large tracker providers like Google Analytics, DoubleClick, and Facebook often collect data on vast swaths of user traffic-up to 70-80%- highlighting their expansive reach.	• Impact of GDPR on market concentration.
3. Impact of Privacy Regulation on Online Trackers	Goldfarb and Tucker (2011), Peukert et al. (2022), Johnson et al. (2023), Godinho De Matos and Adjerid (2022), Wang et al. (2024), Goldberg et al. (2024), Laub et al. (2024), Lefrere et al. (2024), Miller et al. (2024), Miller and Skiera (2024)	• **Initial decline & rebound:** Several studies report a short-term drop in tracker usage up to 3 months post-GDPR and a rebound thereafter. • **Market concentration**: Larger tracker providers often handle compliance costs more efficiently, potentially reinforcing their dominance. • **Consent mechanisms**: GDPR-compliant banners can raise user consent rates, sometimes enhancing targeted marketing effectiveness rather than diminishing it. • **Limited negative impact on engagement**: Some EU publishers see no decline in user engagement or content provision despite fewer trackers, making the overall effect on the online advertising market uncertain. • **Tension with ad revenues:** Reduced tracker usage and reduced access to user data may lower ad-targeting effectiveness and publisher revenues; some publishers compensate via alternative approaches (e.g., via contextual or first-party data).	• Expand prior studies that relied on web crawlers primarily simulating user behavior. We use data from actual users who automatically reported the trackers they encountered. • Document actual user exposure to different trackers, providing more nuanced assessment of GPDR's effects on user privacy. • Assessment of number of trackers may reflect users' privacy concerns.

Table 3: Description of the Data Sets

Data Set	Contained Information	Period	Purpose
WhoTracks.me	- Publishers' use of trackers - Trackers (e.g., purpose, tracker provider) - Monthly data for 294 publishers over 32 months - Balanced panel of 9,408 observations (294 publishers * 32 months) - Information about publisher types - For each publisher, top-level domain used to categorize as EU vs. Non-EU, in combination with SimilarWeb data	05/2017 - 12/2019	Main data set to empirically describe trackers and measure impact of GDPR's enactment on trackers
SimilarWeb	- Traffic shares from the top five (EU and non-EU) countries - Information on 294 out of 294 (100%) publishers in the balanced panel	08/2021	Augments WhoTracks.me data set to categorize publishers as EU vs. Non-EU based on majority of traffic shares
	- Daily-level information on traffic shares for 7,332 publishers - Traffic shares for US users and specific EU countries - Information on 200 out of 294 (68%) publishers in the balanced panel	01/2018 - 12/2019	Augments public SimilarWeb data set to check the consistency of publisher's website traffic distribution over time
Evidon	- Information on trackers from their privacy policies	03/2021	Augments WhoTracks.me data set to categorize trackers based on tracking

- 724 (76%) matched trackers of 949 unique trackers from WhoTracks.me - 546 (75%) of 724 disclose data collection and sharing practices - 35 (4%) disclose only data sharing, 0 disclose only data collection, 143 (15%) disclose neither practices - 225 (24%) trackers do not match	functionality from their disclosed data collection and sharing practices

Table 4: Steps Taken to Prepare the Sample of 294 Publishers

Step	Number of Publishers	Percent Change
Raw global sample (unbalanced; average number of publishers released monthly)	8,334	
Balanced global sample (May 2017 to December 2019)	962	-88.46%
Raw EU/US sample (unbalanced; average number of publishers released monthly)	7,264	
Balanced EU/US sample (April 2018 to December 2019)	717	-90.13%
Publishers present in both global and EU/US samples	354	-63.20%
Removing outliers in the control group of the global sample (ensuring parallel trends assumption)	294	-16.95%

Table 5: Distribution of Observations (Monthly Publishers) Across Publisher Designation

Publisher Designation	Number and Percentage of Observations
EU publisher[1]	2,144 (22.79%)
Non-EU publisher[2]	7,264 (77.21%)
∑	9,408 (100.00%)

[1]A publisher is designated as an "EU publisher" if (1) the publisher uses an EU top-level domain (e.g., .de) or (2) the publisher receives more traffic from EU users than non-EU users. [2]A publisher is designated as a "non-EU publisher" if (1) the publisher uses a non-EU top-level domain (e.g., .com) and (2) the publisher receives more traffic from non-EU users than EU users.

Notes: The cells in this table show the number and percentage of observations in our sample corresponding to each case. The cell belonging to the control group—where GDPR does not apply—is colored gray, and the cell belonging to the treatment group—where GDPR applies—is not colored. In total, 23% (N observations = 2,144) of all observations (N observations = 9,408) belong to the treatment group and 77% (N observations = 7,264) to the control group.

Table 6: Distribution of the Average Number of Trackers per Publisher By Categorizations of Trackers

	Number of Trackers per Publisher Across All Months			
Categorization of Trackers by Purpose and Necessity	mean	SD	min	max
Essential:	*4.432*	*3.232*	*0*	*20*
Privacy-Friendly Analytics	0.032	0.215	0	3
Tag Managers, Error Reports and Performance	0.774	0.886	0	5
Consent	0.139	0.438	0	4
Content Delivery Network (CDN)	2.931	2.127	0	12
Hosting	0.556	0.759	0	4
Non-Essential:	*12.258*	*11.125*	*0*	*92*
Advertising	7.257	8.103	0	76
Analytics	2.864	2.386	0	18
Social Media	0.659	0.937	0	8
Comments	0.064	0.248	0	2
Audio Video Player	0.408	0.731	0	5
Miscellaneous	0.452	0.828	0	6
Customer Interaction	0.404	0.785	0	6
Unknown	0.181	0.502	0	5
Categorization of Trackers by Tracking Functionality				
Not Collecting PII	1.872	2.641	0	25

	Number of Trackers per Publisher Across All Months			
Categorization of Trackers by Purpose and Necessity	mean	SD	min	max
Collecting PII	3.258	3.553	0	31
Collecting and Sharing PII	8.424	6.764	0	46
Unknown (Undisclosed or No Match)	4.227	3.519	0	32
Categorization of Trackers by Type of Publisher				
News Publishers:	*28.902*	*19.044*	*1*	*111*
News & Portals	28.902	19.044	1	111
Non-News Publishers:	*15.353*	*12.013*	*1*	*103*
E-Commerce	24.971	13.166	1	71
Recreation	18.781	10.405	1	51
Business	18.682	13.991	1	77
Entertainment	16.947	12.694	1	103
Reference	13.803	11.282	1	78
Adult	9.545	5.075	1	33
Government	7.281	3.429	2	11
Categorization of Trackers by Size of Tracker Provider				
Trackers of Providers with High Market Share	8.351	5.708	0	30

	Number of Trackers per Publisher Across All Months			
Categorization of Trackers by Purpose and Necessity	mean	SD	min	max
Trackers of Providers with Low Market Share	8.338	9.087	0	83

Notes: This table displays descriptive statistics for the number of trackers per publisher across all months and types of tracker categorizations. Italicized labels represent grouped variables, where category descriptives (e.g., "Essential:") are followed by descriptives for subcategories within that group (e.g., "Privacy-Friendly Analytics"). Multiplying the number of publishers (N publishers = 294) and the number of months (T = 32 months) yields the number of observations (N observations = 9,408).

Table 7: Distribution of the Average Number of Trackers per Publisher By Categorizations of Trackers in the Treatment and Control Groups

	Treatment Group	Control Group	Difference	(%)
Number of Trackers per Publisher Across All Months	20.457	15.577	4.879	(31.32%)
Categorization of Trackers by Purpose and Necessity				
Essential:	5.078	4.241	0.838	(19.75%)
Privacy-Friendly Analytics	0.124	0.005	0.119	(2,380.00%)
Tag Managers, Error Reports and Performance	0.857	0.750	0.107	(14.28%)
Consent	0.145	0.137	0.008	(6.11%)
CDN	3.280	2.828	0.452	(15.98%)
Hosting	0.673	0.522	0.151	(28.96%)
Non-Essential:	15.378	11.337	4.042	(35.65%)
Advertising	10.170	6.397	3.773	(58.97%)
Analytics	3.049	2.810	0.239	(8.51%)
Social Media	0.505	0.704	-0.199	(-28.25%)
Comments	0.075	0.061	0.015	(23.97%)
Audio Video Player	0.443	0.397	0.045	(11.45%)
Miscellaneous	0.545	0.425	0.120	(28.19%)
Customer Interaction	0.424	0.398	0.025	(6.34%)

	Treatment Group	Control Group	Difference	(%)
Number of Trackers per Publisher Across All Months	20.457	15.577	4.879	(31.32%)
Categorization of Trackers by Purpose and Necessity				
Unknown	0.292	0.149	0.143	(95.71%)
Categorization of Trackers by Tracking Functionality				
Not Collecting PII	3.064	1.520	1.544	(101.56%)
Collecting PII	3.779	3.104	0.676	(21.77%)
Collecting and Sharing PII	9.623	8.070	1.553	(19.24%)
Unknown (Undisclosed or No Match)	5.888	3.737	2.150	(57.54%)
Categorization of Trackers by Type of Publisher				
News Publishers:	34.833	22.547	12.286	(54.49%)
News & Portals	34.833	22.547	12.286	(54.49%)
Non-News Publishers:	16.309	15.119	1.190	(7.87%)
E-Commerce	25.554	24.461	1.093	(4.47%)
Recreation	18.711	18.875	-0.164	(-0.87%)
Business	29.823	18.134	11.689	(64.46%)
Entertainment	17.545	16.887	0.658	(3.90%)

	Treatment Group	Control Group	Difference	(%)
Number of Trackers per Publisher Across All Months	20.457	15.577	4.879	(31.32%)
Categorization of Trackers by Purpose and Necessity				
Reference	13.812	13.800	0.012	(0.09%)
Adult	12.040	8.322	3.718	(44.67%)
Categorization of Trackers by Size of Tracker Provider				
Trackers of Providers with High Market Share	9.925	7.887	2.039	(25.85%)
Trackers of Providers with Low Market Share	10.531	7.691	2.840	(36.93%)
Publisher Characteristics				
Share of Traffic from EU Users	48.37%	10.24%		(38.12 pp)
Share of Traffic from Non-EU Users	15.63%	43.28%		(-27.65 pp)
5 Most Common TLDs	com, co.uk, de, fr, net	com, net, org, ru, tv		

Notes: This table shows the average number of trackers for the treatment and control groups across all months and types of tracker categorizations. Italicized labels represent grouped variables, where broad category descriptives (e.g., "Essential:") are followed by descriptives for subcategories within that group (e.g., "Privacy-Friendly Analytics"). The table also shows the average share of traffic from (non)-EU users and the five most common TLDs for treatment and control groups. Percent differences are displayed as percentage points (pp) for shares of traffic from (non)-EU users. The Government publisher has been deliberately omitted from this analysis, given that only a single publisher of this type was present in the control group of our sample.

Table 8: Average (Monthly) Number of Trackers in the Treatment and Control Groups Before and After the GDPR's Enactment

Group	Before GDPR's Enactment	After GDPR's Enactment	Difference
Treatment	16.610	22.765	6.155
Control	9.262	19.366	10.104
Difference	7.347	3.398	-3.949 (14.79% decrease)

Notes: This table shows the average (monthly) number of trackers for the treatment and control groups in periods before (May 2017-April 2018) and after (May 2018-December 2019) GDPR's enactment and the differences in the average (monthly) number of trackers between groups and periods. We use unrounded values to derive the differences. The values in parentheses represent the percent changes for each group from the period before to the period after the GDPR's enactment. The Difference-in-Differences (DiD) as a percentage is calculated by comparing the observed value in the treatment group after GDPR (22.765) with the expected value if the GDPR had not been enacted. The expected value is calculated by adding the pre-GDPR difference between groups (7.347) to the post-GDPR control group value (19.366), which equals 26.714. The percent decrease is then derived from the ratio of the difference between these two values to the expected value: $\text{DiD}\ (\%) = \frac{26.714 - 22.765}{26.714} \times 100 \approx 14.79\%$.

Table 9: Result of Difference-in-Differences (DiD) Analysis for the Number of Trackers

Dependent Variable:	Number of Trackers per Publisher and Month
Model:	(1)
Treatment x PostGDPR	-3.949* [-7.082; -0.816]
Publisher ID-Fixed Effects	✓
Month ID-Fixed Effects	✓
N Observations	9,408
R^2	0.744

Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Two-way standard errors are clustered at the publisher and month levels; 95% confidence intervals are reported in brackets.

Notes: This table shows the difference-in-differences coefficient (Treatment x PostGDPR) from the OLS regression. We assign treatment to each publisher according to the publisher's designation (EU or non-EU). Multiplying the number of publishers (N publishers = 294) and the number of months (T = 32 months) yields the number of observations (N observations = 9,408).

Table 10: Summary of Robustness Tests

Robustness Test	Fundamental Concern	Summary of Result	Web Appendix
Treatment assignment based on server location	Misclassification of publishers into treatment (EU) and control groups (non-EU) based on publisher's website traffic shares and top-level domain (TLD)	GDPR reduced the number of trackers by 3.867 per publisher with treatment assignment based on server location	9.2.1
Treatment assignment based on publisher designation and user location	Misclassification of publishers into treatment (EU) and control groups (non-EU) based on publisher's website traffic shares and top-level domain (TLD)	GDPR reduced the number of trackers by 1.692 per publisher instance with treatment assignment based on publisher designation and user location	9.2.2
Parallel trends assumption	Treatment and control groups do not follow same trends in the pre-treatment period (violation of parallel trends assumption)	Development of monthly DiD coefficients and placebo tests confirm the assumption likely holds	9.3.1
Spillover effects	GDPR spillovers affect control group (= violation of stable unit treatment value assumption)	GDPR reduced the number of trackers by 2.922 per publisher instance in the "cleanest" comparison between treatment (EU-located users visiting EU publishers) vs. control (US-located users visiting non-EU publishers) groups	9.3.2
Impact of GDPR on user behavior	GDPR inadvertently affects behavior of Ghostery users rather than publishers' use of trackers	No significant change in the number of Ghostery users (Chrome and Firefox) after GDPR	9.3.3
Anticipation and external shocks (early 2018)	Bias from publishers' early willingness to comply with GDPR (= anticipation assumption) or shocks unrelated to the GDPR (e.g., Cambridge Analytica)	GDPR reduced the number of trackers by 4.523 per publisher when removing the months of March, April, May and June 2018	9.3.4
Skewness of the dependent variable	Skewness in the distribution of the number of trackers	GDPR reduced the logged number of trackers by 0.490 per publisher	9.4.1
Stability of publishers' website traffic shares	Misclassification of publishers due to potential changes in website traffic distributions over time when using a single point-in-time SimilarWeb data set	The average difference for EU publishers' website traffic shares was 11.08 pp, indicating stable website traffic distributions between public (single point in time) and proprietary (over time) SimilarWeb data sets	9.4.3

Robustness Test	Fundamental Concern	Summary of Result	Web Appendix
Generalized synthetic control method	Potential model misspecifications in the difference-in-differences (DiD) analysis	GDPR reduced the number of trackers by 5.303 per publisher	9.5.1
Unbalanced panel	Potential lack of representativeness due to excluding a large number of publishers from the balanced panel	GDPR reduced the number of trackers by 1.081 (treatment assignment based on TLD) and 0.825 (treatment assignment based on server location) per publisher in the unbalanced panel of 29,735 unique publishers	9.5.2

Table 11: Summary of Empirical Findings on Description of Online Trackers and Their Conclusions

Analysis	Summary of Findings (Full Oberservation Period Before and After GDPR)	Conclusions
Average Effect	• Average number of trackers per publisher (~17) • Most publishers use 1-10 trackers (Min= 1, Max = 111)	• Some publishers strongly rely on trackers. • Distribution of trackers is heavily right-skewed.
	Differences across categorizations of online trackers	
Trackers by Necessity	• 27% essential trackers per publisher (~4) • 73% non-essential trackers per publisher (~12)	• Publishers use three times as many non-essential than essential trackers. • Users are exposed to privacy risks from non-essential trackers.
Trackers by Purpose	• Top essential trackers: ○ 66% content delivery (~3)	• Advertising, analytics, and content delivery trackers are most often used.

Analysis	Summary of Findings (Full Oberservation Period Before and After GDPR)	Conclusions
	○ 17% tag managers (~1) ○ 13% hosting (~1) • Top non-essential trackers: ○ 59% advertising (~7) ○ 23% analytics (~3) ○ 5% social media (~1) • Among essential trackers: ○ 0.7% privacy-friendly analytics (<1)	• Publishers rarely use privacy-friendly analytics trackers.
Trackers by Functionality	• 11% of trackers do not collect personal data (~2) • 66% of trackers collect personal data ○ 28% of those trackers do not share personal data (~3) ○ 72% of those trackers share personal data (~8)	• Most trackers are highly privacy-invasive as they collect and share personal data.
Trackers by Type of Publisher	• 67% of trackers belong to news publishers (~30) • 33% of trackers belong to non-news publishers (~15)	• News publishers use twice as many trackers as non-news publishers. • News publishers rely on trackers to enhance and monetize their content through advertising.
Trackers by Size	• 50% of trackers belong to providers with a high market share (~8) • 50% of trackers belong to providers with a low market share (~8)	• Publishers use a similar amount of trackers from tracker providers with a high or low market share. • Across all trackers, our study does not find evidence for market concentration of large or small trackers.

Table 12: Summary of our Empirical Findings of Impact of GDPR on the Number of Online Trackers and Their Conclusions

Analysis	Summary of Findings[A]	Conclusions
Average Effect	• Average reduction of trackers (~4)	• GDPR reaches its intended consequence and decreases trackers by 14.79% compared to expectations without GDPR. • Although trackers increased over time across EU and non-EU publishers before and after the GDPR, the increase is much smaller for EU publishers.
	Differences across categorizations of online trackers	
Trackers by Necessity	• Average reduction of essential trackers (~1) • Average reduction of non-essential trackers (~3)	• GDPR led to the unintended consequence of decreasing essential trackers. • GDPR reached its intended consequence of decreasing non-essential trackers. • GDPR did not reach its intended consequence of decreasing advertising trackers and increasing privacy-friendly analytics trackers.

Trackers by Purpose	• Average reduction of essential trackers
	  ○ Content Delivery (~1)
	  ○ Hosting (~1)
	  ○ Privacy-friendly analytics (~0)
	• Average reduction of non-essential trackers
	  ○ Analytics (~1)
	  ○ Social Media (~1)
	  ○ Advertising (~0)	

Trackers by Functionality	• Average reduction of trackers that do not collect personal data (~0) • Average reduction of trackers that do collect personal data ○ Tracker does not share personal data (~1)	• GDPR achieved its intended consequence of decreasing highly privacy-invasive tracking.

Analysis	Summary of Findings[A]	Conclusions
	○ Tracker shares personal data (~2)	
Trackers by Type of Publisher	• Average reduction of trackers of news publishers (~0) • Average reduction of trackers of non-news publishers (~6) ○ Recreation (~14) ○ Business (~9) ○ E-commerce (~0) ○ Entertainment (~0)	• GDPR reached its intended consequence and decreased tracking of non-news publishers. • GDPR led to the unintended consequence of not decreasing trackers of news, e-commerce, and entertainment publishers.
Trackers by Size	• Reduction of average number of trackers of providers with high market share (~2) • Reduction of average number of trackers of providers with low market share (~0)	• GDPR reached its intended consequence of decreasing trackers of high market share tracker providers and did not increase market concentration. • GDPR reached the unintended consequence of not decreasing trackers of low market share tracker providers.

Notes: A) The summary of findings refers to the average reduction of trackers per EU publisher.