The Impact of the GDPR on German Online Behavior: An Analysis of Traffic, Cookie Compliance and Online Harms

Carla Dausend

Dissertation written under the supervision of
Professor Miguel Godinho de Matos.

Dissertation submitted in partial fulfillment of requirements for the MSc in Business Analytics at the Universidade Católica Portuguesa.

January 4, 2023

The Impact of the GDPR on German Online Behavior: An Analysis of Traffic, Cookie Compliance and Online Harms

Carla Dausend

January 4, 2023

Keywords: Privacy regulation, GDPR, Website traffic, Cookie Compliance, Online Harms

Abstract

The European General Data Protection Regulation (GDPR) was implemented in May 2018 to enhance privacy protection. Its implementation has changed the digital landscape and increased awareness of data privacy rights. This study assesses the impact of the GDPR on German online traffic. We analyzed traffic data from The Nielson Company for the top 1000 websites in Germany and manually created a dataset to track cookie compliance for these websites. We find no evidence that German traffic changed. The study used a difference-in-differences approach to compare three traffic patterns: (1) changes in German traffic to websites that show cookie consent notices compared to those that do not; (2) comparison of German traffic to the United States; and (3) comparison of German traffic inside to outside the EU. The results showed a decrease in pageviews per person and time per person when comparing German traffic to the United States. However, statistical analysis did not reveal a significant difference. The study also found that compliance with the GDPR occurred gradually over time, with non-compliance being more prevalent among websites with higher harm scores.

Acknowledgements

I would like to take this opportunity to express my gratitude to all those who provided continuous support during the creation of my master's thesis. Especially I want to thank my professor, Miguel Godinho de Matos, for his guidance, motivation and constructive feedback.

O Impacto do GDPR no Comportamento Online Alemão: Uma Análise do Tráfego, Cookie Compliance e Danos Online

Carla Dausend

January 4, 2023

Palavras-chave: Regulação da privacidade, GDPR, Tráfego do website, Cookie Compliance, Danos Online

Resumo

O Regulamento Geral Europeu de Protecção de Dados (GDPR) foi implementado em Maio de 2018 para reforçar a protecção da privacidade. A sua implementação alterou o panorama digital e aumentou o conhecimento dos direitos à privacidade dos dados. Este estudo avalia o impacto do GDPR no tráfego em linha alemão. Analisámos os dados de tráfego da The Nielson Company para os 1000 principais websites na Alemanha e criámos manualmente um conjunto de dados para acompanhar a conformidade dos cookies para estes websites. Não encontramos provas de que o tráfego alemão tenha mudado. O estudo utilizou uma abordagem por diferenças para comparar três padrões de tráfego: (1) alterações no tráfego alemão a websites que mostram avisos de consentimento de cookies em comparação com os que não mostram; (2) comparação do tráfego alemão com os Estados Unidos; e (3) comparação do tráfego alemão dentro para fora da UE. Os resultados mostraram uma diminuição de pageviews por pessoa e tempo por pessoa ao comparar o tráfego alemão com os Estados Unidos. No entanto, a análise estatística não revelou uma diferença significativa. O estudo concluiu também que o cumprimento do GDPR ocorreu gradualmente ao longo do tempo, sendo o não cumprimento mais prevalecente entre os websites com pontuações de danos mais elevadas.

Agradecimentos

Gostaria de aproveitar esta oportunidade para expressar a minha gratidão a todos aqueles que prestaram apoio e motivação durante a criação da minha tese de mestrado. Quero agradecer especialmente ao meu professor, Miguel Godinho de Matos, pela sua orientação e revisão minuciosa do meu trabalho. Estou profundamente agradecido pelas valiosas sugestões e pelo feedback construtivo que ele deu ao longo de todo o processo de preparação desta dissertação.

Contents

1 Introduction

On May 25th, 2018, the European Parliament passed the new General Data Protection Regulation (GDPR) to improve privacy protection and maintain the benefits of data processing. A standardized regulation was necessary to deal with the immense increase in online traffic and the associated data collection. The GDPR defines the use of personal data of businesses and public administration. It also grants individuals more rights related to their data, such as the right to access or erasure. This study investigates the consequences of this new law on online traffic.

Previous research has investigated changes in online consumer behavior due to the GDPR. These studies have generally examined overall website traffic rather than traffic broken down by country. The GDPR applies to any organization that processes the personal data of European citizens. Websites must implement design and functionality changes, such as adding consent banners or privacy settings. These changes led some websites outside the EU to block European users to avoid direct compliance with the GDPR, such as the Los Angeles Times[1] (see Figure 1). It negatively impacted online traffic from the supply side, meaning traffic went down as EU citizens could not access the websites anymore. This study examines the impact from the opposite perspective - the demand side.

Figure 1: Example of EU user blockage on a website due to GDPR

We specifically focus on the effects of the GDPR on German consumer traffic because Germany it is the most populous member of the European Union and thus subject to the regulation. We compare online traffic before and after the implementation. Our analysis is based on an aggregated dataset of the 1000 most popular websites from The Nielsen Company. It is a global information, data, and measurement company that provides market research and insights about what people browse on the internet, watch, and buy. The used dataset includes the online behavior of Germans and Americans from January 2018 (prior to GDPR)

[1] https://www.theguardian.com/technology/2018/may/25/gdpr-us-based-news-websites-eu-internet-users-la-times

to November 2018 (post-GDPR). We focus on the immediate short-term effects, analyzing the first six months after the enforcement of the regulation. To examine the impact, we contrast three different traffic patterns: (1) a comparison of German website traffic between sites that display cookie consent notices and those that do not; (2) a comparison of German traffic to websites in the United States; and (3) a comparison of German traffic within the EU versus outside of it. Figure 2 presents a short outline of our thesis topic development.

Figure 2: Thesis Topic Development

Instead of assuming that all websites would directly follow the regulations of the GDPR, we examined the actual implementation. We manually generated a dataset of websites checking if they display cookie notices (defined here as 'compliance'). We found that about half the websites instantly fulfilled the requirements, but the rest only followed later. Additionally, we assigned scores to the websites indicating if they were harmful. Harms include, for example, copyright infringement or spam. We discovered that non-compliance was higher for websites that contained harm.The findings help identify gaps in privacy regulation across the internet and improve compliance in the future.

To assess the impact of the GDPR on online traffic, we apply a difference-in-differences regression approach. We found no significant changes in German traffic since the GDPR went into effect. Statistical analysis did not show a difference in traffic between GDPR compliance and non-compliance. Because our compliance sample is relatively small, we also compared German to American traffic. It shows a slight decrease in pageviews and time per person but not in audience or pageviews per person. Lastly, comparing German traffic inside the EU to traffic outside the EU does not show an effect.

We give an overview of the traffic situation for website owners and consumers in Germany. The results can also give an outlook on what other countries can expect regarding website traffic in the future.

Our paper is structured as follows: Section 2 reviews the data privacy legislation. Section 3 contains previous literature and our research questions. Afterward, Section 4 presents the data. Section 5 describes our difference-in-differences model. In Section 6, we elaborate on the impact of GDPR on online traffic. Finally, Section 7 discusses findings and limitations.

2 Legislative framework

This section provides a brief overview of the evolution of privacy law in Europe and the United States, focusing on the most significant legislative changes.

2.1 European legislations

Privacy regulations in Europe have advanced significantly in the past years; Figure 3 depicts its evolution through time.

The first principles started with the 'European Convention on Human Rights in 1950. This convention grants the right to respect one's private and family life, home, and correspondence, for example, emails, phone calls, and messages (Council of Europe, 1950).

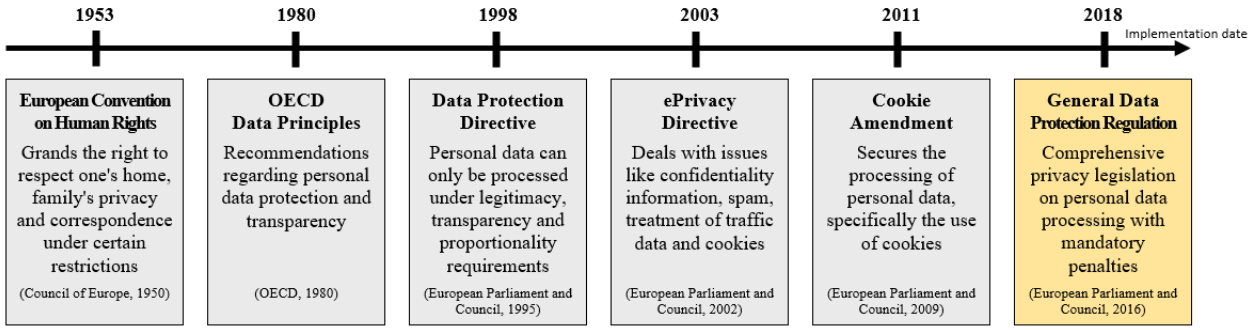

Figure 3: Timeline of privacy regulations in the European Union

Later in 1980, the Organisation for Economic Co-operation and Development (OECD) issued eight 'OECD Data Principles' to improve data protection. These principles state that one should place transparent protection. They also include suggestions for data access, erasure, and objection. However, these rules were only recommendations and not binding (OECD, 1980).

The 'Data Protection Directive' is the first obligatory law regulating the processing of personal data in the EU. It was implemented in 1998 and established fundamental rights that still apply today. The directive says that personal data should only be processed if the entities processing the data are transparent or have a legitimate reason for doing so. It is important to note that the EU directives are meant for member states and are usually not enforced against individuals (European Parliament and Council, 1995).

The Privacy and Electronic Communications Directive sometimes referred to as the "ePrivacy Directive," went into effect in 2003. It expanded the Data Protection Directive and dealt

with internet data traffic, consent, and cookies. The law was updated later to keep up with changes in technology. According to the Directive, users should know what information is stored and why. It became mandatory to ask for consent to store and process data. (European Parliament and Council, 2002).

Finally, the General Data Protection Regulation is known to be the strictest data privacy law worldwide. It was adopted by the EU in April 2016 and went into effect in May 2018. The GDPR applies to any data processing involving EU or EEA citizens. The directive is complex and contains a total of 99 articles and 173 recitals. It applies to all organizations that process data of natural persons, whether they are based inside or outside the EU. According to the GDPR, sensitive data includes a person's genetic, economic, cultural, or social identity. More specifically personal data are for example email address, social security number, IP address, telephone number, location, and birth date. In contrast to previous regulations, it introduces mandatory penalties for providers who violate the GDPR. The maximum fee is 4% of a company's global annual revenue or €20 million. The GDPR includes not only guidelines for companies and public institutions, but also specific instructions on how controllers and processors should conduct themselves in relation to data protection (European Parliament and Council of the European Union, 2018).

The GDPR strengthens and standardizes data protection laws across the EU. It gives individuals more control over their personal data and its use. The GDPR applies to various data processing activities, including collecting, storing, using, and sharing personal data (Li, Yu, & He, 2019). It sets out specific rights for individuals, such as the right to access, rectify, erase, or restrict the processing of their data, the right to object to the processing of their data, and the right to data portability (Osano, Inc., 2022). In our research, we are specifically interested in the impact of GDPR on online traffic. Those rights can affect web browsing behavior in several ways. For example, the right to be informed may cause individuals to be more cautious about providing personal information online. People may read privacy policies and terms of service before agreeing to share their data. The right to access their data can encourage individuals to review the information collected about them. They can then request that any incorrect or outdated information may be corrected or deleted. The right to erasure and the right to restrict processing may also lead individuals to be more selective about the information they share online. The right to data portability makes it easier for individuals to move their data between different service providers, which could increase competition in the market. Finally, the right to object may make individuals more selective about the types of communication they receive. They may be more likely to opt out of receiving marketing materials. Overall, the GDPR might impact web browsing behavior, but the extent may vary depending on individual circumstances and preferences.

Over the past decades, Europe has significantly improved its privacy regulations. As technology advances, policies will tighten to protect data and privacy rights. The GDPR, in particular, has established high standards for privacy protection worldwide. In the next section, we will compare this to the situation in the United States.

2.2 United States legislations

The United States is also concerned about data privacy but has yet to pass a national law. The EU and the US are dedicated to respecting each other's privacy rights. However, there have been economic tensions regarding the security dimensions. The GDPR draws attention to some of these disparities and can make it challenging for US companies to do business in Europe. The EU personal data protection and confidentiality as fundamental rights, while the US lacks a comprehensive law. Namely, the United States does not have a single policy but covers specific categories of data (Congress Research Service, 2020).

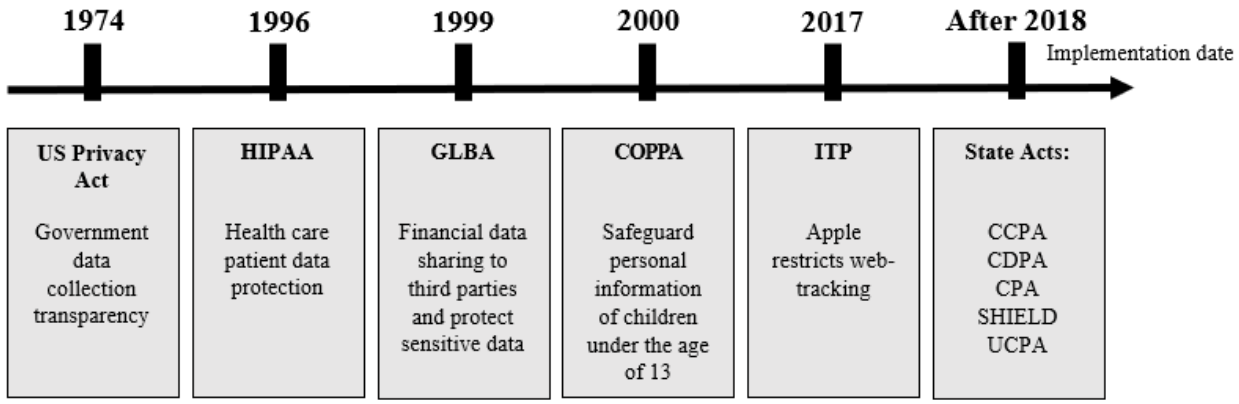

Figure 4: Timeline of privacy regulations in the United States (Osano, Inc., 2022)

In 1974 the 'US Privacy Act' was enacted. It was in response to concerns that using government databases would affect people's right to privacy. Any records kept on a person had to be made available upon request. The next significant shift in data privacy came in 1996. The 'Health Insurance Portability and Accountability Act' (HIPAA) outlines patient data protection and confidentiality. In 1999 a regulation came into place to secure financial information, the 'Gramm-Leach-Bliley Act' (GLBA). Financial institutions must disclose how they share customer information with third parties and protect sensitive data. Furthermore, the US also had to protect children with the rise of the internet. In 2000, the ' Children's Online Privacy Protection Act' (COPPA) was passed to safeguard the personal information of children under 13.

Furthermore, companies like Apple added privacy measures. In 2017, Apple introduced 'Intelligent Tracking Protection' (ITP). ITP was created to restrict the ability of businesses to track customers' online behavior when they go to other websites. In 2018, the GDPR came into effect, which also influenced some firms in the US. Companies that had a close relationship with the European Union also started to follow the guidelines. After GDPR, certain states started implementing stricter acts, such as 'California Consumer Privacy Act' (CCPA), 'Virginia's Consumer Data Protection Act' (CDPA), 'Colorado Privacy Act' (CPA), 'New York SHIELD Act' (SHIELD), 'Utah Consumer Privacy Act' (UCPA), 'Connecticut's Data Privacy Law' (CTDPA). As of May 2022, legislation is also in committee in Alaska, Louisiana, Massachusetts, Michigan, North Carolina, New Jersey, New York, Ohio, Pennsylvania, Rhode Island, and Vermont. In addition, the US is working on a national regulation, the 'American Data Privacy Protection Act' (ADPPA). However, it is still being determined whether it will be implemented and will take time (Osano, Inc., 2022).

In summary, while the EU and the US claim to prioritize the protection of personal data and respect for individual privacy, the EU has implemented a comprehensive data privacy law in the form of the GDPR. The United States, on the other hand, has regulated data privacy at a sectoral level and has yet to pass a national law addressing the issue. Despite this, various states in the US have implemented their data privacy laws, but only after the GDPR in 2018. Therefore the US acts as a relevant country for comparison to see a potential effect in Germany.

3 Theoretical foundations

3.1 Research questions

To define the scope of our research, we conducted a review of existing research on the GDPR. Through this evaluation, we identified three main themes that have been commonly addressed in previous research.

- The behavioral consequences of GDPR on consumers include studies that aim to answer how the regulation causes individuals to behave differently after GDPR. Behaviors analyzed are for example online traffic, information sharing or privacy awareness (Goldberg, Johnson, & Shriver, 2019; Congiu, Sabatino, & Sapi, 2022; Godinho de Matos & Adjerid, 2022).
- The choice architecture for consent elicitation focuses on cookie control; since the adoption of the GDPR, most websites in Europe display cookie consent forms (Bauer,

Bergstrøm, & Foss-Madsen, 2021; Sanchez-Rola et al., 2019; Utz, Degeling, Fahl, Schaub, & Holz, 2019).

- The impact of GDPR on market outcomes and firm strategy; for example, Johnson, Shriver, and Goldberg (2021) analyzed the vendors' market. They found that GDPR reduces online data sharing, but it also had the unintended consequence of increasing market concentration among web technology vendors. Peukert, Bechtold, Batikas, and Kretschmer (2022) documented short-run changes in the websites and web technology industry, focusing on Google.

For the first topic, there are still only limited research papers available, and we see great potential for further research. Therefore we complement the academic work primarily on behavioral consequences of the GDPR. Specifically, literature in the past has mainly investigated changes in online consumer behavior by looking at the unfiltered traffic data of websites. For example, Goldberg et al. (2019) used Adobe data that gives the total website visits without considering consumer traffic from specific countries. We, however, investigate changes in web browsing behavior only for Germans. We chose Germany because it is the most populous member of the European Union and thus subject to the GDPR. Germany is also one of the economically strongest countries in the EU and has pushed businesses and public administrations to put a focus on the GDPR.[2]

Studies on behavioral change due to GDPR typically assumed that all European websites immediately complied with the new regulations (Aridor, Che, & Salz, 2021; Schmitt, Miller, & Skiera, 2020). We question this assumption and look into compliance in more detail. We provide a descriptive overview of which websites comply with the regulation and when. We also examine whether there are any differences in the effects of regulation on online traffic for websites that comply versus those that do not comply, which leads us to the first of three research questions:

"How did the General Data Protection Regulation affect German web browsing behavior for websites that complied compared to websites that did not comply?"

Due to data limitations for our compliance sample, we also compare German traffic to American traffic. We use the United States as a country of comparison because it is outside the EU but shares similar values and culture. The comparison brings us to our second research question:

[2] https://www.datenschutz.de/die-datenschutz-grundverordnung-ds-gvo/

"How did the General Data Protection Regulation affect web browsing behavior in Germany compared to the United States?"

Because US and German consumers may differ in demographic characteristics, which could affect internet browsing behavior, we also study German traffic differences inside and outside the EU. Here we assume that websites outside the EU take longer to comply with the GDPR. Our third research question is:

"How did the General Data Protection Regulation affect web browsing behavior in Germany within the European Union compared to outside the European Union?"

We will briefly introduce studies about online privacy perceptions before we describe how the GDPR affected consumers and firms.

3.2 Privacy perceptions

Various studies have been conducted on the impact of websites' privacy handling on consumer attitudes and website usage behavior. The outcome of the results differs. Some claim that stricter privacy laws positively impact user behavior when sharing data. Others assert the opposite. Surveys show that privacy is a primary concern for citizens in the digital age.[3] Nevertheless, privacy notices, such as cookies, help to increase consumer trust. People feel more in control and thus believe that the websites are legitimate (Tang, Hu, & Smith, 2008). Trust is also influenced by the number of alternatives and choices for privacy controls; the more options available, the less concerned people are (K. Martin, 2015). Several studies have found that this positively influences people's willingness to share data with the website (K. Martin, 2015) or purchase more products (K. D. Martin, Borah, & Palmatier, 2017; Tsai, Egelman, Cranor, & Acquisti, 2011). Other studies show that if companies give their customers more control over how their personal information is used, they are more likely to limit its use (Mandić, 2009; Kim, Barasz, & John, 2019). However, one can often observe that individuals rate their privacy concerns higher than they might seem when looking at their online behavior, a phenomenon known as the 'privacy paradox' (Kokolakis, 2017). Our study investigates the relationship between this privacy paradox and online traffic, specifically whether people are more likely to increase or decrease online browsing behavior.

[3] https://www.sas.com/content/dam/SAS/documents/marketing-whitepapers-ebooks/sas-whitepapers/en/data-privacy-110027.pdf

3.3 Behavioral consequences

We continue with a literature review that summarizes what we currently know about the behavioral effects of the GDPR. Based on the literature, the regulation has decreased traffic and website engagement. For example, Goldberg et al. have written two papers on this topic. Their first paper primarily focuses on the economic aspects of the GDPR, while the second paper looks at the early impact on web traffic and e-commerce (Goldberg et al., 2019; Goldberg, Johnson, & Shriver, n.d.). They have discovered that recorded total pageviews on Adobe and revenues declined by roughly 10-12%. Aridor et al. (2021) find similar results; the authors detect a 12.5% decrease in observed consumers on travel websites, yet the remaining consumers are more persistently recognizable. Moreover, Schmitt et al. (2020) investigated the short- and long-term impacts. They identify that the GDPR negatively affects user quantity (e.g., number of total visits) and usage intensity (e.g., page impressions per visit). About two-thirds of websites continue to be negatively affected by the GDPR in the long term. Congiu et al. (2022) also find a long-term overall traffic reduction of 15% and a measurable decline in website engagement.

Some relate this decrease in traffic to the rise in costs for gathering customer data. Privacy restrictions make it more expensive for companies to match customer preferences (Goldberg et al., 2019, n.d.). Others say the negative impact is due to limitations of online tracking by changing the website content offered to visitors. The findings imply that websites experienced a decrease in third-party cookies and tracking replies. However, the GDPR does not hurt the volume of content that websites can publish or the average level of social media engagement and interaction with such content (Lefrere, Warberg, Cheyre, Marotta, & Acquisti, 2020).

We contribute to the existing literature by focusing on a specific subgroup, namely German traffic behavior and how they respond to the GDPR; most studies have been conducted on a broader level, looking at the overall traffic of websites. In contrast to the findings of these previous studies, our research found that there have been no changes in German traffic since the GDPR went into effect.

One area of focus in the study of the GDPR is the choice architecture for consent elicitation. It refers to how websites present data collection and use options to users and seek their consent. Utz et al. (2019) investigated the influence of notice position, type of choice, and content framing on consent. They find that seemingly small implementation choices can significantly impact how individuals interact with cookies. Their research demonstrated the importance of regulations that allow users to make informed decisions about their data. Bauer et al. (2021) also showed that proper choice architecture could increase consent rates by up to 17 percentage points.

Sanchez-Rola et al. (2019) analyzed tracking information and found that the GDPR impacted website activity globally. They discovered that third-party website opt-out services restricted tracking. Some cookies could, however, bypass these services and continue following user activity. Their research highlighted the challenges of obtaining informed consent for data collection and the necessity of having clear guidelines. Godinho de Matos and Adjerid (2022) researched the impact of the GDPR on consumer consent using data from a large telecom company. They found that opt-in rates increased following the implementation of the GDPR. They observed that consumers tend to be selective about granting permission to use their personal information.

Our study focuses on the impact of cookie pop-ups on online traffic rather than specific types of cookie consent notices. Previous research on traffic has generally assumed that all European websites complied with the GDPR (Aridor et al., 2021; Schmitt et al., 2020), but we aimed to investigate the actual effect of compliance. For this, we manually identified websites that implemented cookie pop-ups after the GDPR introduction.

The impact of the GDPR has mainly shown a decrease in online traffic behavior. Factors influencing the impact include website type, consumer trust, control level, and the costs to companies of collecting and using personal data. Further research is needed to fully understand the short and long-term effects of the GDPR on online traffic behavior. These theoretical foundations provide a strong foundation for addressing our research questions and conducting our analysis.

4 Data

4.1 Data description

We will use data from three different data sources:

- Website traffic collected by The Nielsen Company,
- Online harm engagement from OpSec, White Bullet, Virus Total, and Domain tools,
- Cookie compliance of websites manually created by us.

4.1.1 Website traffic

Two dataset's including website traffic from The Nielsen Company, serve as the foundation for our primary analysis. The first contains a summary of German web browsing behavior to the 1000 most popular German websites, while the second includes American online usage to the 1000 most popular US websites. These domains come from various industries, including

retail, media, and news. A list of the most popular websites can be found in the Appendix. The data for this study covers the period from May 2017 to November 2018. However, We will only consider January to November 2018 in our regression analysis, but we will use the entire period for the synthetic control method. The popularity list is accurate as of May 2017; the number of websites included in our sample decreases in the following months as websites that no longer qualify for the list is removed. We will look at only the domains for which we have data for all periods. We examine the audience, pageviews per person, sessions per person, and hours per person for each website aggregated monthly. We present the definitions of the before-mentioned metrics in Table 1. They are taken from Google Analytics[4] as we did not get a glossary from The Nielsen Company. The data scientists of The Nielsen Company create smaller groups that mimic the demographics and estimate the behavior of the larger overall population[5]. To ensure high data accuracy, we will filter our dataset for websites with a sample size greater than 60.

The GDPR applies to all EU citizens, meaning any website that wants to access European personal data should comply with GDPR standards. However, we assume that websites outside the EU will take longer to implement changes. To ensure that we capture the effect, we examine the country code top-level domain; for German traffic inside the EU, we only consider European codes {.at,.de,.eu,.uk}, and for the United States, we consider {.gov,.la,.us}, which leaves us with 485 domains for Germany and 32 for the US. We also compare our German traffic inside the EU with the German traffic outside the EU, which includes 275 domains.

Table 1: Definitions of variables of interest

Variable	**Definition**
Audience	The total number of visitors to a website in a month
Pageviews per person	Average number of times a person views the website in a month. An additional pageview is also recorded if a user reloads after reaching the page.
Sessions per person	Average number of sessions initiated by users to the website within a month. Inactivity for 30 minutes or more stops a session.
Hours per person	Average amount of time a visitor spends on a website page. The data is presented in hours (0.013 hours = 46,8 seconds)

[4] https://support.google.cm/analytics

[5] https://www.The Nielsen Company.com/about-us/The Nielsen Company-panels/

4.1.2 Online harms

We investigate GDPR compliance for websites that engage in illegal or harmful activities, such as copyright infringement or phishing. For various reasons, harmful websites may not follow the GDPR. They might need to be made aware of their responsibilities or intentionally ignore them. Additionally, some websites are designed to evade detection or avoid being tracked by law enforcement. They are making it hard to ensure compliance with privacy regulations.

To determine whether the domains in our sample engage in online harm, we turned to multiple information sources. We use data from OpSec, White Bullet, Virus Total, and Domain tools to identify websites involved in online harms such as copyright infringement, cyber scams, or malware. The information is from July 2020. See Table 2 for specific definitions and distributions. After cross-referencing the harm dataset with the traffic dataset from The Nielsen company, we were left with a subset of the websites. In our sample, the Copyright Infringement category accounts for 30.0% of the total, with four subcategories: CI, CI2, CI3, and UC. The Malicious category represents 9.4% of the total, with two subcategories: ML and SU. The Phishing category makes up 10.5%.

Table 2: Definitons of online harms and their website share

Category	Subcategory	Definition	Share
Copyright Infringement (Total Share 30.0%)	CI	Content infringing IP rights or copyright with at least one instance since 01.2019	22.2%
	CI2	Content infringing IP rights or copyright in the top quartile since 01.2019	26.6%
	CI3	Content infringing IP rights or copyright with at least medium risk since 01.2019	5.2%
	UC	Presence of unauthorized content (patent, trademark, counterfeit, illegal import)	12.0%
Malicious (Total Share 9.4%)	ML	Number of VirusTotal partner trackers that identify a website as malicious	8.2%
	SU	Number of VirusTotal partner trackers that identify website as suspicious	1.1%
Phishing	PS	Presence of phishing or BEC scams	10.5%

4.1.3 Cookie compliance

Cookies are small pieces of data stored on the computer or mobile device when visiting a website. Cookies are an essential tool that can provide businesses with a wealth of information about their users' online behavior. Websites use them to store information about the visit, such as preferences and history. Some cookies are essential for the operation of a website, while others are used to personalize the user experience or to track user behavior. Nowadays, one can accept or decline cookies by modifying browser settings. Given the amount of information that cookies can store, they may qualify as personal data and fall under the purview of the GDPR. Websites must obtain users' consent before using any cookies except those that are strictly necessary.[6]

We created a dataset using the Wayback Machine - Internet Archive[7] that contains information on whether a website complies with the cookie regulation. The Wayback Machine is a collection of snapshots of websites available on the internet as they appeared on specific dates in the past. We manually checked 466 websites to see if they had a cookie notice for the dates: June 13$^{\text{th}}$, 2018, and November 30$^{\text{th}}$, 2018. If no snapshot was available on that date, we used the observation that was closest to it. We were not able to access 43 websites. We only considered websites for which information was available regarding whether they entail harmful content or not (see Subsection 4.1.2).

In Table 3, we show how many websites complied in our sample. We see for June 2018 that 49% of websites with German traffic had a cookie pop-up. Compliance increased to 57% by November 2018. Even though US traffic technically does not require cookie notices, we see 19% and 23% for June and November, respectively. We only consider German traffic for our analysis.

Table 3: Cookie compliance per country

	June 2018	November 2018
Germany Traffic		
Compliance	151	175
Non-Compliance	157	133
United States Traffic		
Compliance	22	27
Non-Compliance	93	88

[6] https://gdpr.eu/cookies/

[7] https://archive.org/web/

4.2 Summary statistics

Table 4 presents summary statistics for Nielsen online traffic data. The table includes five samples: German traffic to websites that show cookie consent pop-ups, German traffic inside the EU, German traffic to websites that do not show cookie consent pop-ups, American traffic inside the US, and German traffic outside the EU. Each sample is divided into two periods: before and after implementing the GDPR. The "Observations" column indicates the number of data points in each sample. The "Websites" column indicates the number of websites included in each sample. The number of observations is larger in the after period, as we are considering a larger time period.The traffic metrics are the audience, pageviews per person, sessions per person, and hours per person.

First, we take a look at our treatment samples. The compliance dataset entails websites with a higher audience of 1,358,626 visitors per month before the GDPR and afterward 1,339,078 visitors per month, so we see a decrease of 19,548 visitors. On average, the websites got 11.737 pageviews per person compared to after GDPR 11.554 pages per person. We again observe a potential decrease due to the GDPR. In contrast, looking at sessions per person, we see an increase of 0.014 in the average value, from 2.670 to 2.684. There is no movement in time; the average stays at 0.10 hours per person for both periods.

For German traffic inside the EU, the audience mean was 1,154,424 visitors per month before the GDPR compared to 1,126,393 afterward, which indicates a decrease of about 28 thousand visitors. For pageviews per person, the mean in Germany is 15.313 pages before GDPR and 15.034 afterward. The delta of -0.279 indicates a downward adjustment over time. The median number of page views per person is almost half the size due to a few outliers with significant views. The median pageviews of 8.220 before GDPR decreased by 0.334 afterward. We are moving on to the number of sessions per person. On average, in Germany, one user has 2.999 sessions on a website before GDPR and 2.982 afterward. This slight downward shift is only visible in the mean; for median sessions, we see a tiny increase of 0.06. Next, we discuss the statistics of the number of hours people usually spend on each website. For Germany, we see a minimal decrease from 0.013 to 0.012 hours.

For the Control Samples, non-compliance has an audience of 1,632,792 visitors per month before and 1,656,359 visitors afterward. The increase could point towards the possibility that consumers switched from websites that comply to websites that do not comply. Some people may prefer websites that do not comply with the GDPR because they do not require users to accept terms and conditions or provide personal information. It can be more convenient for users, especially if they only visit the website for a short time. Pageviews per person are 14.313 before, 15.613 after, and sessions are 3.733 and 3.758, so both pages and sessions

increase. The average hours per person decreased slightly from 0.019 to 0.018.

American traffic inside the US has a higher average audience, 1,718,018 and 1,366,767 before and after GDPR. Both control and treatment show a downward trend, which needs further investigation to determine if the GDPR influences this. The people viewed, on average, a website 20.784 times (median 18.100) in the month before the data privacy laws were implemented in Europe. Afterward, the average was 22.667 (median 19.664). This increase is the opposite trend compared to Germany. In the United States, the mean number of sessions per person is 3.571 before GDPR and 3.673 afterward, and the average hours per person do not change in the US.

Finally, German traffic outside the EU shows an average audience increase per month, 1,309,882 before and 1,333,205 after the regulation. This boost could be due to a shift from European to Non-European websites. Furthermore, there are, on average, 14.445 and 14.635 pages per person, so we also see an increase of 0.19. The metric average sessions per person, however, decline from 3.395 to 3.364, whereas hours per person again do not change.

Table 4: Summary statistics for Nielson Online Traffic

Samples	Period	Observations	Websites	Audience	Pages pp.	Sessions pp.	Hours pp.
Treatment Samples							
German traffic - Compliance	Before GDPR	540	135	1,358,626	11.737	2.670	0.010
	After GDPR	945	135	1,339,078	11.554	2.684	0.010
German traffic inside the EU	Before GDPR	1,940	485	1,154,424	15.313	2.999	0.013
	After GDPR	3,395	485	1,126,393	15.034	2.982	0.012
Control Samples							
German traffic - Non-Compliance	Before GDPR	544	136	1,632,792	14.313	3.733	0.019
	After GDPR	952	136	1,656,359	15.613	3.758	0.018
American traffic inside the US	Before GDPR	128	32	1,718,918	20.784	3.571	0.016
	After GDPR	224	32	1,366,767	22.667	3.673	0.016
German traffic outside the EU	Before GDPR	1,028	275	1,309,882	14.445	3.395	0.017
	After GDPR	1,799	275	1,333,205	14.635	3.364	0.017

In conclusion, the summary statistics in Table 4 show that the implementation of the GDPR had a varied impact on online traffic in the different samples. Further investigation is needed to determine the full impact of the GDPR on online traffic and to understand the reasons behind the observed trends.

5 Difference-in-differences approach

This section describes the method employed to explain and estimate the impact of GDPR on online traffic behavior.

5.1 Groups of comparison

To determine the effect of the GDPR on German online traffic user behavior, we compare three usage patterns before and after the GDPR was implemented:

- **German traffic to compliant websites vs. non-compliant websites**: To determine the actual impact, we compare the change in German online traffic for websites that complied (treatment) - had a cookie pop-up - to websites that did not (control).
- **German traffic vs. American traffic**: German traffic to European websites will serve as our treatment group because the websites belong to the European Union and, therefore, must comply with the GDPR. American traffic to US websites will act as the control group as the US is outside the GDPR region.
- **German traffic inside vs. outside the EU**: Many foreign websites did not comply with GDPR initially, even though they are required to when dealing with European personal data. We compare German traffic inside the EU (treatment) vs. German traffic outside the EU (control).

5.2 Methodology

The difference-in-differences (DID) approach is a widely used method for evaluating the impact of a policy change on a particular outcome (Congiu et al., 2022; Schmitt et al., 2020), which we will also utilize in our study. This method involves comparing the change in the outcome of interest in the treatment group (in our case, websites that are subject to GDPR) with the change in the outcome in a control group (in our case, websites that are not subject to GDPR) over the same period. DID allows us to control for any underlying trends or changes that may affect the outcome and isolate the effect of the treatment (Card & Krueger, 1994). Our general DiD specification is:

$$log(y_{i,t}) = \beta_0 + \beta_1 TG_i + \beta_2 GDPR_t * TG_i + \alpha_i + \mu_t + \epsilon_{i,t} \tag{1}$$

where $y_{i,t}$ is the variable of interest (audience, pages per person, sessions per person and time per person) for website i and time t in months. $GDPR_t$ is a dummy which refers to

whether the observation occurred after GDPR implementation in May 2018 (1) or before (0). Another indicator variable is TG_i which states if the web traffic comes from the treatment group (1) or the control group (0). Finally, α_i denotes website fixed effects, μ_t stands for the period fixed effects and $\epsilon_{i,t}$ is a mean-zero error term.

Our main coefficient of interest is β_2, representing the average causal impact of the GDPR on website traffic. For example, the estimated coefficient measures the average (percentage) difference in the audience for the treatment group and the control group caused by GDPR.

To ensure our control groups can be used appropriately, we need to identify if they follow the same pre-treatment pattern as the treatment group - we need to check for parallel trends. We test this assumption by including the interaction of the treatment group dummy variable with the ordinal variable *Period*, which represents the observation's time.

$$log(y_{i,t}) = \beta_0 + \beta_1 TG_i + \beta_2 GDPR_t * TG_i + \beta_3 Period_t * TG_i + \alpha_i + \mu_t + \epsilon_{i,t} \tag{2}$$

We accept different pre-trends for the treated and control group as long as they are linear. Therefore, in regression (3), all period-interactions before treatment (β_4, β_5, β_6) should be insignificant.

$$\begin{aligned} log(y_{i,t}) &= \beta_0 + \beta_1 TG_i + \beta_2 GDPR_t * TG_i + \beta_3 Period_t * TG_i + w_{i,t} + \alpha_i + \mu_t + \epsilon_{i,t} \\ where \quad w_{i,t} &= \beta_4 Feb2018_t * TG_i + \beta_5 Mar2018_t * TG_i + \beta_6 Apr2018_t * TG_i \end{aligned} \tag{3}$$

6 Results

6.1 German traffic to compliant websites vs. non-compliant websites

In the following section, we present the findings on the impact of GDPR compliance on German traffic for websites with and without cookie pop-ups. In Figure 5, we can see the traffic development over time. All four plots show similar and consistent trends for compliant and non-compliant websites, with some fluctuations over the months. Websites without consent notices have more visits and higher engagement than websites with cookie pop-ups. However, there was no noticeable change in any of the variables after the implementation of GDPR.

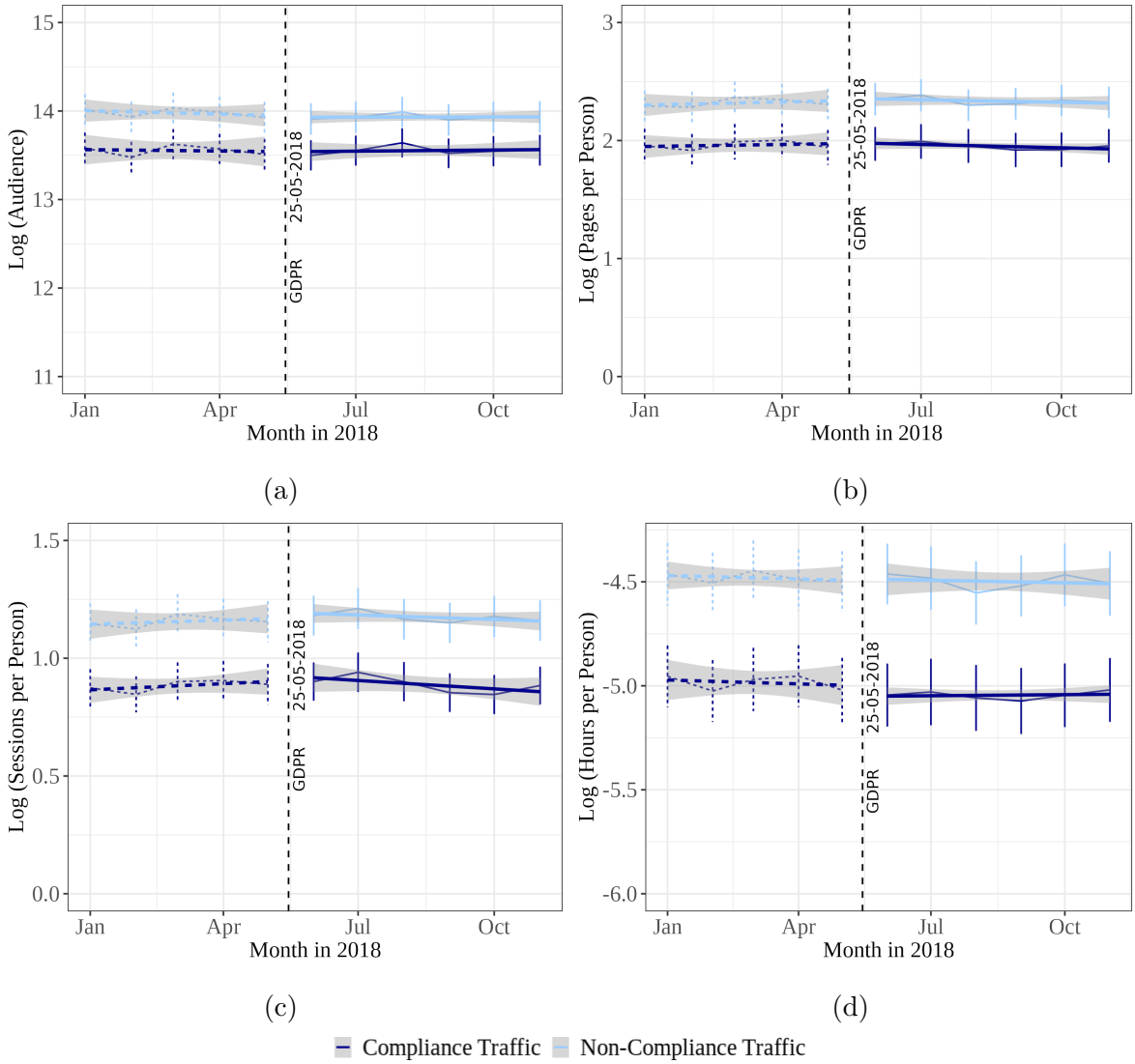

Figure 5: Time-series plots: German traffic to compliant websites vs. non-compliant websites

Furthermore, Table 5 contains the regressions investigating the impact of GDPR compliance on our four variables of interest. Since the data exhibits parallel trends, we only consider the $GDPR * Cookie$ coefficient. The coefficient for the audience was positive, whereas pageviews, sessions, and hours per person were negative, but none were statistically significant. Overall, the results of the regressions were consistent with the trends observed in the graphical analysis.

Regarding compliance, we also investigate whether harmful websites are less likely to comply with the GDPR. To do this, we conduct four logit regressions. We present the average partial effects (APR) in Table 6. The first and third regressions examine the likelihood of harmful websites displaying a cookie pop-up in June and November 2018, respectively. The

Table 5: DiD regressions: German traffic to compliant websites vs. non-compliant websites

	Log (Audience)	Log (Pageviews per Person)	Log (Sessions per Person)	Log (Hours per Person)
GDPR*Cookie	0.021	−0.040	−0.001	−0.027
	(0.027)	(0.029)	(0.015)	(0.034)
Website FE	Yes	Yes	Yes	Yes
Period FE	Yes	Yes	Yes	Yes
Observations	2,981	2,981	2,981	2,981
R^2	0.956	0.920	0.938	0.904
Adjusted R^2	0.952	0.911	0.931	0.894

Note: *p<0.1; **p<0.05; ***p<0.01
Cluster robust standard errors in (): Errors clustered by website

second and fourth regressions specifically focus on the impact of copyright infringement, malicious behavior, and phishing on compliance.

Our results show that harmful websites are 11.2% less likely to have a cookie notice in June, as indicated by the negative coefficient in regression (1). This effect becomes even more pronounced in November, with harmful websites being 21.6% less likely to display a cookie notice. These findings suggest that compliance with the GDPR increases over time, but harmful websites resist. In addition, we find that the observed effect is primarily driven by websites that violate copyright, while the other forms of harm studied are insignificant regarding their impact on compliance.

Table 6: Logit regressions: Online harms on GDPR compliance

	Cookie in June		Cookie in November	
	(1)	(2)	(3)	(4)
Harmful website	-0.112^{*}		-0.216^{***}	
	(0.058)		(0.056)	
Copyright Infringment		-0.151^{**}		-0.254^{***}
		(0.063)		(0.062)
Malicious		−0.017		−0.040
		(0.094)		(0.092)
Phishing		−0.072		−0.151
		(0.096)		(0.095)
Observations	308	308	308	308

Note: *p<0.1; **p<0.05; ***p<0.01

6.2 German traffic vs. American traffic

Next, we look at the overall impact of the GDPR on German traffic within the EU compared to American traffic within the US.

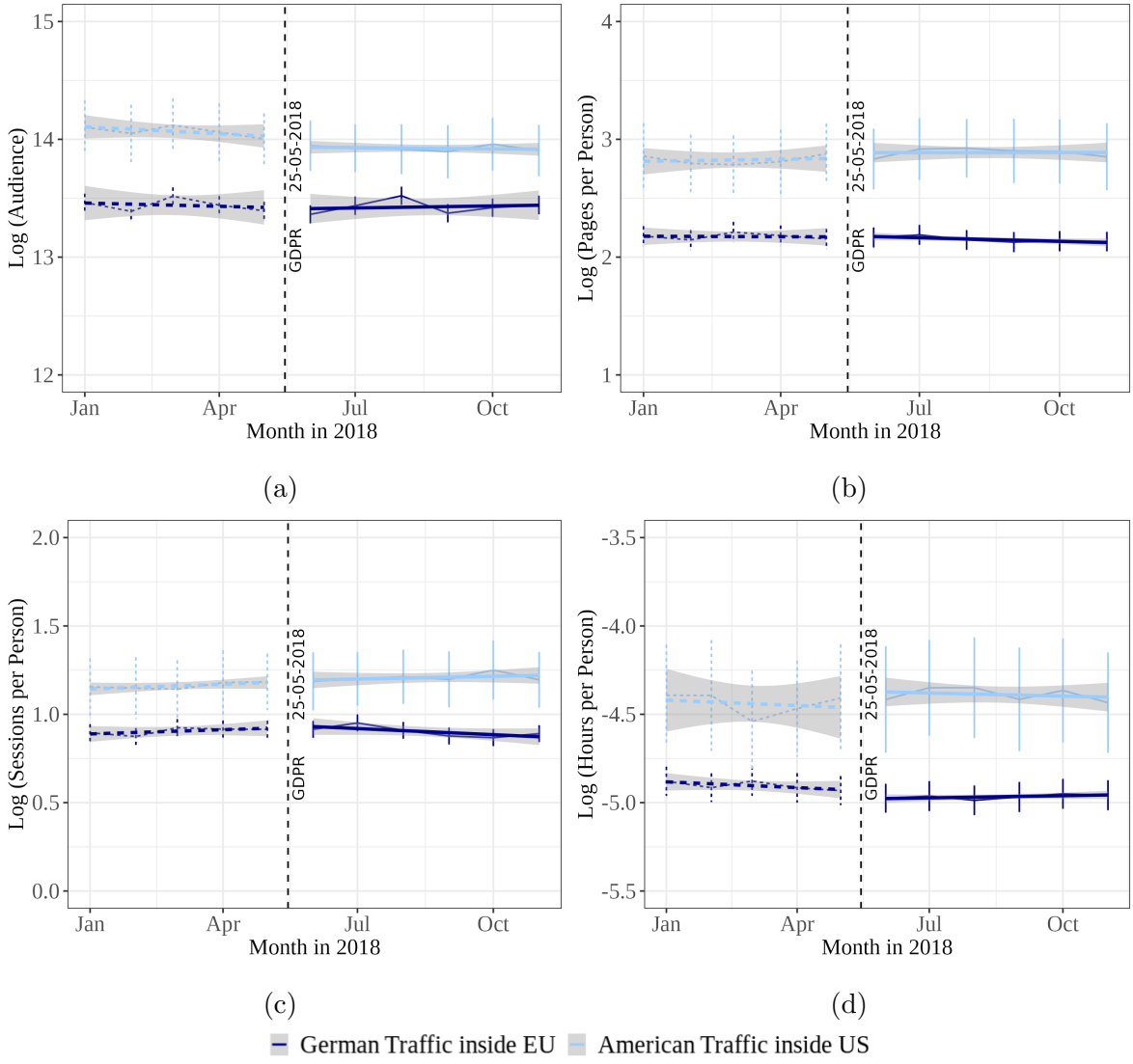

Figure 6: Time-series plots: German traffic inside the EU vs. American traffic inside the US

In Figure 6, plot (a) shows that the general audience in the United States is slightly declining before the GDPR and stays relatively constant afterward. In Germany, there is more volatility, though the overall trend stays around the same value. Since the GDPR was implemented, we have not noticed any significant changes in audience trends. Graph (b) depicts the evolution of pageviews per person, with the US showing an increasing trend and Germany showing a decreasing trend. Again, there is no noticeable change in trend following GDPR. For sessions per person, we see a relatively flat line for Germany and an increasing

line for the US in the plot (c). Graph (d) illustrates fluctuations in the amount of time spent per person on a website in the United States. In contrast, in Germany, there was a decrease in time spent in the months preceding the implementation of the GDPR, but a slight increase was observed afterward. In conclusion, our graphical analysis of website traffic data for Germany and the United States revealed some differences; we did not observe any significant changes in these trends following the implementation of the GDPR. The time trends may violate the parallel assumptions for each variable.

The regression results are presented in Table 7, where regressions (1)-(3) are regarding the audience. The coefficients of $Period * Germany$ in regressions (2) and (3) are positive and significant. They indicate a violation of the parallel trend assumption between Germany and the US. The trend is linear, as the pre-period interactions are all insignificant. When considering this linear trend, we do not find a significant effect of the GDPR on the German audience compared to the US audience. Next, we examine page views per person. Regressions (5) and (6) show that the parallel assumption holds; the coefficients for $Period * Germany$ are insignificant, and the pre-period interactions are not significantly different from 0 at a 5% level. Regression (4) shows that the GDPR significantly negatively affects pageviews per person for Germans visiting EU websites, with a 9.8% decrease. Regressions (5) and (6) also show an adverse effect but are not statistically significant.

We observe a linear difference in the trend for sessions per person in regression (8) and (9). German users reduce the average number of sessions per person more than users in the United States. However, the GDPR has no significant effect on sessions per person in Germany compared to the US, according to all three regressions. Finally, we find a parallel trend for hours per person and a significant negative effect of GDPR on Germany. Regressions (10) and (11) suggest a significant decrease of 11.48% and 13.84%, respectively, at the 5% level.

Summarizing, the regression analysis shows a slight decrease in pageviews and time per person but not in audience or pageviews per person.

Table 7: DiD Regressions: German traffic inside the EU vs American traffic inside the United States

	Log (Audience)			Log (Pageviews per Person)			Log (Sessions per Person)			Log (Hours per Person)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Period*Germany		0.017***	0.018**		−0.003	−0.008		−0.012**	−0.014**		0.005	0.0003
		(0.006)	(0.007)		(0.010)	(0.011)		(0.005)	(0.006)		(0.013)	(0.015)
February*Germany			−0.039			0.036			0.0001			−0.036
			(0.034)			(0.063)			(0.034)			(0.092)
March*Germany			0.001			0.122*			0.071			0.149*
			(0.032)			(0.073)			(0.044)			(0.090)
April*Germany			−0.033			0.076			0.039			0.037
			(0.045)			(0.066)			(0.052)			(0.098)
GDPR*Germany	0.117***	0.025	0.002	−0.104**	−0.086	−0.001	−0.046	0.017	0.060	−0.122***	−0.149**	−0.086
	(0.039)	(0.044)	(0.057)	(0.042)	(0.058)	(0.092)	(0.029)	(0.031)	(0.057)	(0.043)	(0.062)	(0.115)
Website FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Period FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,687	5,687	5,687	5,687	5,687	5,687	5,687	5,687	5,687	5,687	5,687	5,687
R^2	0.941	0.941	0.941	0.916	0.916	0.916	0.918	0.918	0.918	0.881	0.881	0.881
Adjusted R^2	0.934	0.935	0.934	0.907	0.907	0.907	0.910	0.910	0.910	0.869	0.869	0.869

Note: *p<0.1; **p<0.05; ***p<0.01

Cluster robust standard errors in (): Errors clustered by website

6.3 German traffic inside vs. outside the EU

We are moving on to our third usage pattern comparison - German traffic within and outside the EU. The time series plots are shown in Figure 7.

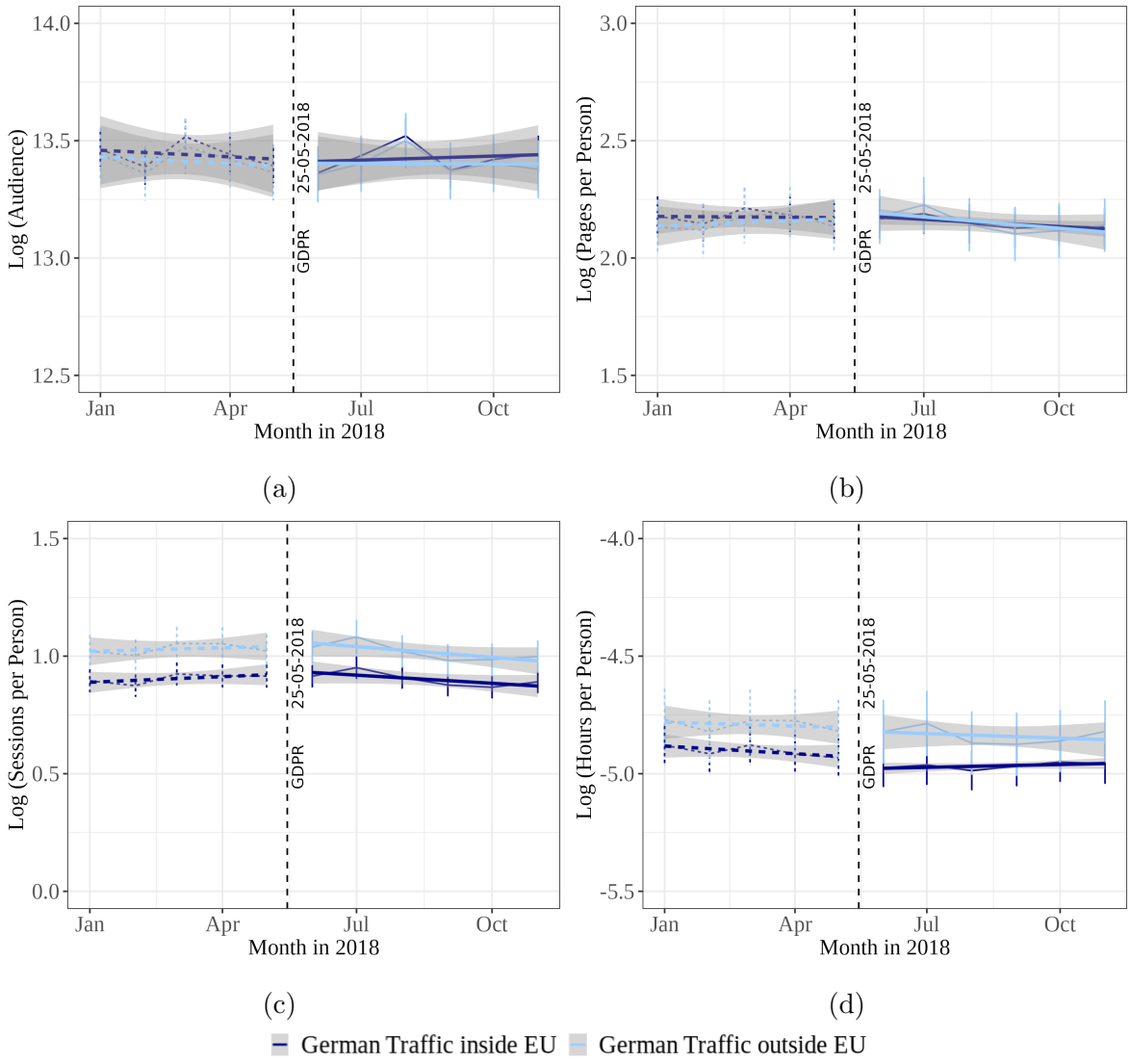

Figure 7: Time-series plots: German traffic inside vs. German traffic outside the EU

All four plots show a similar trend for traffic both within and outside the EU. Additionally, all four graphs exhibit fluctuations. It is worth noting that the audience and pages per person are higher within the EU, while the websites outside the EU have more sessions and time spent. The implementation of the GDPR has not caused significant changes in any of the variables except for minor trend shifts.

In Table 8, we present our regressions; we again only focus on the coefficient for $GDPR*InsideEU$, as the data for all of the regressions displayed a parallel trend. Utilizing the DiD

approach, we find that the implementation of the GDPR had no significant effect on any of the metrics analyzed. The coefficients for audience, pageviews per person, and hours were negative, sessions was positive, but none were statistically significant. Overall, the results of the regressions were consistent with the trends observed in the graphical analysis.

Table 8: DiD regressions: German traffic inside vs German traffic outside the EU

	Log (Audience)	Log (Pageviews per Person)	Log (Sessions per Person)	Log (Hours per Person)
GDPR*Inside EU	−0.004	−0.023	0.016	−0.013
	(0.016)	(0.021)	(0.012)	(0.023)
Website FE	Yes	Yes	Yes	Yes
Period FE	Yes	Yes	Yes	Yes
Observations	8,162	8,162	8,162	8,162
R^2	0.946	0.912	0.919	0.888
Adjusted R^2	0.940	0.903	0.911	0.877

Note: *p<0.1; **p<0.05; ***p<0.01
Cluster robust standard errors in (): Errors clustered by website

6.4 Robustness Checks

To enhance the reliability of our analysis, we also employed multiple robustness checks. First, to improve the compliance results, we rerun the DID regression for German traffic to compliant websites vs. non-compliant websites. We looked at the websites with a cookie pop-up in November instead of June. It made sure to not include websites that complied during the observation period in our control sample. Second, we implemented a synthetic control method (SCM). This technique involves constructing a weighted combination of websites used as controls to which the treatment group is compared. SCM helps to improve the similarity between the control and treatment conditions (Xu, 2017). Further details on the methodology and results can be found in the Appendix. Our results obtained are consistent with the findings from our difference-in-difference analysis.

7 Conclusion

The paper finishes with a discussion of the essential findings and summarizes the results to address the research questions posed at the investigation's outset. It is important to note that the current study has certain limitations that should be considered. From the results, we draw other ideas for future research.

7.1 Summary and research outcome

This paper studies the behavioral consequences of the GDPR, investigating the effect on German online traffic. Although the GDPR is one of the strictest data privacy laws worldwide, we did not see significant changes in web browsing behavior. Our analysis is crucial as it contrasts previous studies that suggested adverse effects (Goldberg et al., n.d.; Congiu et al., 2022). The available research was performed on a general level using website traffic from multiple countries. We, however, focused on the changes only in German behavior.

To answer our first research question, we collected data on whether websites had cookie consent pop-ups. The adoption of GDPR compliance measures happened gradually because the GDPR is complex and requires significant changes to data collection and usage practices, which may require new technology and specialized staff. We observed that websites with harmful content, such as those that engage in copyright infringement, were less likely to comply with the GDPR. Non-compliance may happen because they are more focused on profit, willing to take legal risks, and have lower visibility. This research may help politicians identify gaps in online privacy laws and promote compliance in the future. To evaluate the impact of the GDPR, we employed a difference-in-differences approach. We did not find any significant differences between compliant websites and those not.

To answer the second research question, we contrasted German and US behavior. Our analysis revealed no significant change in Germany's average number of audience and sessions per person due to the GDPR. However, we did observe a decrease in pageviews and hours per person after the regulation. This downward trend could be attributed to the GDPR. However, it could also be influenced by other factors, such as the types of websites and industries in the two countries, consumer behavior, or marketing strategies.

For our third research question, we compared German consumers accessing websites within and outside of the EU and found no statistically significant effect of the GDPR.

The reason why we see no effect might be because websites in Germany were already compliant with similar privacy laws before the implementation of GDPR. Hence, the new regulations had minimal impact on their traffic. Additionally, German consumers are cautious about sharing their data online and might not be affected by the new regulations. Other factors that contributed to the lack of changes in German online traffic are the timing of marketing strategies that are likely to attract and retain visitors.

In conclusion, the results suggest that GDPR does not impact traffic in Germany. It may encourage websites to prioritize and improve their online privacy practices without worrying about lower engagement. The findings also provide valuable insight into the potential effects of privacy laws in other countries.

7.2 Limitations and future research suggestions

When we attempt to investigate the impact of GDPR, we need to consider some biases and limitations. The following paragraphs list these restraints and present various extensions for future research.

Our study is based on a limited dataset consisting of a selection of the most popular websites in Germany and the United States. Our results may not represent low-traffic websites or websites in other countries. It would be interesting to investigate the impact of these factors on traffic in multiple countries, as cultural differences may influence the extent to which people are sensitive to privacy issues. Additionally, our analysis only covers the period from January to November 2018, so the long-term effects of the factors we examined cannot be inferred from our findings. Further research is needed to confirm the validity of our results over the long term.

Additional analysis could provide insight into the factors that impact website traffic and engagement. This study did not consider other variables that may have affected the results. For example, changes in website content or marketing strategies drive traffic. It would also be valuable to study direct GDPR impacts per consumer groups, such as age groups or occupations. This information could be helpful for governments as they strive to understand individual needs and tailor privacy regulations accordingly.

Furthermore, limited research has been conducted on GDPR compliance. We looked at cookie pop-ups, but many other criteria could be considered concerning website traffic. Measures could be data minimization, accuracy, and retention. Future research could review the connection between these factors and harm, as we have only just begun to explore the risks associated with compliance.

In the past, databases like 'Whois' included information about domain name owners and their contact information. Web owners could purchase to conceal this information; the GDPR enabled it for free. Websites owners might keep this information private to reduce spam, prevent identity theft, avoid controversy and hide their physical location. Our analysis found a significant increase in websites withholding their identity after the GDPR. Due to the limited data, we could not perform a meaningful analysis of the impact of the GDPR on internet transparency.

Finally, our comprehensive analysis demonstrates that there has been no significant impact on online behavior in Germany. Although our study is based on a small data sample, it is a solid foundation for future research.

References

Aridor, G., Che, Y.-K., & Salz, T. (2021). The effect of privacy regulation on the data industry: Empirical evidence from gdpr. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3563968

Bauer, J. M., Bergstrøm, R., & Foss-Madsen, R. (2021). Are you sure, you want a cookie?– the effects of choice architecture on users' decisions about sharing private online data. *Computers in Human Behavior*, *120*. doi: 10.1016/j.chb.2021.106729

Card, D., & Krueger, A. B. (1994). Minimum wages and employment: A case study of the fast food industry in new jersey and pennsylvania. *American Economic Review*, *84*. doi: 10.1257/aer.90.5.1397

Congiu, R., Sabatino, L., & Sapi, G. (2022). The impact of privacy regulation on web traffic: Evidence from the gdpr. *SSRN Electronic Journal*. doi: 10.2139/ssrn.4025033

Congress Research Service. (2020). Eu data protection rules and u.s. implications. *https://sgp.fas.org/crs/row/IF10896.pdf*.

Council of Europe. (1950). European convention on human rights. *https://www.coe.int-/en/web/human-rights-convention/the-convention-in-1950*.

European Parliament and Council. (1995). Data protection directive. *https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A31995L0046*.

European Parliament and Council. (2002). Privacy and electronic communications directive. *https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32002R0178*.

European Parliament and Council of the European Union. (2018). General data protection directive. *https://gdpr.eu/*.

Godinho de Matos, M., & Adjerid, I. (2022). Consumer consent and firm targeting after gdpr: The case of a large telecom provider. *Management Science*, *68*. doi: 10.1287/mnsc.2021.4054

Goldberg, S., Johnson, G., & Shriver, S. (n.d.). Regulating privacy online: An economic evaluation of the gdpr. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3421731

Goldberg, S., Johnson, G., & Shriver, S. (2019). Regulating privacy online: The early impact of the gdpr on european web traffic & e-commerce outcomes. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3421731

Johnson, G., Shriver, S., & Goldberg, S. (2021). Privacy & market concentration: Intended & unintended consequences of the gdpr. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3477686

Kim, T., Barasz, K., & John, L. K. (2019). Why am i seeing this ad? the effect of ad transparency on ad effectiveness. *Journal of Consumer Research*, *45*, 906–932. doi:

10.1093/jcr/ucy039

Kokolakis, S. (2017). Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers and security*, *64*, 122–134. doi: 10.1016/j.cose.2015.07.002

Lefrere, V., Warberg, L., Cheyre, C., Marotta, V., & Acquisti, A. (2020). The impact of the gdpr on content providers. *The 2020 Workshop on the Economics of Information Security*.

Li, H., Yu, L., & He, W. (2019). The impact of gdpr on global technology development. *Journal of Global Information Technology Management*, *22*, 1–6. doi: 10.1080/1097198X.2019.1569186

Mandić, M. (2009). Privacy and security in e-commerce. *Market-Tržište*, *21*(2), 247–260.

Martin, K. (2015). Privacy notices as tabula rasa: An empirical investigation into how complying with a privacy notice is related to meeting privacy expectations online. *Journal of Public Policy & Marketing*, *34*, 210–227. doi: 10.1509/jppm.14.139

Martin, K. D., Borah, A., & Palmatier, R. W. (2017). Data privacy: Effects on customer and firm performance. *Journal of Marketing*, *81*, 36–58. doi: 10.1509/jm.15.0497

OECD. (1980). Oecd guidelines on the protection of privacy and transborder flows of personal data. *https://www.oecd.org/digital/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm*.

Osano, Inc. (2022, Sep). Data privacy laws: What you need to know in 2022: Articles. *https://www.osano.com/articles/data-privacy-laws::text=Despite%20numerous 20proposals%20over%20the,health%20information%2C%20credit%20information%2C%20 financial*.

Peukert, C., Bechtold, S., Batikas, M., & Kretschmer, T. (2022). Regulatory spillovers and data governance: Evidence from the gdpr. *Marketing Science*, *41*. doi: 10.1287/mksc.2021.1339

Sanchez-Rola, I., Dell'Amico, M., Kotzias, P., Balzarotti, D., Bilge, L., Vervier, P.-A., & Santos, I. (2019). Can i opt out yet? gdpr and the global illusion of cookie control. *AsiaCCS 2019 - Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security*, 340–351. doi: 10.1145/3321705.3329806

Schmitt, J., Miller, K. M., & Skiera, B. (2020). The impact of privacy laws on online user behavior. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3774110

Tang, Z., Hu, Y., & Smith, M. D. (2008). Gaining trust through online privacy protection: Self-regulation, mandatory standards, or caveat emptor. *Journal of Management Information Systems*, *24*, 153–173. doi: 10.2753/MIS0742-1222240406

Tsai, J. Y., Egelman, S., Cranor, L., & Acquisti, A. (2011). The effect of online privacy

information on purchasing behavior: An experimental study. *Information systems research*, *22*, 254–268. doi: 10.1287/isre.1090.0260

Utz, C., Degeling, M., Fahl, S., Schaub, F., & Holz, T. (2019). (un) informed consent: Studying gdpr consent notices in the field. *Proceedings of the 2019 acm sigsac conference on computer and communications security*, 973–990. doi: 10.1145/3319535.3354212

Xu, Y. (2017). Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis*, *25*, 57–76. doi: 10.1017/pan.2016.2

A Appendix

A.1 Extract of most popular websites

Table 9 shows the top five websites for each of the samples we use in our analysis in terms of average audience size. We take this information from The Nielsen Company as described in the Data section.

Table 9: Top five websites per sample in terms of audience

	Treatment Samples		Control Samples		
Rank	German traffic inside the EU	Compliance June 2018	American traffic inside the US	German traffic outside the EU	Non-Compliance June 2018
1.	google.de	amazon.de	irs.gov	google.com	google.de
2.	amazon.de	facebook.com	ca.gov	youtube.com	google.com
3.	ebay.de	ebay.de	ssa.gov	facebook.com	youtube.com
4.	web.de	paypal.com	nih.gov	paypal.com	bing.com
5.	ebay-kleinanzeigen.de	chip.de	ed.gov	wikipedia.org	t-online.de

A.2 Robustness: German traffic to compliant websites vs. non-compliant websites

To improve the compliance results, we rerun the difference-in-difference regression for German traffic to compliant websites vs. non-compliant websites. In this study, we focused on websites that displayed a cookie pop-up in November rather than June. This ensured to excluded any websites that were compliant during the observation period from our control sample. By doing this, we were able to assess the impact of the cookie pop-up on compliance and control for any confounding factors.

The results are presented in Table 10. As in the previous analysis, we don't find a statistically significant effect for any of the four variables of interest.

Table 10: DiD regressions: German traffic to compliant websites vs. non-compliant websites in November

	Log(Audience)	Log(Pageviews per Person)	Log(Sessions per Person)	Log(Hours per Person)
GDPR*Cookie	0.015	−0.038	0.002	−0.007
	(0.030)	(0.032)	(0.016)	(0.035)
Website FE	Yes	Yes	Yes	Yes
Period FE	Yes	Yes	Yes	Yes
Observations	2,959	2,959	2,959	2,959
R^2	0.956	0.919	0.937	0.904
Adjusted R^2	0.952	0.911	0.931	0.892

Note: *p<0.1; **p<0.05; ***p<0.01
Cluster robust standard errors in (): Errors clustered by website

A.3 Synthetic control group method

A.3.1 Methodology

Additionally, we utilize the synthetic control method (SCM) to increase the robustness of our analysis. This method involves creating a synthetic control group based on the average of a group of potential control units similar to the treatment group in terms of pre-treatment characteristics. This helps to improve the similarity between the control and treatment conditions (Xu, 2017). The synthetic control group is then compared to the treatment group to assess the impact of the GDPR. For our traffic pattern comparisons, it is important to consider the potential for spillover effects, as websites in the control group may choose to comply voluntarily with the GDPR. To address this concern, we also perform an SCM regression.

Y_it is our variable of interest with website i at time t (We observe data for T periods). We have two website sets, treatment T and control C with a total number of websites $N = N_{tr} + N_{co}$, where N_{tr} and N_{co} are the numbers of treated and control domains, respectively. The EU implemented the GDPR policy at period $T_0 + 1$, so we denote the number of pre-treatment periods for website i as $T_{0,i}$.

We presume that Y_it is given by:

$$Y_{it} = \delta_{it} D_{it} + \epsilon_{it} \tag{4}$$

where D_{it} is a binary variable, which is 1 if a domain has to follow the GDPR and 0 otherwise. So, δ_{it} is the heterogeneous treatment effect on unit i at time t.

We want to evaluate the average treatment effect on the treated (ATT) at time t ($t > T_0$)

$$ATT_{t,t>T_0} = \frac{1}{N_{tr}} \sum_{i \in T} [Y_{it}(1) - Y_{it}(0)] = \frac{1}{N_{tr}} \sum_{i] \in T} = \delta_{it} \tag{5}$$

In order to estimate the treatment effect, we must determine the appropriate weights for each unit in the control group. So the synthetic control estimator becomes:

$$\hat{ATT}_{t,t>T_0} = \frac{1}{N_{tr}} \sum_{i \in T} [Y_{it}(1) - \hat{Y}_{it}(0)] \tag{6}$$

It is important to note that the SCM method also relies on the assumption of parallel trends, which means that the treatment and control groups should have had similar trends in the outcome of interest prior to the treatment or intervention.

A.3.2 Results

The results of the SCM method for three traffic pattern comparisons are presented in Table 11 and Figures 8, 9 and 10. The results are consistent with those of the difference-in-difference analysis.

For the cookie compliance comparison, we did not find a significant effect for any of the four variables of interest.

For the comparison of Europe vs. America, the SCM results show significant effects on Log (Pageviews per Person) and Log (Hours per Person) after the implementation of the GDPR. Specifically, pageviews per person decreased by 7.03% and hours per person decreased by 8.6%. There was no significant effect on audience or sessions per person.

In the comparison between German traffic inside and outside the EU, the SCM did not find any evidence of a change in traffic due to GDPR.

Table 11: Synthetic control method outcomes

Dependent variable	ATT Estimate	S.E.	CI.lower	CI.upper	p.value
German traffic to compliant websites vs. non-compliant web- sites					
Log (Audience)	-0.063	0.058	-0.178	0.052	0.281
Log (Pageviews per Person)	-0.138	0.123	-0.379	0.103	0.261
Log (Sessions per Person)	0.019	0.044	-0.067	0.105	0.662
Log (Hours per Person)	0.067	0.091	-0.111	0.246	0.459
German traffic vs. American traffic					
Log (Audience)	0.033	0.026	-0.018	0.084	0.207
Log (Pageviews per Person)	-0.076	0.035	-0.144	-0.009	0.027
Log (Sessions per Person)	-0.006	0.017	-0.038	0.027	0.738
Log (Hours per Person)	-0.090	0.022	-0.133	-0.047	0.000
German traffic inside vs outside the EU					
Log (Audience)	-0.069	0.081	-0.226	0.089	0.394
Log (Pageviews per Person)	-0.080	0.078	-0.234	0.073	0.304
Log (Sessions per Person)	-0.031	0.030	-0.090	0.028	0.3061
Log (Hours per Person)	0.093	0.056	-0.203	0.017	0.099

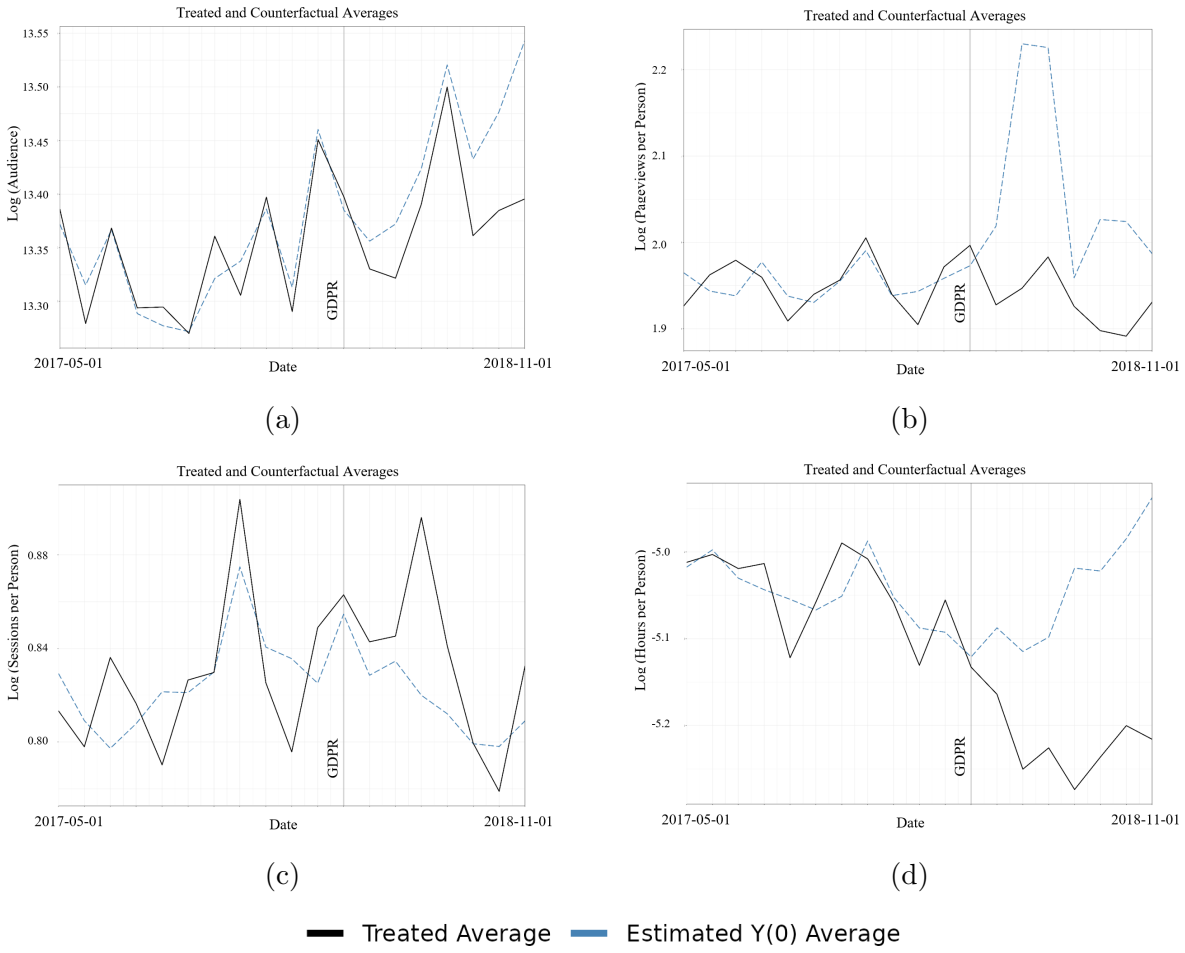

Figure 8: Synthetic control method: German traffic to compliant websites vs. non-compliant websites

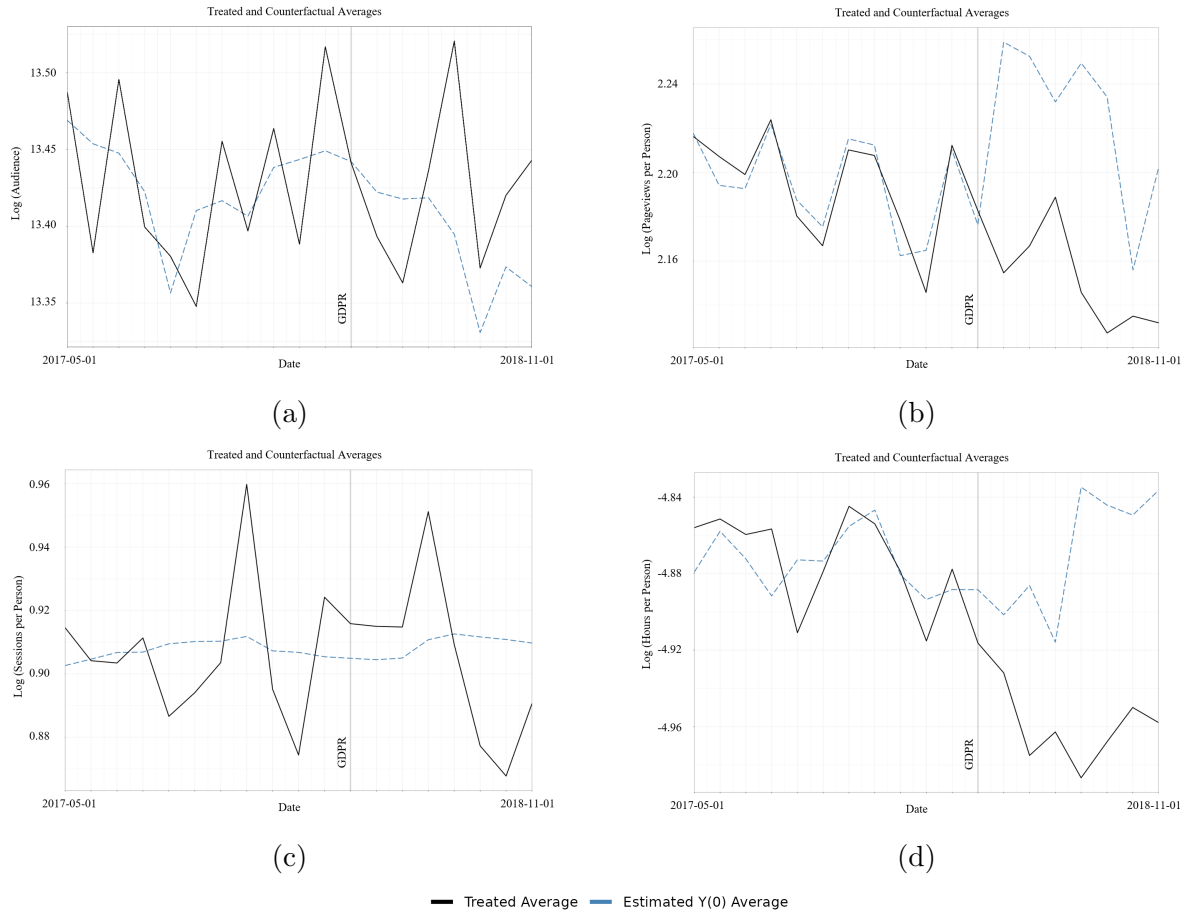

Figure 9: Synthetic control method: German traffic vs. American traffic

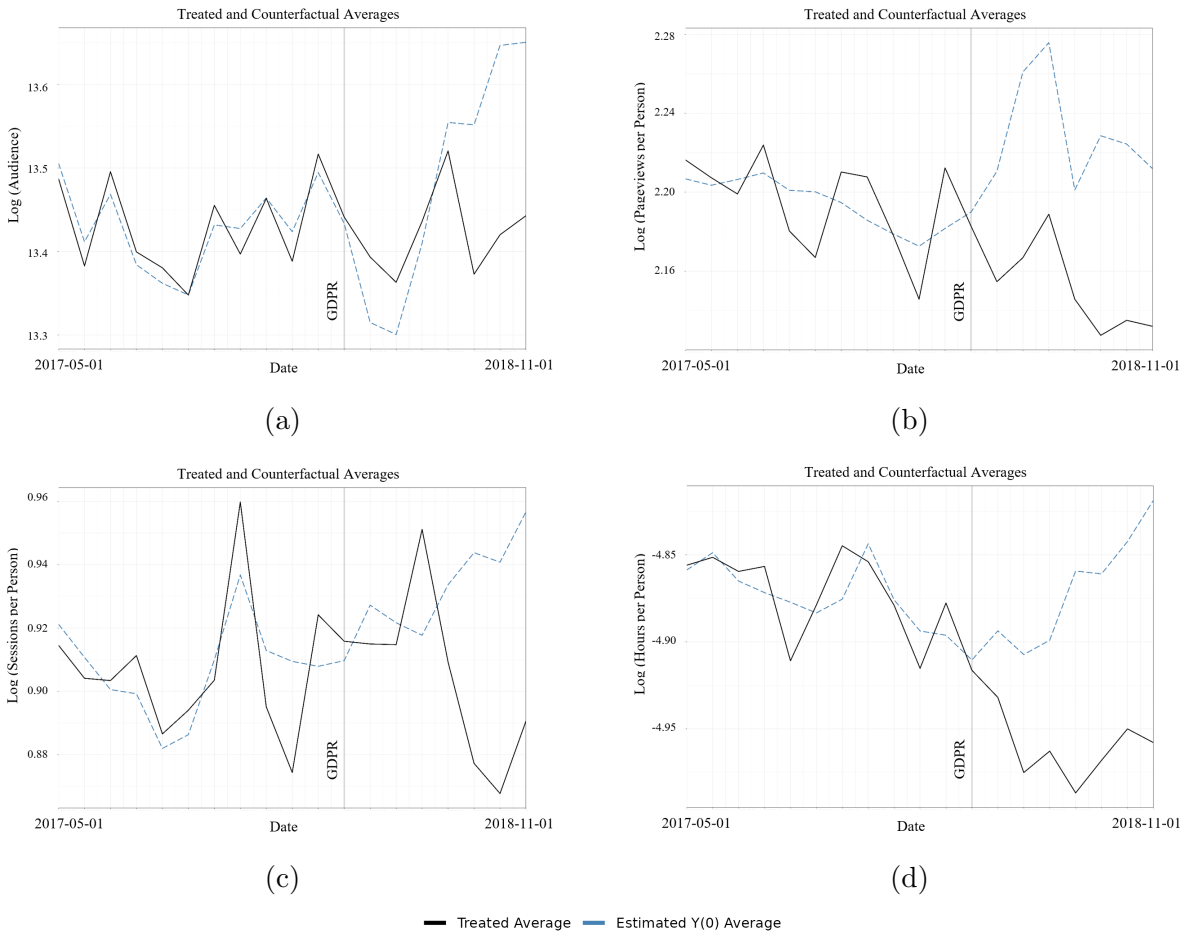

Figure 10: Synthetic control method: German traffic inside vs. outside the EU

A.4 R Code and Data

The analysis was carried out in R version 4.2.0. Our code and cookie compliance dataset can be found on GitHub at

https://github.com/carladausend/Master-Thesis-Carla-Dausend-CLSBE.

Please contact carladausend@gmail.com with any questions or requests for data access.