Privacy-Enhanced versus Traditional Retargeting: Ad Effectiveness in an Industry-Wide Field Experiment

Shunto J. Kobayashi, Garrett A. Johnson, & Zhengrong Gu*

September 30, 2024 Work in progress

Abstract

An advertiser can use retargeting to target its site visitors with ads offsite, often to push users to complete a purchase. Retargeting is controversial because it raises both user privacy concerns and questions about its effectiveness for advertisers. In this study, we partner with an advertiser intermediary to measure retargeting effectiveness across more than 2,000 advertisers, leveraging an industry-wide experiment to evaluate both traditional and privacy-enhanced retargeting approaches. Google's Privacy Sandbox allows advertisers to retarget users without tracking cross-site browsing by moving ad selling onto the user's device. We provide broad-based evidence that retargeting lifts advertisers' baseline conversions by 4.6%. While removing third-party cookies significantly reduces ad clicks and click-through conversions, implementing Privacy Sandbox recovers 46.3% of lost ad clicks, and 43.5% of lost click-through conversions. Importantly, when adjusting for ad expenditure, the performance gap between privacy-enhanced and traditional retargeting narrows: Sandbox's click per dollar and click-through conversion per dollar achieve 86.4% and 81.8% of traditional counterparts, respectively. We provide additional evidence exploring time heterogeneity and advertiser heterogeneity in treatment effects, suggesting that the limited overall performance of Privacy Sandbox may be due to the lack of supply-side adoption of Privacy Sandbox.

Keywords: privacy, online advertising, privacy-enhancing technologies, ad effectiveness

*Questrom School of Business, Boston University: Kobayashi <sjkobaya@bu.edu>; Johnson <garjoh@bu.edu>; and Gu <gzr@bu.edu>. Authors are listed in reverse alphabetical order. We thank seminar participants at Questrom School of Business and Brett Gordon for their helpful comments, and we thank our industry partners for their assistance with the data and our research. We thank the NET Institute (www.NETInst.org) for financial support.

1 Introduction

Digital advertising relies on cross-website and cross-app user identity to measure, target, and optimize ads. However, this dependence raises significant consumer privacy concerns, prompting regulations like the European Union's General Data Protection Regulation (GDPR) and various state laws in the United States. In response, major tech companies are adapting their business practices to improve privacy. Google plans to limit third-party cookies in 2025 and introduced Privacy Sandbox, a suite of technologies designed to maintain ad functionality while enhancing user privacy. Privacy Sandbox includes tools for ad targeting (Protected Audience API, Topics API), measurement (Attribution Reporting API), and fraud detection (Private State Tokens API). These use privacy-enhancing technologies such as data aggregation, k-anonymity, differential privacy, and on-device computation. These efforts represent a foundational change in ad technology, which requires marketers to adapt.

We study whether advertising that uses privacy-enhancing technologies effectively delivers advertiser performance. We focus on the retargeted advertising, a strategy in which advertisers target their site visitors with ads offsite. For instance, advertisers try to persuade users to purchase by targeting ads at users who browsed products or added a product to their online cart. Retargeting is controversial because it raises both user privacy concerns and questions about its effectiveness for advertisers. We partner with an advertiser intermediary to measure retargeting's effectiveness across more than 2,000 advertisers, leveraging an industry-wide experiment to evaluate both traditional and privacy-enhanced approaches to retargeting.

The British Competition and Markets Authority (CMA) is investigating Google's Privacy Sandbox: its effectiveness, its consequences for competition, as well as its consequences for user privacy. The CMA and Google devised an industry-wide experiment in which a subset of Chrome users are randomly assigned to three different experimental groups. The Status Quo group retains third-party cookies, the Privacy Sandbox group uses Privacy Sandbox but not third-party cookies, and the Cookieless group uses neither. This experiment represents an unusual *ex ante* approach both to industrial policy evaluation and technology adoption. The experiment is also unusual in its global, industry-wide scale and its complex consequences for the interdependent adtech industry. We examine the strengths and limitations that arise in this landmark experiment.

The effectiveness of retargeting is controversial. While some industry studies and research suggest high sales for consumers exposed to retargeted ads, this association may not be causal due to the self-selection of these consumers who may have purchased anyway. Additionally, retargeting could provoke consumer reactance, reducing ad effectiveness. Extant evidence on retargeting reflects single advertisers (Bleier & Eisenbeiss, 2015; Lambrecht & Tucker, 2013; Johnson et al., 2017b; Sahni et al., 2019). Our research contributes to this body of knowledge by providing generalizable evidence on the effectiveness of retargeting across more than 2,000 advertisers, free from industry and academic publication bias. Our experiment shows that increasing retargeted ad exposure increases the advertiser conversions by 4.6%. This intent-to-treat estimate compares all site visitors in the Status Quo and Cookieless groups. This result builds on recent efforts to measure ad effectiveness across many advertisers without selecting for advertisers that choose to run ad experiments.

Traditional retargeting depends on user identifiers like third-party cookies, though Privacy Sandbox uses on-device ad selling to enable retargeting with tracking user browsing. The experiment reveals that removing third-party cookies reduces ad clicks by 88.3% and ad-attributed conversions by 88.9% and 88.9%, relative to the status quo. The experiment also shows that relying on Privacy Sandbox instead of third-party cookies yields 47.4% fewer ad clicks, and 50.2% fewer ad conversions, relative to the status quo. We also find that the Cookieless group has 4.4% fewer conversions, while the Sandbox group has 3.55% fewer conversions relative to the status quo, although these effects are statistically indistinguishable from each other. Thus, the Privacy Sandbox recovers 46.3% of the lost ad clicks, and 43.5% of the lost in ad conversions. Despite this, the expenditure-adjusted performance gap between the Status Quo and Sandbox groups is narrower: Sandbox's click per dollar and click-through conversion per dollar reach 86.4% and 81.8% of the Status Quo group's levels, respectively. This suggests that the limited supply of Privacy Sandbox impressions helps explain the disparity in performance. Our heterogeneity results further reinforce this point, as Sandbox performance is related to its market penetration.

We contribute to several strands of literature. First, academics from various disciplines explore the Privacy Sandbox. Jerath & Miller (2023) surveyed subjects for their perceived privacy violations in online targeting, finding modest improvements in privacy perceptions with Privacy Sandbox. Alcobendas et al. (2023) used a structural model to simulate the impact of privacy re-

strictions, suggesting that Privacy Sandbox mitigates revenue loss for publishers and advertisers but could harm competition by favoring firms with substantial first-party data. Cooper et al. (2023) found consumer preference for Privacy Sandbox over identity-based replacements, aligning with Miller's findings on improved privacy perceptions. Additionally, research from Google reported that users felt better about privacy and control with topic-based personalization compared to third-party cookies (Lachner et al., 2023). Legal scholars have also discussed the competitive, governance, and regulatory consequences of Privacy Sandbox (Geradin et al., 2021; Nottingham, 2021; Olejnik, 2023, 2024). This body of work highlights the potential benefits and challenges of Privacy Sandbox, informing the present research on its effectiveness and market impact.

Second, our research provides generalizable evidence of the effectiveness of retargeting across more than 2,000 advertisers, free from industry and academic publication bias. This complements the work of Shapiro et al. (2021) and Wernerfelt et al. (2022), who explored similar themes with different advertising media. Shapiro et al. (2021) examined the effect of television advertising across 288 consumer packaged goods categories. They drew attention to the generalizability issue in advertising effectiveness research and highlighted potential biases in prior analyses. Wernerfelt et al. (2022) conducted a large-scale study involving hundreds of thousands of advertisers on Meta, focusing on the benefits of shared data across applications. Their findings illuminate the impact of online advertising on even small-scale advertisers and in sectors like CPG, retail, and e-commerce. Several past papers in marketing provide evidence of both television and online ad effectiveness across large collections of advertisers, though advertisers that choose to run such experiments may be selected.

Retargeting is a marketing strategy that targets users who have previously interacted with an advertisement or visited a website. The effectiveness of retargeting is controversial. While some industry studies and research suggest high sales for consumers exposed to retargeted ads, this association may not be causal due to the self-selection of these consumers who may have purchased anyway. Additionally, retargeting could provoke consumer reactance, reducing ad effectiveness. Extant evidence on retargeting is anecdotal. For example, Lambrecht & Tucker (2013) and Bleier & Eisenbeiss (2015) examined the effectiveness of different retargeting creatives that varied the featured product or category. Johnson et al. (2017b) provided novel evidence that retargeting can be effective when compared with a control group that was held out from retargeted advertising,

demonstrating significant lifts in website visits, transactions, and sales. Furthermore, recent studies by Sahni et al. (2019) and Li et al. (2021) confirm that retargeting can be effective. Sahni et al. (2019) showed that retargeting increases site visits in high-cost and high-involvement product categories, particularly effective in the days following a user's site visit. Li et al. (2021) demonstrated that email retargeting can boost purchases among online cart abandoners. Our research contributes to this body of knowledge by providing generalizable evidence on the effectiveness of retargeting across more than 2,000 advertisers, free from industry and academic publication bias. This extensive and unbiased dataset allows us to offer robust conclusions about retargeting's true impact—both using traditional cookie-based targeting and using privacy-centric alternatives—advancing the understanding of its role in digital advertising.

This article is organized as follows. The next section provides background on Privacy Sandbox. Section 3 explains the design of the industry-wide experiment and its inherent limitations. Section 4 describes our advertiser intermediary data. Section 5 explains our estimation methodology and Section 6 presents our results.

2 Background: Privacy Sandbox

Google Chrome remains one of the last major browsers to support third-party cookies. Due to privacy concerns, Safari and Firefox blocked these cookies by default in 2018 and 2019, respectively. Third-party cookies contain pseudonymous identifiers that facilitate tracking users across websites, which enables online behavioral advertising, user segmentation, retargeting, and ad effectiveness measurement. Adtech vendors use these cookies to optimize advertising campaigns. Websites also use third-party cookies more generally for user authentication, content personalization, and fraud prevention.

In response to privacy concerns, Google developed replacement technologies for third-party cookies under Google's "Privacy Sandbox" initiative. This aims to improve user privacy while retaining the benefits of personalized advertising. Google's Privacy Sandbox includes various technologies for specific advertiser use cases: targeting (Protected Audience API and Topics API), measurement (Attribution Reporting API), and fraud prevention (Private State Tokens API). These technologies employ privacy-enhancing techniques such as on-device computing, differential pri-

vacy, and zero-knowledge proofs. Additionally, Privacy Sandbox features APIs that limit information flow across websites and prevent covert tracking methods like fingerprinting.

Our research focuses on the Protected Audience API (PA API), because this facilitates retargeting. PA API reconfigures how online ad transactions are conducted by moving the process to the user's device. Traditionally, retargeting allows an online advertiser and publisher to recognize a user across different sites, which poses privacy risks. With PA API, when a user visits a retailer's site, the browser adds them to an "interest group" for ad targeting. This interest group encodes the advertiser's bidding preferences and links to the ad creative. Later, when the user visits a publisher's site, the publisher's site instructs the browser on the ad auction rules, and the browser conducts the auction using the stored interest groups. This method enhances privacy by keeping the data on the user's device and limiting cross-site tracking. Notably, Microsoft announced its Ad Selection API, which is compatible with PA API but makes some different design decisions. Should these standards come into alignment, this could be the basis of a new browser standard for privacy-centric advertising.

Early adoption and testing of PA API by adtech vendors like RTB House and Criteo in 2022 provided valuable insights. RTB House, during the summer of 2022, added 1.2 million users globally to PA interest groups and displayed over 7 million PA impressions from nearly 400 advertisers across 4,172 advertiser domains (Rumiński et al., 2022). They also collected data on interest group creation from 13,354 sites, noting that Google, RTB House, and Criteo were the primary adtech vendors involved. By the end of 2022, Criteo was adding approximately 80 million users to interest groups daily across 14,000 websites and receiving about 10 million PA bid requests each day (Höring, 2022). However, challenges such as low adoption by supply-side platforms and technical issues like latency and debugging were identified as hurdles for broader PA API implementation.

3 Experimental Design

3.1 Overall design

The experimental design involves randomizing 2% of all Chrome users at the browser level, beginning in January 2024. The design aims to help both industry and regulators evaluate the performance of Privacy Sandbox by experimentally turning on or off the Privacy Sandbox APIs and/or

Table 1: Experimental treatment group summary

	3rd Party Cookies	Privacy Sandbox	Share†
Status Quo	✔	✘*	1%
Privacy Sandbox	✘	✔	0.75%
Cookieless	✘	✘	0.25%

Note: *Google left on the targeting (Topics, PA) and measurement (attribution reporting) APIs in the Privacy Sandbox (see footnote 1). †Share refers to eligible Chrome users who do not block third-party cookies.

third-party cookies. Participating users are Chrome users on both desktop and mobile (Android) with recent browser versions that support both Privacy Sandbox and sharing the experimental labels. Note that this excludes iOS users, as Chrome-branded browsers on iOS are actually built on Safari technology.

The study involves three experimental groups: Status Quo (using third-party cookies and limited Privacy Sandbox functionality[1]); Privacy Sandbox (using Privacy Sandbox, not third-party cookies); and Cookieless (using neither).[2] The summary table below explains that these groups represent 1%, 0.75%, and 0.25% of eligible Chrome users respectively. To be eligible, Chrome users must not block third-party cookies.

Compliance is enforced at the level of the browser through technical means. Websites can potentially request consent before using the Sandbox APIs, though websites that do so would apply the same logic to third-party cookie consent.[3] Adtech vendors choose to what extent they adopt Privacy Sandbox tools, which creates interdependence in executing ad campaigns that rely on Privacy Sandbox. Johnson & Neumann (2024) show that adoption rates among vendors and websites are moderate. They provide recent statistics on a public dashboard: https://app.sincera.io/privacysandbox. Even conditional on adoption, vendors also vary by their level of investment in Privacy Sandbox tools.

The strengths of this design include its ambitious, industry-wide scope. The experiment is unprecedented in scale and methodology. It addresses industry-wide coordination problems by randomizing and enforcing compliance at the browser level. This allows for observing the real-

[1]In the Status Quo group, Google left on the targeting (Topics, PA) and measurement (attribution reporting) APIs in the Privacy Sandbox. This was done to facilitate ARA-based measurement between the Status Quo and Sandbox groups. However, Google encouraged vendors not to use the targeting APIs for this group to better capture the status quo. In particular, Google Ad Manager—which has 90%-100% in the publisher ad server market (CMA 2020)—did not use the targeting APIs for this group.

[2]Google and the CMA instead refer to these groups as control_1, treatment_1, and control_2 respectively.

[3]Thus, users that opt out at the site level should be symmetric across the Status Quo and Privacy Sandbox groups for websites that enable Privacy Sandbox opt out.

world effects and market-level consequences of Privacy Sandbox, in contrast with past industry tests that were limited in scope because they either relied on simulations[4] or on isolated tests of individual APIs.[5]

However, the design has several limitations owing to its complex and ambitious nature. Experiments assess the impact of a treatment holding all else equal. The radical nature of the Privacy Sandbox treatment in this experiment poses multiple challenges because experiments are predicated on the notion that all else but the treatment is held equal. However, Privacy Sandbox changes both how adtech vendors bid on impressions and how they measure campaign performance. This in turn creates several challenges for maintaining an "apples-to-apples" comparison between treatment groups. We discuss these challenges and our solution below.

3.2 ITT experimental design & measurement challenges

Our experimental design builds on the overall experiment to measure advertiser performance using an Intent-to-Treat (ITT) comparison of advertiser-side outcomes. In so doing, we overcome two challenges in measuring advertiser performance. First, ad buyer's own data would be incomplete because the buyer does not observe the experimental labels assigned to all users: buyers rely on upstream ad sellers to forward this information. Second, the experiment actually varies the measurement technology: these measurement technologies have inherent limitations and cannot be used in all treatment groups. By instead relying on data obtained directly from advertisers, we avoid these issues and measure clicks and conversions from each experimental group. Below, we describe these issues and our experimental approach in depth.

3.2.1 Selective label forwarding by upstream firms

The industry-wide nature of the design creates interdependence between firms. This interdependence creates a challenge for the experiment: selective forwarding of experimental labels by upstream partners. Website owners can observe their users' experimental labels, but ad buyers

[4] "Criteo's First Look at the Attribution Reporting API" (medium.com/criteo-engineering/criteos-first-look-at-the-privacy-sandbox-attribution-reporting-api-event-level-f96f42537b9c).

[5] "Results from Google Ads' interest-based advertising testing" (blog.google/products/ads-commerce/iba-test-results/), "Topics API: Criteo's First Look at Google's Interest-Based Advertising Solution" (medium.com/criteo-engineering/is-googles-topics-api-a-viable-replacement-for-interest-based-advertising-297076192bd).

Table 2: Supported measurement technologies by experimental group

Group	Cookie-based	Attribution Reporting API	Advertiser-side data
Status Quo	X	X	X
Sandbox		X	X
Cookieless			X

rely on upstream websites and ad sellers to pass this information along. If certain adtech partners do not forward the labels, the resulting data will not be representative: observed data labels will be selected due to differing competitive conditions across experimental groups. Criteo documents large imbalances between groups on this basis: Criteo estimates that 17% of its spending in the Sandbox group arose in unlabeled impressions compared to 45% in the Status Quo group (Selman, 2024).

We solve this difficult problem by obtaining onsite ad performance data from advertisers whose websites can directly obtain the experimental label from their visitors. We further supplement our advertiser-level performance data by collecting advertisers' aggregate ad spend and ad quantity data from our DSP. The label forwarding problem is acute at the point where ad buyers bid on an impression as they rely on upstream firms for this information. If the buyer's ad appears on the page, the buyer can interact with the user's browser at this time to query the user's experimental group. In this way, our DSP is able to measure each advertiser's ad quantity and ad spend by experimental groups.

3.2.2 Difference in measurement technologies across the treatment groups

The experiment is unusual in that it changes how the experimental outcomes are measured, which creates a second issue. The table below summarizes three measurement technologies and whether each is supported by the three experimental treatment groups.

In the status quo, ad measurement relies on third-party cookies (and adjacent technologies like pixels) to connect a user's ad exposure to that user's conversion outcomes on the advertiser's website. Cookies thereby allow post-exposure conversion tracking regardless of whether the user clicks on the ads. However, the Sandbox and Cookieless groups deactivate third-party cookies, so that ad buyers and advertisers can no longer attribute conversions in this way.

The Privacy Sandbox offers the Attribution Reporting API (ARA) as an alternative measure-

ment technology. ARA provides feedback to buyers on ad clicks and conversions. ARA and cookie-based technologies measure the same underlying user behavior, though their counting will diverge for several reasons. First, ARA adds differential privacy noise to its measurement reports. Second, ARA caps the total conversion events per user (e.g., multiple purchases), whereas cookie-based measurement does not impose a cap. Third, ARA coarsens possible conversion events to a maximum of eight possible conversion types per user, whereas cookie-based conversion event types are not capped.

As the above table indicates, ARA is turned on in the Status Quo group. Ad buyers can use ARA to compare the Status Quo and Sandbox groups—in theory. However, ARA still can not be used for the Cookieless group, which deactivates all Privacy Sandbox technologies. The Cookieless group is important here as it present a low-ad baseline that helps identify the incremental effect of retargeted ads.

Measurement is necessary but not sufficient for analyzing the experiment: ad buyers need to be able to connect outcomes to experimental labels. As discussed above, ad buyers may not observe the experimental label when it is not shared by upstream firms. This problem is acute at the time of bidding, though buyers can try to collect the labels post-exposure. Thus, appending experimental labels to cookie- or ARA-based measurement poses a non-trivial challenge for ad buyers.

3.2.3 ITT approach using advertiser-side data

To circumvent the above measurement issues, we rely on advertiser-side data to measure ad performance. As the above table indicates, this is the only measurement technology that works with all three experimental groups. Advertisers can track several performance outcomes using their proprietary onsite data alone. In particular, an advertiser can see which users arrive on its website by clicking on a display ad by encoding this channel information in the ad's target URL. Thus, while advertisers do not observe all users that view its ads, advertisers can observe the subset of users that click on its ads.

Advertisers can also track onsite conversion events like adding products to cart and purchases. However, advertiser-side conversion data are disconnected from ad exposure and the data include

(but do not distinguish between) both incremental conversions that are caused by ad exposure and baseline conversions that would occur without ad exposure. Advertisers can identify the subset of conversions by users who arrived on site via an ad click: i.e., click-through conversions. Here too, advertiser-side click-through conversion data excludes users that convert after viewing but not clicking ads (i.e., view-through impressions).

Our experimental approach relies on an ITT comparison of total outcomes across all users by treatment group. In particular, we do not observe which users are exposed to ads in each group and exposed users are not symmetric across groups due to different ad exposure levels. Our design therefore resembles the ITT design of Johnson et al. (2017b), which also considers aggregate outcomes by treatment group without knowing the quantity of treatment eligible users.

Though click-through conversions and (unobserved) incremental conversions only overlap partially, they provide complementary ad effectiveness measures. In particular, Gordon et al. (2023b) find that the two measures—in aggregate, at the ad campaign-level—are correlated. Online ads exhibit notoriously low statistical power (Johnson, 2023; Lewis & Rao, 2015). Our advertiser estimates have low power because they contain at most 2% of Chrome users, though we pool across more than 2,000 advertisers and multiple weeks. Our ITT estimates reveal the incremental effect of cookie- or Sandbox-based retargeting on conversion. Since this is imprecisely estimated, our heterogeneity analyses often rely on click and click-through conversions as more precise proxy measures of ad performance.

Third-party firms like our DSP can gain access to advertiser-side data through a process referred to as tagging. Websites can set up a vendor-specific tag on their website such that the vendor receives a live stream of onsite events. Websites can set up tagging to help their ad buyers to measure ad effectiveness and to maintain user remarketing lists for targeting. By default, this data would not contain experimental group information. However, our DSP was able to leverage its tagging presence on advertisers' sites to query for users' experimental labels and thereby avoid the aforementioned label-forwarding problem.

Several researchers that study privacy restrictions on advertiser performance use similar data to avoid the mechanical problems that privacy restrictions pose for ad measurement. Our data most closely resemble those in Aridor et al. (2023, 2024), which obtain advertiser-side data from intermediaries that partner with multiple advertisers. In contrast, related data types like URL

toolbar data (e.g., Budak et al., 2016) and site analytics data (e.g., Goldberg et al., 2024) typically exclude experimental labels. Note that we may not observe the data of users that do not consent to sharing data with our DSP (see e.g., Goldberg et al., 2024), though we expect that consent rates are high and symmetric across experimental groups.

3.3 Fundamental caveats of the experiment

The experiment has fundamental limitations due to the complexity of the advertising market and the experimental treatment. We explain below how these limitations may impact our results

3.3.1 Internal validity

We do not observe user-level data, so we cannot independently validate the experiment's internal validity. We assume that Google, which is experienced with running experiments, properly randomized Chrome users into treatment groups.

Nevertheless, internal validity challenges can arise from bot traffic with the experimental labels, as this creates imbalanced group sizes. These imbalances seem to be an unavoidable feature of a field experiment conducted on the open web rather than a locked-down platform. In related work, we obtain publisher data containing over a billion ad impressions from this experiment. This data shows economically small imbalances between treatment groups. These differences are statistically significant given this large sample size, so the experiment fails randomization checks.

We expect that bot traffic explains the small experimental imbalances observed elsewhere. Nevertheless, bots do not make purchases, so any experimental imbalance arising from bots should not bias our conversion results.

3.3.2 Advertiser strategic considerations

To facilitate comparison of ad performance, in principle, we would like ad buyers to hold elements of their bidding fixed between groups wherever possible. In practice, this is difficult or impractical to execute for several reasons.

Privacy Sandbox restricts an advertiser's targeting technologies to Topics API and PA API. Many advertisers must therefore adjust their campaign strategy. This adjustment is an interesting

dimension of evaluating Privacy Sandbox. However, our data instead exclusively consists of retargeting campaigns, which readily translate to PA API. Moreover, our DSP maintains consistent campaign setups between the Status Quo and Sandbox groups. Our setting therefore stands out in that it maintains the campaign targeting strategy between these groups. Nonetheless, technical restrictions in PA API may affect campaign performance in the Sandbox group. PA API imposes limitations on the ad components (e.g., products) that can be included in the ad: extant research suggests these differences affect ad performance (Bleier & Eisenbeiss, 2015; Lambrecht & Tucker, 2013). Moreover, PA API has significant latency issues with ad load times (as currently implemented) and PA interest groups expire after a maximum of 30 days.[6]

Maintaining consistent ad spending or ad performance between groups is difficult and/or impractical. First, market conditions and available supply vary between conditions. Some ad sellers do not participate in PA API auctions, so some users are not reachable in the Sandbox group. Furthermore, ad prices will diverge between groups owing to varying Privacy Sandbox API participation by adtech vendors on all sides of the market. Second, treatment groups vary in the measurement technologies that they support, as discussed above. Third, Privacy Sandbox reporting of ad performance, reach, and frequency are nontrivial to set up and are typically reported with noise and delay. Fourth, the experiment consists of less than 2% of Chrome users: resulting advertising figures may be imprecisely estimated at the advertiser level. Nevertheless, one DSP (Criteo) sought to maintain ad performance between Privacy Sandbox and Status Quo groups in its test; however, Criteo relied on advertiser site visits as a proxy performance outcome to improve statistical precision (Selman, 2024).

Given these challenges, we treat ad campaign performance as an outcome variable. In practice, our DSP spends less per user in the Sandbox group owing to more limited PA API supply. We therefore provide complementary performance results where we normalize by ad quantity or ad spend.

A buyer that adopts PA API must create a new bidding system that may compete with the buyer's traditional buying system. The buyer's traditional buying system may operate in all three experimental groups. The buyer can bid on these impressions without user-level data by using

[6]PA therefore excludes certain retargeting campaigns like reactivation campaigns targeting inactive users after more than 30 days. Our results omit this comparison. For a discussion of latency issues, see this issue from RTB House: https://github.com/WICG/turtledove/issues/1207.

contextual information about the ad placement alone (e.g., the website URL).[7] The buyer's traditional bidding system may also use alternative identifiers, which can a range from probabilistic (e.g., fingerprinting) to deterministic identifiers (e.g., hashed email-based). The buyer may also adapt its traditional bidding system to ingest and integrate Topics API category data. In principle, buyers should therefore also optimize the interplay and budget allocation between their traditional and PA API buying systems. In practice, buyers may forgo engineering sophisticated solutions to this problem given the uncertainty surrounding Privacy Sandbox and third-party cookie deprecation on Chrome.

Our DSP largely avoids these problems by focusing on (third-party) cookie-based retargeting. As such, our DSP intended to only buy using its PA API bidding system in the Sandbox group, thereby simplifying the comparison between groups. Despite this, we see below that experimental contamination disrupted this plan. Note that our DSP also does not make use of Topics API for the campaigns that we consider.

3.3.3 Experimental contamination

Ad experiments may experience various forms of contamination. As with other ad experiments, our experimental design may miss cross-device conversions, leading to potential biases in the results (Lin & Misra, 2022). However, our experiment exhibits a novel form of contamination. If cookie-based ad buying persists in the Sandbox & Cookieless groups, we would say this contaminates the experiment. In particular, some adtech companies may circumvent the loss of third-party cookies—for instance, by using fingerprinting to track users covertly—and thereby artificially inflate the performance of the experimental groups where third-party cookies are deprecated. Our DSP does not use fingerprinting or other covert means to track users.

Nevertheless, a seller behavior referred to as "ID bridging" leads to contamination in the experiment. "ID bridging" describes a phenomenon whereby the seller misrepresents that an ad has a given 3rd-party (cookie-based) identifier when that identifier is actually absent. Here, the seller can use its 1st-party user data to recover 3rd-party identifiers when these are deleted in the Cookieless & Sandbox groups. For instance, example.com can recognize a user (e.g., based on its first-party cookie ID of "ABC123") that visited its site prior to the experiment and know that

[7] The publisher may also opt to share a first-party ID to, for instance, help the DSP manage ad frequency.

this user has a third-party cookie ID (e.g., "EFG456") associated with one of the site's ad buyers. After the experiment, example.com can "bridge" its 1st-party ID to reconstitute this user as having third-party ID of "EFG456" and declare the user as such in example.com's bid request. This process is debatably fraud if it is not clearly communicated to the buyer. ID bridging thereby leads to contamination in the Sandbox & Cookieless conditions.

Again, our partner DSP expected that its cookie-based buying system would not purchase impressions in the Sandbox & Cookieless groups. Nevertheless, our DSP inadvertently did so in cases where the seller uses ID bridging. The contamination problem is therefore limited in scope to sellers engaging in ID bridging.

One contribution of our study is to illuminate the empirical scale and the decay rate of ID bridging. The utility of ID bridging should decay over time both as sites' first-party user data degrades (e.g,. users clear first-party cookies) and the value of past user behavioral information decays for advisers.

3.3.4 External validity

The performance of Privacy Sandbox depends on industry adoption and investment. The reverse is also true, though performance is unknowable beforehand. This creates a "chicken and egg" problem for industry and regulators. In particular, extrapolating our findings to the long-run equilibrium is challenging: we observe the market scale and Privacy Sandbox API implementation maturity during the first half of 2024, which are expected to evolve. Moreover, the market conditions within the Privacy Sandbox group may be softer owing to moderate adoption of these tools. Nevertheless, our research focuses on whether privacy-centric advertising can be effective. The policy implications depend on effect size, and our study provides directional evidence to inform this evaluation.

4 Data

The data for this research is sourced from an anonymous Demand Side Platform (DSP), and we henceforth refer to it simply as the "DSP." A Demand Side Platform (DSP) is a technology platform that enables advertisers to purchase digital advertising inventory from multiple ad exchanges and

data sources through a single interface, allowing for efficient management and optimization of ad campaigns. Our DSP specializes in retargeting, a marketing strategy that targets users who have previously interacted with an advertisement or visited a website. This DSP uses Privacy Sandbox's PA API for retargeting purposes. Though adoption of Privacy Sandbox is relatively modest (see market level statistics at https://app.sincera.io/privacysandbox), our DSP is a leading adopter of Privacy Sandbox technologies.

The dataset consists of weekly aggregated outcomes for advertisers, segmented by device and experimental group. The outcomes measured include clicks, click-through conversions, and overall conversions. Conversions usually refer to purchases on e-commerce sites, though some represent clicks on affiliate links. Data also include advertiser characteristics, such as country and industry, that allow us to capture advertiser heterogeneity. In addition, we have other outcome variables, such as impressions, that can be used for further analysis.

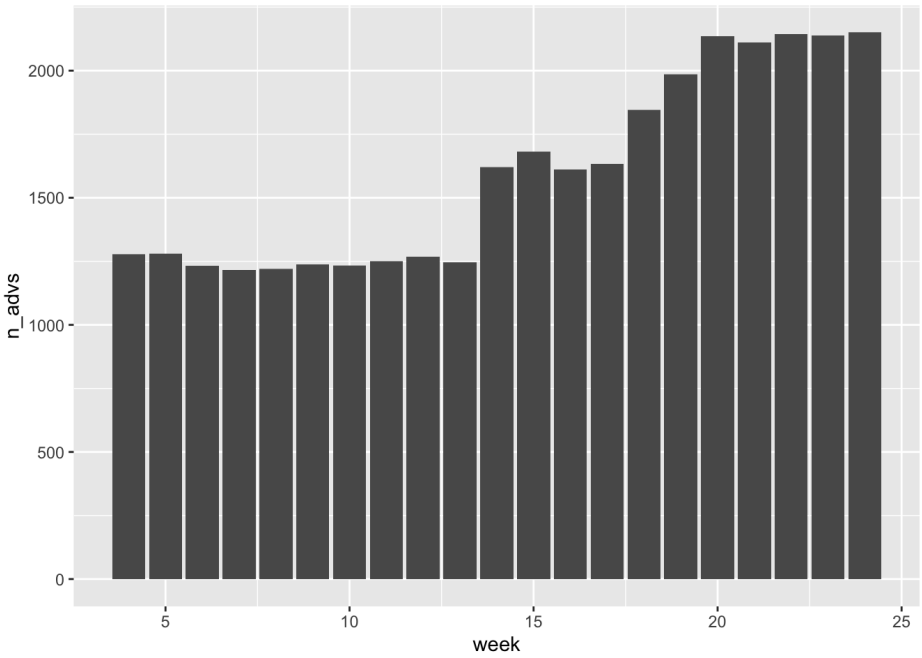
Our DSP data encompass a variety of e-commerce industries, including travel, beauty, electronics, and clothing. The sample size gradually increases as the DSP incorporates support for the advertiser's preferred creative formats, and there is some week-to-week attrition as advertisers modify their campaign configurations. Data collection starts at the end of January with approximately 1,400 active advertisers, expanding to a little over 2,000 by mid-June. Figure 1 shows the number of advertisers who actively participate in PA API per week, and Table 3 shows summary statistics for our key advertiser outcome variables (excluding spending variables for confidentiality), where the unit of observation is an advertiser-device-week from every experimental group. Our data contains 201,060 such observations to date.

5 Estimation Methodology

The goal of our analysis is to understand the effect of removing third-party cookies or replacing them with Privacy Sandbox for an average Chrome user. Because the treatments are randomly assigned across Chrome users, we do not have endogeneity issues unlike non-experimental studies. However, because our data are aggregated by experimental groups and they vary in size, we need to account for these size differences to capture the treatment effects.

For each advertiser k in each week t, we observe an aggregate outcome of interest Y_{ktg} gener-

Figure 1: Weekly number of advertisers who advertise through PA API



ated by Chrome users in experimental group $g \in \{StatusQuo, Cookieless, Sandbox\}$. The variable Y_{ktg} can be ad clicks, post-click conversions, and all conversions. Note that we also observe outcomes at the device type level (i.e., desktop versus mobile), but omit this below to economize on notation. Users are assigned at different rates into the experimental groups, so the outcome variables—which are aggregated measures—are mechanically influenced by the group sizes. Nevertheless, the average relative treatment effect is identified by leveraging the randomized treatment assignment and using the information on the treatment rate.

Suppose there are N treatment-eligible Chrome users. Each user i is assigned with treatment with probability r_g and generates y_i number of clicks, conversions, etc. We have an i.i.d. sample of $Y_g = \sum_{i:g_i=g} y_i$ for each g, so we can estimate the left-hand side of the following:

$$E[Y_g] = N r_g E[y_i \mid g_i = g].$$

Using this identity, we can identify the relative average treatment effect at the user level as

$$\%ATE_g = \frac{E[y_i \mid g_i = g] - E[y_i \mid g_i = StatusQuo]}{E[y_i \mid g_i = StatusQuo]} = \frac{\frac{r_{StatusQuo}}{r_g} E[Y_g] - E[Y_{StatusQuo}]}{E[Y_{StatusQuo}]}$$

Table 3: Summary Statistics

Variable	Mean	sd	Min.	p25	Median	p75	Max.
Ad clicks	30.28	166.67	0	0	2	15	8595
Post-click conversions	2.39	42.44	0	0	0	0	2703
All conversions	59.29	609.28	0	0	3	12	35116
Classic impressions	1825.52	9831.73	0	3	62	928	849862
Classic clicks	15.88	123.75	0	0	0	2	8592
Classic post-click conversions	1.12	29.24	0	0	0	0	2581
PAAPI impressions	645.19	4052.83	0	0	0	156	335485
PAAPI clicks	4.62	46.95	0	0	0	0	4070
PAAPI post-click conversions	0.31	9.81	0	0	0	0	955

for each $g \in \{Cookieless, Sandbox\}$. Since N cancels out, knowing the total number of Chrome users is not necessary for identification.

We employ Poisson quasi-maximum likelihood (i.e., Poisson regression) to estimate the relative average treatment effect. Poisson regression is suitable for our analysis for a few reasons. First, our outcome variables involve counts (e.g., clicks, conversions, etc). Second, we are interested in identifying treatment effects in percentage terms, and Poisson regression circumvents the well-known issue that arises in linear regression with a logarithmized response variable while also having similar robustness properties (Chen & Roth, 2024; Wooldridge, 2010). Finally, Poisson regression has a natural way to account for the difference in group sizes by using exposure variables.

We use r_g as the exposure variable and analyze the following regression:

$$E[Y_{ktg}] = \exp\left(\log(r_g) + \tau_{Cookieless}\mathbb{1}\left(g = Cookieless\right) + \tau_{Sandbox}\mathbb{1}\left(g = Sandbox\right) + \xi_k + \xi_t\right),$$

where $\mathbb{1}\left(\right)$ is the indicator function ξ_k and ξ_t and are advertiser and week fixed effects, respectively. Having the logarithmic treatment rate as an offset in the regression allows us to account for the fact that the experimental groups have different sizes (Cameron & Trivedi, 2013). Interacting the treatment dummy variables with other variables, such as week, can capture potential heterogeneous treatment effects.

Table 4: Relative Average Treatment Effects and Recovery Share

	Cookieless effect	Sandbox effect	Recovery Share
Ad clicks	-88.3% (-89.7%, -86.9%)	-47.4% (-49.7%, -44.7%)	46.3% (43.6%, 49.4%)
Click-through conversions	-88.9% (-92.6%, -84.7%)	-50.2% (-56.0%, -45.4%)	43.5% (35.4%, 49.2%)
All conversions	-4.38% (-6.90%, -1.40%)	-3.55% (-5.28%, -1.97%)	18.7% (-95.2%, 43.3%)

Note: 95% bootstrap confidence intervals are obtained with advertisers as clusters.

Our main objects of interest are the relative treatment effects, which can be obtained by transforming coefficients τ_g. Given the form of the conditional mean, the percentage effect from a status-quo subject receiving treatment is

$$\%ATE_g = \exp\left(\tau_{Cookieless}\mathbb{1}\left(g = Cookieless\right) + \tau_{Sandbox}\mathbb{1}\left(g = Sandbox\right)\right) - 1$$

We are also interested in how much of the performance drop from removing third-party cookies is recovered by introducing the privacy-friendly technology, so we analyze the *recovery share*, defined as the relative difference between the sandbox and cookieless treatment effects:

$$\text{Recovery Share} = \frac{\%ATE_{Cookieless} - \%ATE_{Sandbox}}{\%ATE_{Cookieless}}.$$

We bootstrap with advertisers as clusters to obtain robust standard errors for the regression coefficients and the other objects of interest.

6 Results

We find that removing third-party cookies significantly negatively impacts advertisers' key performance metrics due to the reduced ability to effectively retarget cookieless users. Nevertheless, using the Protected Audience API (PAAPI) provided by Privacy Sandbox can recover some of this loss. Furthermore, we observe heterogeneous treatment effects based on advertiser characteristics, illustrating how limited access to PAAPI-eligible impressions can constrain the extent of recovery.

Table 4 shows the treatment effects and recovery share for each outcome variable. Removing third-party cookies significantly hurts the amount of ad clicks and post-click conversions that advertisers receive from users. We find that removing third-party cookies from an average Chrome users results in 88.3% fewer ad clicks and 88.9 fewer click-through conversions from such users. Given that the DSP specializes in retargeting, theoretically we should not observe ad clicks and click-through conversions from cookieless users, but we find some of the users in the Cookieless treatment group had their cookies recovered by ID bridging, where a publisher or its intermediary assigns a cookieless user with their potential past cookie by using some linkage method. Nevertheless, we show in the next section that ID bridging becomes less effective over time.

In addition, we provide novel evidence of the effectiveness of retargeting in driving conversions. Specifically, advertisers experience approximately 4.4% fewer conversions from cookieless users compared to status quo users, which is a substantial impact given that this outcome measure includes baseline conversions through other channels (such as recurring customers). Since third-party cookies are essential for retargeting, and our DSP specializes in this strategy, this finding also means that retargeting lifts advertisers' baseline conversions by around 4.6%.[8] This offers generalizable insights into the impact of retargeting, based on data from around 2,500 advertisers. This evidence contributes to the growing literature that provides broad-based experimental evidence of ad effectiveness (e.g., Lodish et al., 1995; Gordon et al., 2023a; Johnson et al., 2017a) as well as ad effectiveness estimates that do select for advertisers that run experiments (Shapiro et al., 2021; Wernerfelt et al., 2022).

The estimated Privacy Sandbox effects are negative for all the outcome variables, but they are much smaller in magnitude compared to the cookieless effects, suggesting that Privacy Sandbox allows advertisers to retarget consumers in a privacy-friendly fashion. We find that an average Chrome user without third-party cookies but with PAAPI enabled generates 47.4% fewer ad clicks, 50.2% fewer click-through conversions, and 3.55% fewer conversions compared to a status-quo user. In fact, we find that the recovery share is around 50% for clicks and post-click conversions (statistically significant at 5% level), and while it is not statistically significant and imprecisely measured, the recovery share is around 20% for conversions. These findings demonstrate that

[8]This relative effect is obtained by using the total conversions from the Cookieless group as the denominator rather than the Status-Quo group.

introducing PAAPI to cookieless users helps mitigate the negative effects of third-party cookie removal.

6.1 Temporal Evolution

As the experiment spans multiple months, market participants have the opportunity to learn and adapt their strategies for utilizing Privacy Sandbox; hence, it is important to analyze the time trend in treatment effects. To analyze how these effects evolve, we interact the treatment indicators with the "week" variable in a Poisson regression with the specification given in Section 5. By illustrating the trend in each treatment effect in Figure 3a, we find that both the cookieless effect and the smaller sandbox effect deteriorate over time for clicks and click-through conversion. We find that the cookieless effects on ad clicks and click-through conversions are initially around -75%, but over the course of twenty weeks, they decline to nearly -90% by the end of the sample period. The worsening cookieless effect may reflect the limitations of ID bridging, where a publisher or its intermediary assigns a cookieless user with their potential past cookie by using some linkage method. This suggests that identifying cookieless users becomes increasingly difficult as time without third-party cookies progresses. As a result, the long-term impact of third-party cookie deprecation could be more detrimental than our short-term estimates imply. This finding serves as a caution for industry participants relying on ID bridging to mitigate the effects of third-party cookie removal.

The sandbox effects on ad clicks and click-through conversions exhibit non-monotonic patterns but follow an overall downward trend in Figure 3a, starting around -40% and falling to -50%.[9] Given that the DSP specializes in retargeting and has invested significant efforts and resources into utilizing PAAPI, it is unlikely that the decline is resulting from changes in their effectiveness in advertising through PAAPI. It is more likely that the downward trend results from a shrinking supply of PAAPI eligible impressions; in fact, we find a similar downward trend in the number of impressions and expenditure in Section 6.2. Generating clicks and click-through conversions through PAAPI relies on the ability to purchase a sufficient number of impressions via PAAPI. Indeed, our findings in Sections 6.2 and 6.3 suggest that the ad clicks and click-through conversions

[9] Although advertisers who adopt PAAPI later have a lower recovery share, as discussed in Section 6.3.4, the temporal evolution result remains robust even when focusing only on the approximately 1,400 initial adopting advertisers.

generated through PAAPI depend on the market conditions affecting the advertiser's ability to purchase impressions through PAAPI.

Figure 3b shows the overall trend of the recovery share with 95% confidence bands. The weekly estimated recovery shares (dotted curves) indicate an improvement in the early periods, but a decline starting around week 12, leading to a downward average trend (solid curves computed via Poisson regression with week treated as a continuous variable). Given that our data provider is actively investing in improving its implementation of Privacy Sandbox and market conditions change over time (both in the supply and demand of PAAPI eligible impressions), observing the evolution of the recovery share over time helps inform how Privacy Sandbox may perform in the longer run.

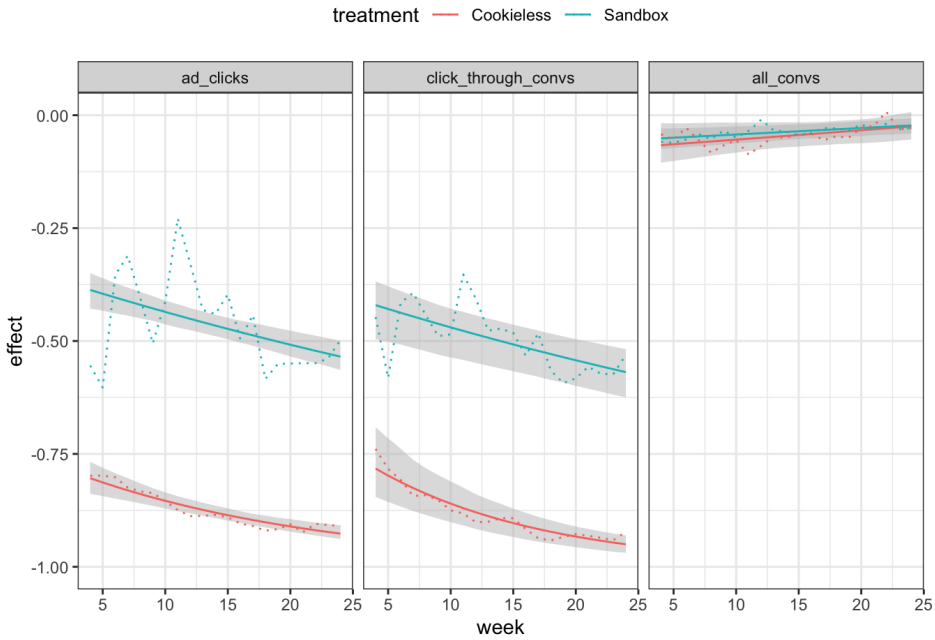
6.2 Expenditure and Impressions

To better understand the factors influencing treatment effects, we analyze impressions and expenditure across treatment groups and between the classic channel and PAAPI. Since impressions and expenditure data became accurately available from week 10, our analysis in this and the next section uses data from that point onward. Figure 3 presents the average outcome variables for each treatment group, normalized by the Status Quo group's averages. Since users in the Sandbox group can be reached through both the classic channel and PAAPI, we decompose the outcomes to identify which channel drives performance.

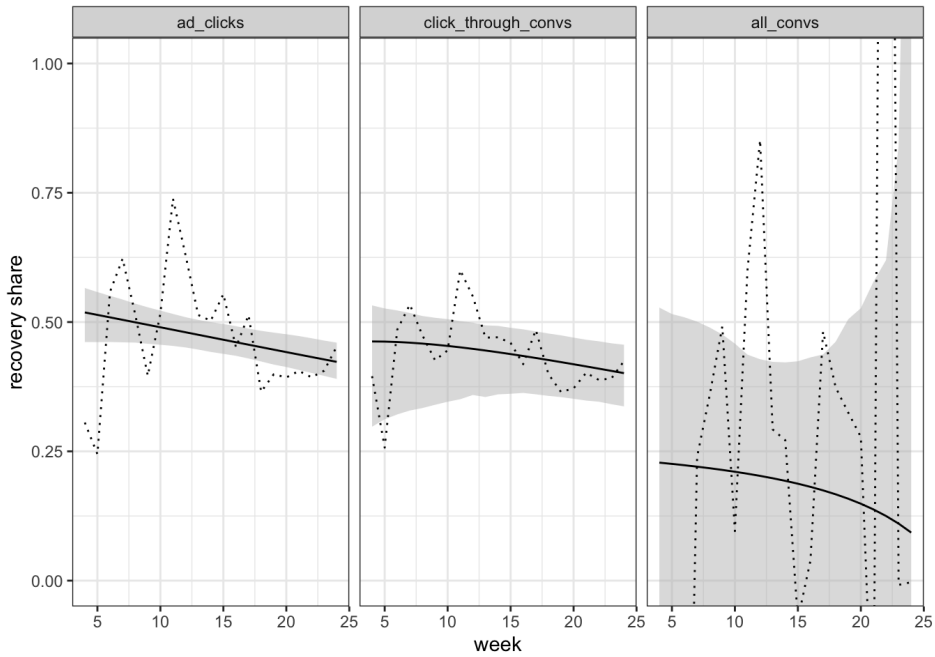
The figure indicates that the superior performance of the Sandbox group, compared to the Cookieless group, is primarily driven by ad opportunities purchased through PAAPI. Echoing results from Table 4, impressions and ad expenditure associated with the Cookieless group are approximately 90% lower than those of the Status Quo group. As previously mentioned, some impressions from the Cookieless group were inadvertently purchased by the DSP due to ID bridging, which matches these users with past cookies that may have belonged to them. The DSP's retargeting strategy in the classic channel relies on cookies, so without ID bridging, these impressions theoretically should not have been purchased.

We find that the Sandbox group achieves similar levels of outcomes to the Cookieless group through the classic channel but generates significantly more outcomes through the PAAPI channel.

Figure 2: Relative ATEs and Recovery Share over time



(a) Relative ATEs



(b) Recovery Share

Note: The dotted curves are estimates per week, and the solid curves are estimated via Poisson regression (with 95% bootstrap confidence bands).

This results in the Sandbox group attaining around half of the ad clicks, click-through conversions, impressions, and spending of the Status Quo group. Figure 3 highlights another important point: the sandbox effects on impressions and expenditure are comparable to those on ad clicks and click-through conversions. This suggests that the performance gap between the Status Quo and Sandbox groups is not primarily due to a lack of efficacy in advertising through PAAPI.

Figure 4 further supports this observation, showing that the impressions/dollar-adjusted performance of PAAPI advertising in the Sandbox group is on par with that of the Status Quo group. Using the average of ad clicks, click-through conversions, impressions, and spending, we compute ratio metrics such as clicks per dollar, conversions (click-through) per dollar, clicks per impression, conversions (click-through) per impression, and conversions (click-through) per click. These metrics achieve over 80% of the Status Quo group's performance. Notably, the click-through rate and conversion rate metricsc are mostly statistically indistinguishable from Status-Quo. Additionally, when examining dollar-adjusted figures, we find that the Sandbox's click per dollar and click-through conversion per dollar reach 86.4% and 81.8% of the Status Quo group's levels, respectively. This suggests that the underperformance of the Sandbox relative to the Status-Quo may largely stem from limitations in obtaining advertising opportunities via PAAPI, rather than a lack of effectiveness in retargeting.

Lastly, Figure 5 shows the temporal evolution of the treatment effects on impressions and expenditure. Similarly to the effects on ad clicks and click-through conversions, we find a downward trend in the sandbox effects; as time goes on, the sandbox effects on impressions and expenditure worsen over time. Initially, the sandbox effects on impressions and expenditure were around -25%, but they drop to approximately -50% by the end of the sample period. Given that the DSP consistently invested resources into utilizing PAAPI, we believe that the decline is coming from shrinking supply of PAAPI-eligible impressions, which ultimately contributed to the worsening sandbox effects on ad clicks and click-through conversions.

6.3 Advertiser Heterogeneity

We now explore how various advertiser characteristics—country, size, business type, industry, cohort, and data quality—influence treatment effects. Overall, we find that both cookieless effects

Figure 3: Channel-specific average outcomes

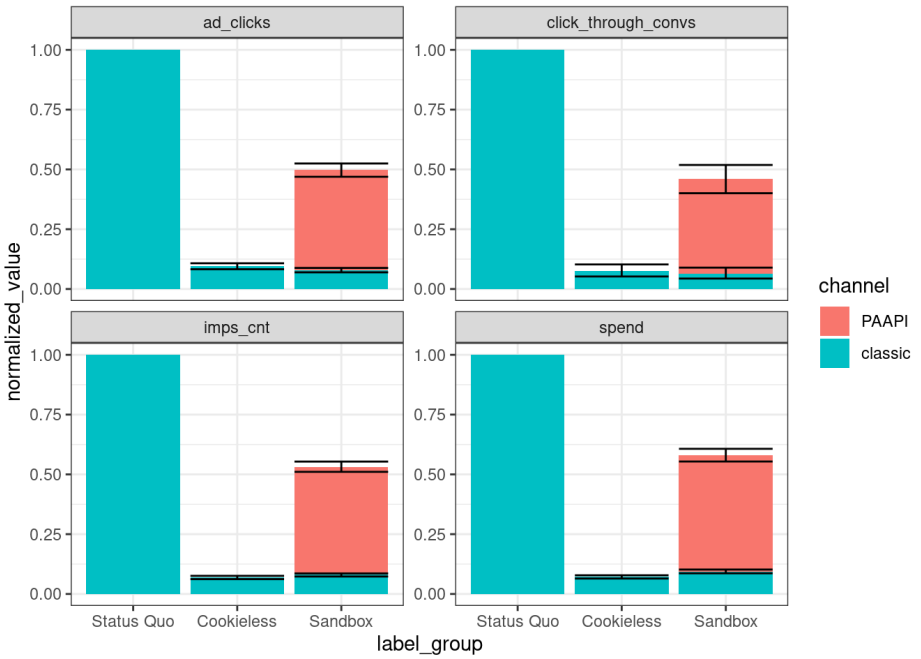


Figure 4: Performance metrics

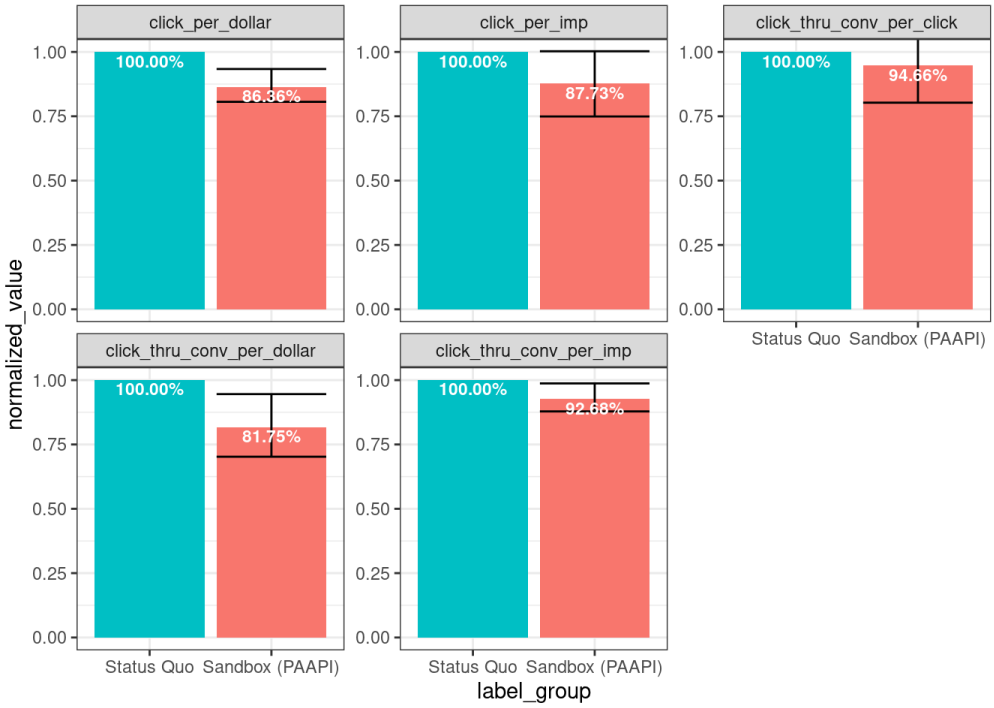
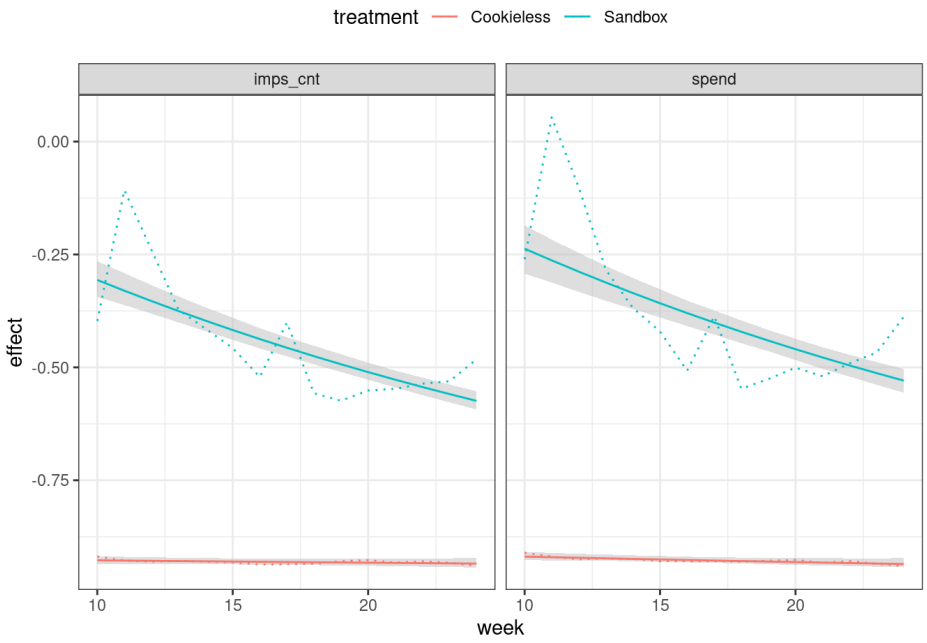


Figure 5: Relative ATEs on impressions and expenditure over time

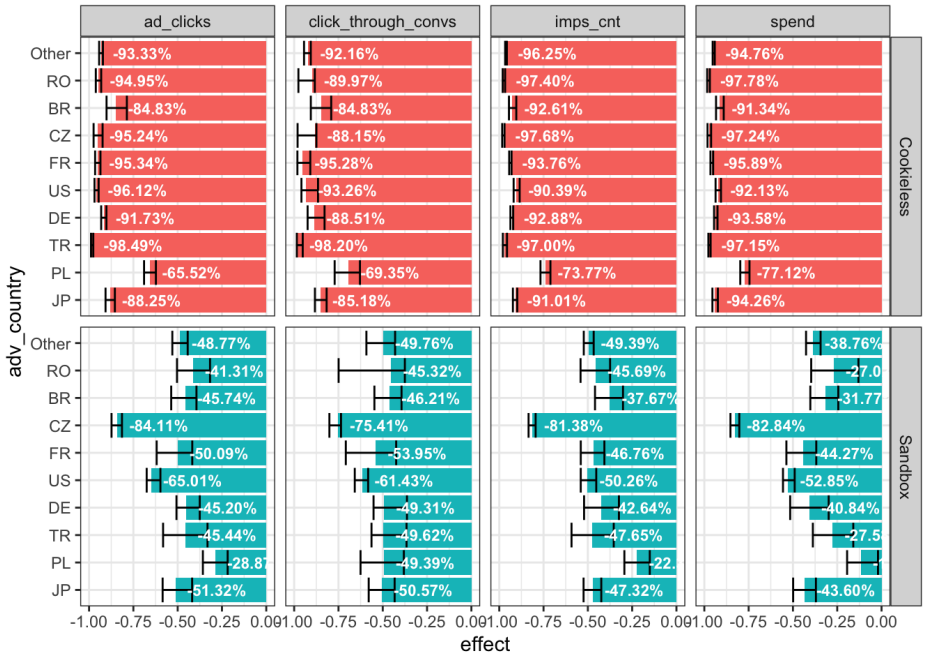


and sandbox effects are significantly impacted by these characteristics. A key finding is that the effects on ad clicks and click-through conversions are heavily influenced by how the supply side of the market adopts PAAPI and/or ID bridging.

6.3.1 Advertiser's country

Figure 6 illustrates the treatment effects by advertiser's country, ordered by the number of advertisers from each country in our dataset. Countries outside the top nine are grouped under "Other." The top nine countries are Japan (JP), Poland (PL), Turkey (TR), Germany (DE), the US, France (FR), Czech Republic (CZ), Brazil (BR), and Romania (RO). We find that, especially for the sandbox effect, there is a considerable heterogeneity across countries. In particular, Czech Republic and Poland stand out. The Czech Republic exhibits the poorest cookieless effect across all response variables. Industry insiders indicate that the Czech Republic has low adoption of Google Ad Manager (GAM) among publishers. GAM is a platform that helps publishers sell impressions through channels like AdX (Google's ad exchange). Given that the cookieless effect for Czech advertisers is around 80% for both ad performance and expenditure, this disparity may be attributed to a lack of PAAPI supply from Czech publishers, as GAM currently provides the best support for

Figure 6: ATEs by advertiser's country



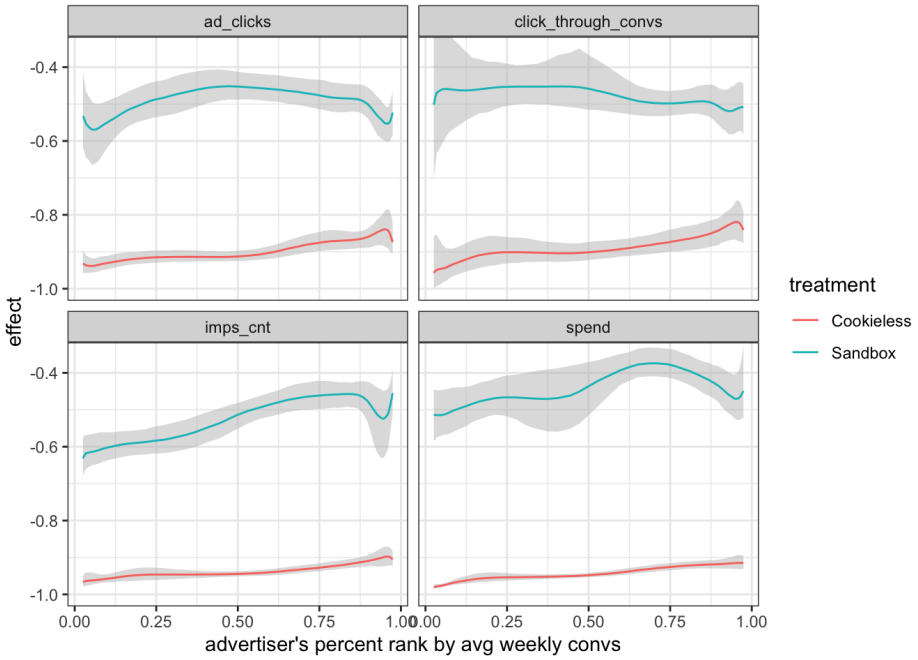
running PAAPI auctions.

Next, we find that Poland exhibits the smallest cookieless effect and sandbox effect, with the exception of click-through conversions. The disparity in cookieless effects between Poland and other countries is particularly striking. While cookieless effects in other countries average around -90%, Poland's effect is approximately 70%. This difference may be attributable to extensive ID bridging by Polish publishers. Industry insiders note that large publishers in Poland have invested heavily in ID bridging solutions. Despite the third-party cookies being disabled for users in the cookieless group, ID bridging allows Polish publishers to reintroduce old cookies, enabling our DSP to retarget these users effectively. Nevertheless, as our results in 6.1 indicate, the long-term effectiveness of ID bridging remains questionable.

6.3.2 Advertiser Size

To assess whether effects vary by advertiser prominence, we use the average weekly conversions from non-experimental users—specifically, Chrome users without experimental labels and non-Chrome users—as a proxy for the advertiser's business size.To ensure the robustness of our results to different functional form assumptions, we employ kernel regression to estimate the treatment

Figure 7: ATEs by advertiser's size



effects. In Figure 7, we illustrate how effects vary according to the percent rank (quantile) of advertisers based on the estimated measure. We find a consistent positive relationship between cookieless effects and advertiser size: smaller advertisers are more adversely affected by the absence of third-party cookies. In our context, this may be due to larger advertisers having better access to technologies that provide user information without relying on third-party cookies, such as ID bridging.

In contrast, the patterns for sandbox effects are less consistent. While larger advertisers appear to purchase more impressions from the sandbox group, the relationship between advertiser size and sandbox effects on ad clicks or click-through conversions is not monotonic. This suggests that the Privacy Sandbox may not necessarily provide a greater benefit to more prominent advertisers using the same DSP.

6.3.3 Business Type and Industry

We now explore the heterogeneous treatment effects by business type and industry. Advertisers can be categorized into four business types: classified, clickthrough, eshop, and travel. This categorization reflects how conversions are defined by the advertiser:

1. Classified: Websites where users search for houses or apartments, with conversions happening when users contact property owners.

2. Clickthrough: Website that provide price comparisons and where conversions occur when users are redirected to e-commerce sites.

3. Eshop: E-commerce websites where a conversion is a product sale; approximately 86% of the advertisers in our data fall into this category.

4. Travel: Websites selling plane tickets, with conversions being ticket sales.

Figure 8 illustrates that while cookieless effects do not show significant differences among business types, travel advertisers experience the worst sandbox effect, particularly on click-through conversions. This suggests that the loss of targeting granularity with the Privacy Sandbox may be particularly detrimental to advertisers in the travel industry.

In Figure 9, we analyze how industry type influences the treatment effects. The figure echoes the finding from Figure 8 that advertisers in the travel industry experience the worst cookieless effects. Although statistically less pronounced, we also observe that sandbox effects vary across other industry categories. For example, the real estate industry appears to perform relatively better via PAAPI.

6.3.4 Cohort

Since advertisers began implementing PAAPI at different times, we explore how the timing affects treatment effects. Around half of the advertisers started using PAAPI from the beginning of the sample collection period. Over time, our DSP adapted to more complex advertising formats used by other advertisers, which allowed them to join PAAPI. In Figure 10, while there is no clear difference in the cookieless effects, we find that the early cohort of advertisers (those who joined from week 4 to week 13) shows better sandbox effects compared to the later cohort (those who joined from week 14 to week 24). This difference could be due to two factors. First, PAAPI may be less effective for more complex ad formats due to its restrictions on the granularity of user targeting. Second, the sandbox effect may evolve over time due to experience and calibration,

Figure 8: ATEs by advertiser's business type

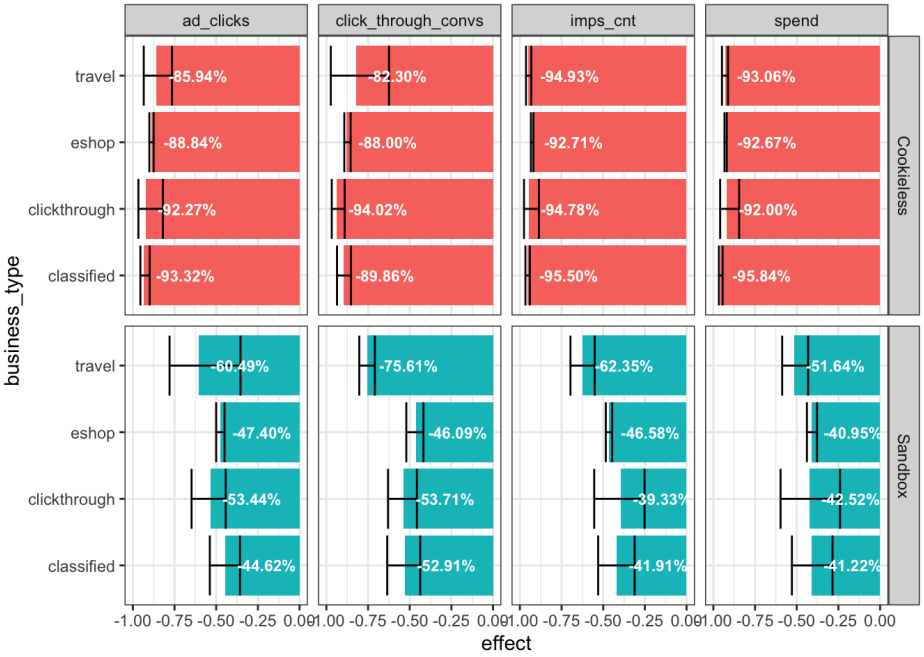


Figure 9: ATEs by advertiser's industry category

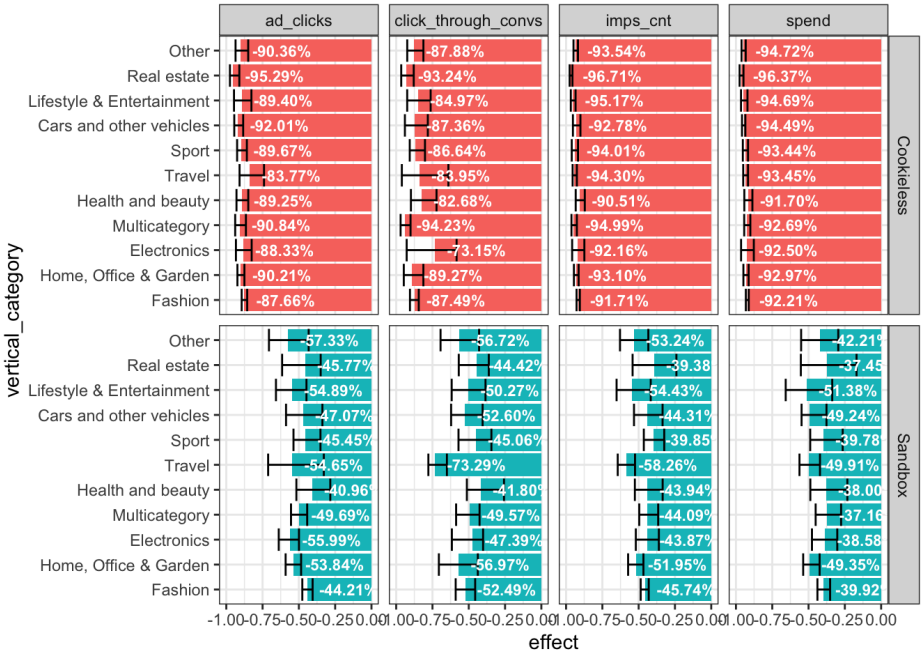
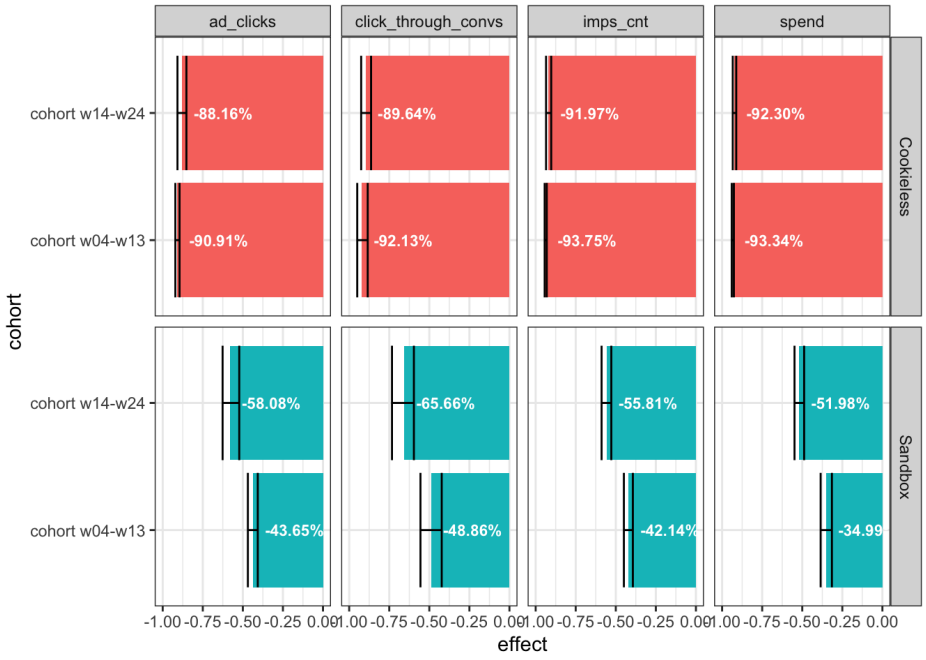


Figure 10: ATEs by advertiser's cohort

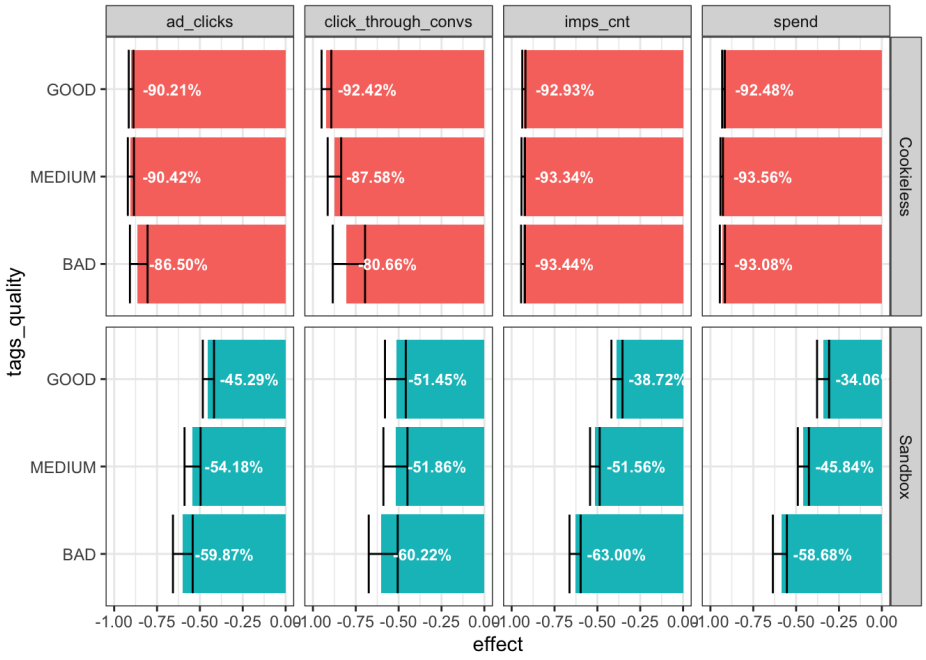


meaning the later cohort may have had less time to optimize their use of PAAPI compared to the early cohort.

6.3.5 Advertiser's Data Quality

As a robustness check and additional heterogeneity exploration, we analyze how the quality of an advertiser's ad tracking data influences treatment effects, as shown in Figure 11. Our DSP determines data quality by comparing click data collected from publisher partners with click data from advertisers, categorizing accuracy into three levels: bad, medium, and good. We find that both the cookieless effect and the sandbox effect remain relatively consistent with our main results even when focusing on data of good quality. Interestingly, we observe that the sandbox effect is worse for advertisers with poorer data quality. This suggests two possibilities: first, estimating the cookieless effect may be sensitive to measurement error, potentially resulting in substantial attenuation bias. Second, better data quality likely leads to improved calibration in PAAPI.

Figure 11: ATEs by advertiser's data quality



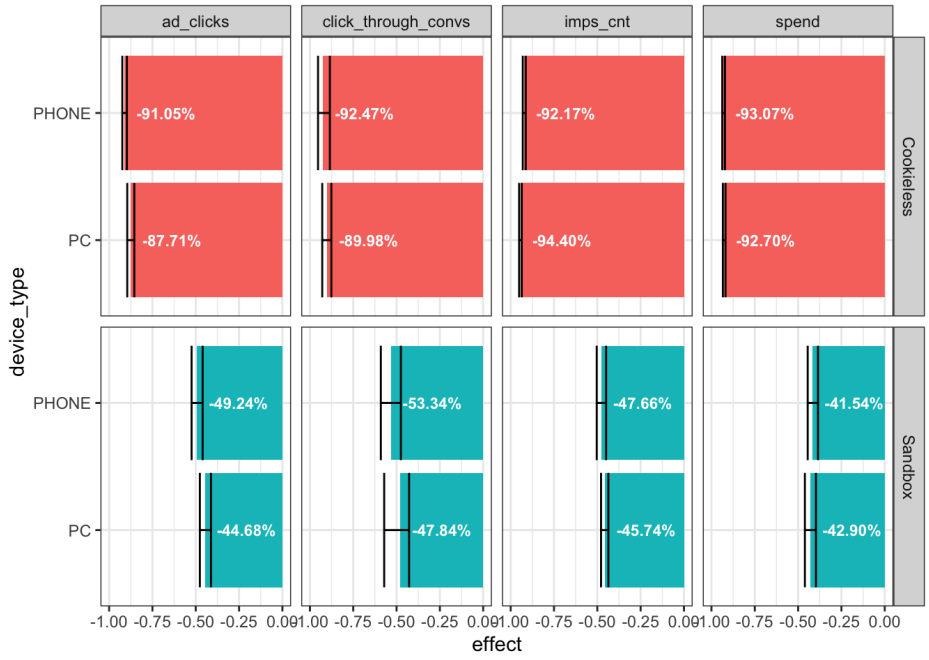
6.4 Device Heterogeneity

Subjects in this experiment may be using Chrome on either computers or smartphones (Android). We explore whether treatment effects differ across users on different devices. In Figure 12, we find that, while not statistically significant, the cookieless effects and sandbox effects on ad clicks and click-through conversions are generally worse for smartphone users. The difference in the sandbox effects may be due to the fact that PAAPI auctions are conducted on-device, and smartphones typically have less computational capacity compared to computers. Indeed, both industry experts and our parallel research on the supply side of this experiment indicate that smartphones experience higher latency in PAAPI auctions.

7 Conclusion

Our research contributes new evidence on the effectiveness of retargeting and privacy-centric online advertising. Retargeting's ad effectiveness is controversial as it targets favorably selected users and may provoke reactance in users. We provide intent-to-treat estimates across more than 2,000 advertisers that retargeting increases advertiser conversions by 4.6%. This contributes to

Figure 12: ATEs by device



the growing literature that provides broad-based experimental evidence of ad effectiveness (e.g., Lodish et al., 1995; Gordon et al., 2023a; Johnson et al., 2017a) as well as ad effectiveness estimates that do select for advertisers that run experiments (Shapiro et al., 2021; Wernerfelt et al., 2022).

We also contribute to ongoing efforts to evaluate new approaches to online advertising that leverage privacy-enhancing technologies (Index Exchange, 2024; Google, 2024; Parsons, 2024). We show that retargeting in Privacy Sandbox retains 46.3% of clicks and 43.5% of click-through conversions relative to traditional, cookie-based retargeting. This performance gap is attributable in part to the lower supply of impressions that were sold using Privacy Sandbox. Adjusted for ad spending, Privacy Sandbox retains 86.4% of clicks and 81.8% of click-through conversions. On the publisher side, the Google-CMA experiment found that Privacy Sandbox had a negligible impact on ad revenue, though low industry participation contributes to this finding. Our advertiser-side findings suggests a much smaller performance gap from Privacy Sandbox for advertisers relative to publishers. This suggests that more advertisers may use Privacy Sandbox in the long run--if cookies are reduced or eliminated--thereby improving publisher revenue.

Our research has several limitations. Many practical challenges arose in this ambitious, industry-wide experiment including experimental contamination and selection problems with observing

treatment group labels. Ad performance is difficult to compare across treatment conditions because these exhibit different market conditions and technological differences in performance measurement. Our findings may be difficult to extrapolate to a long-run equilibrium where Chrome reduces or eliminates third-party cookies: market conditions will change, Chrome and adtech firms will continue innovating on Privacy Sandbox, and increased scale poses new challenges for Privacy Sandbox performance. Our research focuses on retargeting, which translates well to Privacy Sandbox. Future research can examine how ad performance changes when advertisers must adapt tactics that translate less well to Privacy Sandbox.

References

Alcobendas, M., Kobayashi, S., Shi, K., & Shum, M. (2023). The impact of privacy measures on online advertising markets. Available at SSRN 3782889.

Aridor, G., Che, Y.-K., Hollenbeck, B., Kaiser, M., & McCarthy, D. (2024). Evaluating the impact of privacy regulation on e-commerce firms: Evidence from apple's app tracking transparency. Available at SSRN.

Aridor, G., Che, Y.-K., & Salz, T. (2023). The effect of privacy regulation on the data industry: empirical evidence from GDPR. *The RAND Journal of Economics*, 54(4), 695–730.

Bleier, A. & Eisenbeiss, M. (2015). Personalized online advertising effectiveness: The interplay of what, when, and where. *Marketing Science*, 34(5), 669–688.

Budak, C., Goel, S., Rao, J., & Zervas, G. (2016). Understanding emerging threats to online advertising. In *Proceedings of the 2016 ACM Conference on Economics and Computation* (pp. 561–578).: ACM.

Cameron, A. C. & Trivedi, P. K. (2013). *Regression analysis of count data*. Number 53. Cambridge university press.

Chen, J. & Roth, J. (2024). Logs with zeros? some problems and solutions. *The Quarterly Journal of Economics*, 139(2), 891–936.

Competition and Markets Authority (2020). *Online platforms and digital advertising market study*. Technical report, Competition and Markets Authority.

Cooper, D. A., Yalcin, T., Nistor, C., Macrini, M., & Pehlivan, E. (2023). Privacy considerations for online advertising: A stakeholder's perspective to programmatic advertising. *Journal of Consumer Marketing*, 40(2), 235–247.

Geradin, D., Katsifis, D., & Karanikioti, T. (2021). Google as a de facto privacy regulator: analysing the privacy sandbox from an antitrust perspective. *European Competition Journal*, 17(3), 617–681.

Goldberg, S., Johnson, G. A., & Shriver, S. K. (2024). Regulating privacy online: An economic evaluation of the GDPR. *American Economic Journal: Economic Policy*, 16(1), 325–58.

Google (2024). *Testing privacy preserving signals in the absence of third-party cookies on Google's display ads platforms*. Technical report, Google.

Gordon, B. R., Moakler, R., & Zettelmeyer, F. (2023a). Close enough? a large-scale exploration of non-experimental approaches to advertising measurement. *Marketing Science*, 42(4), 768–793.

Gordon, B. R., Moakler, R., & Zettelmeyer, F. (2023b). Predictive incrementality by experimentation (pie) for ad measurement. *arXiv preprint arXiv:2304.06828*.

Höring, F. (2022). An update on fledge chrome testing. Criteo R&D Blog.

Index Exchange product team (2024). Insights from Privacy Sandbox testing. Index Exchange Blog.

Jerath, K. & Miller, K. M. (2023). Using the dual-privacy framework to understand consumers' perceived privacy violations under different firm practices in online advertising. Available at SSRN.

Johnson, G. A. (2023). Inferno: A guide to field experiments in online display advertising. *Journal of Economics & Management Strategy*, 32(3), 469–490.

Johnson, G. A., Lewis, R. A., & Nubbemeyer, E. (2017a). The online display ad effectiveness funnel & carryover: Lessons from 432 field experiments. Available at SSRN 2701578.

Johnson, G. A., Lewis, R. A., & Nubbemeyer, E. I. (2017b). Ghost ads: Improving the economics of measuring online ad effectiveness. *Journal of Marketing Research*, 54(6), 867–884.

Johnson, G. A. & Neumann, N. (2024). The advent of privacy-centric digital advertising: Tracing privacy-enhancing technology adoption.

Lachner, F., Chen, Y., & Olsauskas-Warren, T. (2023). User attitudes towards controls for ad interests estimated on-device by the browser. In *Symposium on Usable Security and Privacy (USEC)*.

Lambrecht, A. & Tucker, C. (2013). When does retargeting work? Information specificity in online advertising. *Journal of Marketing Research*, 50(5), 561–576.

Lewis, R. A. & Rao, J. M. (2015). The unfavorable economics of measuring the returns to advertising. *Quarterly Journal of Economics*, 130(4), 1941–1973.

Li, J., Luo, X., Lu, X., & Moriguchi, T. (2021). The double-edged effects of e-commerce cart retargeting: does retargeting too early backfire? *Journal of Marketing*, 85(4), 123–140.

Lin, T. & Misra, S. (2022). Frontiers: the identity fragmentation bias. *Marketing Science*, 41(3), 433–440.

Lodish, L., Abraham, M., Kalmenson, S., Livelsberger, J., Lubetkin, B., Richardson, B., & Stevens, M. (1995). How TV advertising works: A meta-analysis of 389 real world split cable TV advertising experiments. *Journal of Marketing Research*, 32(2), 125–139.

Nottingham, M. (2021). Playing fair in the privacy sandbox: Competition, privacy and interoperability standards. Available at SSRN.

Olejnik, L. (2023). *Handbook on the Politics and Governance of Big Data and Artificial Intelligence*, chapter On the governance of privacy-preserving systems for the web: should Privacy Sandbox be governed? Elgar Handbooks in Political Science. Edward Elgar Publishing.

Olejnik, L. (2024). *Reconciling Privacy Sandbox initiatives with EU data protection laws*. PhD thesis, University of Edinburgh Law School.

Parsons, T. (2024). Privacy Sandbox testing results show shortfalls to meet CMA requirements. Criteo Blog.

Rumiński, M., Iwańczak, P., & Włodarczyk, Ł. (2022). Findings from the early fledge experiments. *Available at SSRN 4219796*.

Sahni, N. S., Narayanan, S., & Kalyanam, K. (2019). An experimental investigation of the effects of retargeted advertising: The role of frequency and timing. *Journal of Marketing Research*, 56(3), 401–418.

Selman, E. (2024). Criteo's Privacy Sandbox market testing. Criteo R&D Blog.

Shapiro, B. T., Hitsch, G. J., & Tuchman, A. E. (2021). TV advertising effectiveness and profitability: Generalizable results from 288 brands. *Econometrica*, 89(4), 1855–1879.

Wernerfelt, N., Tuchman, A., Shapiro, B., & Moakler, R. (2022). Estimating the value of offsite data to advertisers on meta. Available at SSRN: https://ssrn.com/abstract=4176208.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.