Measuring Risks to Users' Health Privacy Posed by Third-Party Web Tracking and Targeted Advertising

Eric Zeng
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
ericzeng@cmu.edu

Xiaoyuan Wu
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
wxyowen@cmu.edu

Emily N. Ertmann
Perelman School of Medicine
University of Pennsylvania
Philadelphia, Pennsylvania, USA
Emily.Ertmann@pennmedicine.upenn.edu

Lily Huang
Perelman School of Medicine
University of Pennsylvania
Philadelphia, Pennsylvania, USA
Lily.Huang@pennmedicine.upenn.edu

Danielle F. Johnson
Perelman School of Medicine
University of Pennsylvania
Philadelphia, Pennsylvania, USA
Danielle.Johnson1@pennmedicine.upenn.edu

Anusha T. Mehendale
Perelman School of Medicine
University of Pennsylvania
Philadelphia, Pennsylvania, USA
anushamehendale@gmail.com

Brandon T. Tang
Perelman School of Medicine
University of Pennsylvania
Philadelphia, Pennsylvania, USA
brandon.tang@pennmedicine.upenn.edu

Karolina Zhukoff
Perelman School of Medicine
University of Pennsylvania
Philadelphia, Pennsylvania, USA
karolinazhukoff@gmail.com

Michael Adjei-Poku
Perelman School of Medicine
University of Pennsylvania
Philadelphia, Pennsylvania, USA
Michael.Adjei-Poku@pennmedicine.upenn.edu

Lujo Bauer
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
lbauer@cmu.edu

Ari B. Friedman
Perelman School of Medicine
University of Pennsylvania
Philadelphia, Pennsylvania, USA
Ari.Friedman@pennmedicine.upenn.edu

Matthew S. McCoy
Perelman School of Medicine
University of Pennsylvania
Philadelphia, Pennsylvania, USA
mmcco@pennmedicine.upenn.edu

Abstract

Online advertising platforms may be able to infer privacy-sensitive information about people, such as their health conditions. This could lead to harms like exposure to predatory targeted advertising or unwanted disclosure of health conditions to employers or insurers. In this work, we experimentally evaluate whether online advertisers target people with health conditions. We collected the browsing histories of people with and without health conditions. We crawled their histories to simulate their browsing profiles and collected the ads that were served to them. Then, we compared the content of the ads between groups. We observed that the profiles of people who visited more health-related web pages received more health-related ads. 49.5% of health-related ads used deceptive advertising techniques. Our findings suggest that new privacy regulations and enforcement measures are needed to protect people's health privacy from online tracking and advertising platforms.

CCS Concepts

• Security and privacy → Social aspects of security and privacy; **• Information systems → Online advertising**; **• Social and professional topics → Patient privacy**.

CHI '25, April 26-May 1, 2025, Yokohama, Japan

Keywords

Health privacy, targeted advertising, web tracking, deceptive advertising, deceptive patterns

ACM Reference Format:
Eric Zeng, Xiaoyuan Wu, Emily N. Ertmann, Lily Huang, Danielle F. Johnson, Anusha T. Mehendale, Brandon T. Tang, Karolina Zhukoff, Michael Adjei-Poku, Lujo Bauer, Ari B. Friedman, and Matthew S. McCoy. 2025. Measuring Risks to Users' Health Privacy Posed by Third-Party Web Tracking and Targeted Advertising. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26-May 1, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 26 pages. https://doi.org/10.1145/3706598.3714318

1 Introduction

Health privacy is a core value of healthcare. The confidentiality of health information between doctors and patients is a cornerstone of medical ethics [9]. Unwanted disclosure of health information can lead to subjective privacy harms such as shame and embarrassment [15, 55], social stigma [74], and discrimination [59]. In the United States, the confidentiality of patient information is legally protected by the HIPAA Privacy Rule [68].

However, the ecosystem of online tracking and advertising threatens health privacy. Entities like online services, ad networks, and social media platforms track users' online activities to enable advertisers to precisely target users. This data about users, which includes visits to specific pages and aggregate patterns of browsing behavior, can reveal users' interests, demographics, and habits, including sensitive information such as gender, age [45], ethnicity [65], sexual orientation [7], and health conditions [73].

Collecting data on online activities that could be used to infer people's health conditions represents a fundamental privacy risk. Moreover, collection of health-related data exposes people to an array of downstream harms: (i) Such information could be shared with third-parties such as data brokers [43], who in turn could make that information available to pharmaceutical companies [39], insurance companies [4], and other entities that could profit from knowledge of individuals' health status. (ii) People could be targeted by advertising relevant to their health conditions. Although this could potentially help people become aware of treatments for their condition, this could also harm people if they are targeted with misleading or fraudulent health products, or if ads targeting their condition reinforce stigmas or anxieties (e.g., weight loss ads) [30]. Even targeted advertising for legitimate, FDA-approved medications could be harmful; direct-to-consumer marketing has been found to promote overutilization of drugs and contribute to rising drug costs [56].

Our prior work provides evidence that online trackers can observe users' visits to sensitive health-related sites. Online tracking is prevalent on hospital websites [29], health information websites [33], COVID-19-related websites [51], and reproductive healthcare websites [28]. However, to the best of our knowledge, prior work has not shown whether real users with health conditions visit more health-related websites than other users. Additionally, prior work has not shown whether advertising platforms are able to accurately infer users' health conditions from their browsing behavior, nor whether users are targeted by advertisers using these inferences in practice.

In this work, we bridge these key knowledge gaps by experimentally measuring whether people's health conditions, browsing histories, and exposure to tracking lead to targeted health-related advertising. We also quantify the deceptive practices used in health-related advertising. We investigate the following research questions:

- **RQ1:** How much health-related web browsing do users with health conditions engage in?
- **RQ2:** How prevalent are web trackers on health-related web pages in users' web histories?
- **RQ3:** Are users with health conditions targeted with ads relevant to their health conditions?
- **RQ4:** What are the characteristics of deceptive health-related advertising on the web?

To answer these questions, we conducted a measurement study and experiment using web crawlers and web histories collected from real users with health conditions. We collected a dataset of users' 90-day browsing histories, sampling from a population of people with health conditions—such as heart disease, cancer, diabetes, and irritable bowel syndrome—and from a population of people without health conditions. We created simulations of these users' browsing profiles by replaying their browsing histories using web crawlers.

Using these browsing profiles, we conducted an experiment to investigate whether the profiles of users with health conditions are targeted with health-related ads. In this experiment, we investigated *behavioral targeting*—ad targeting based on tracking users' browsing behaviors, rather than *contextual targeting*—ads that are relevant to the content of the site the ad appears on. We used web crawlers initialized with the browsing profiles of our participants, and scraped ads from a sample of 400 web pages, which we kept constant across participants to control for contextual targeting. We conducted statistical analyses to determine whether participants' health status, demographics, and browsing behaviors influenced the amount of health-related advertising served to their profiles. Furthermore, we qualitatively coded the health-related ads that we collected and analyzed which types of products, health conditions, and advertising platforms are associated with deceptive advertising.

Based on our measurements and experiment, we find that:

- Profiles of participants who browsed more health-related pages were served more health-related ads in our experiment: each 100 health-related pages visited increased the number of health-related ads observed by 2.3. This provides initial evidence that advertisers track users' health-related browsing to target health-related advertising.
- Deceptive advertising techniques are common in online health-related advertising: 49.5% of health-related ads that we collected used deceptive advertising techniques, such as overstating the benefits of treatments and citing undocumented testimonials. In some subcategories, like weight loss supplements, nearly all ads employed deceptive techniques.
- Third-party web trackers are prevalent on the health-related pages in users' histories, with over 70% of health-related pages containing trackers. Trackers from companies like Google and Microsoft are present on most health-related sites, indicating that these platforms are in a position to make inferences about users' health statuses.
- 3.1% of participants' web histories consisted of health-related pages, in the 90-day sample of our dataset. We found no significant differences between people with and without health conditions, suggesting that health information seeking about specific conditions is done in a limited timeframe that is not easily captured in research studies.

Our results provide some of the first concrete evidence that advertisers track users' health-related browsing to target them with health-related ads, demonstrating the risks hypothesized by prior work on web tracking. Furthermore, our measurements show that deceptive health advertising is highly prevalent in web advertising, which poses material risks to consumers.

To enable our study, we extended *Adscraper*[1], a crawler for collecting ad content, into a scalable measurement platform for experimentally detecting targeted advertising on the web. With our enhancements, Adscraper can now be used to conduct parallel, distributed web crawls with hundreds of browsing profiles and to collect data like ad landing pages without biasing profiles. These technical contributions enable large-scale future research on targeted advertising on the web.

2 Background

In this section, we provide a brief overview of how privacy issues surrounding online tracking and advertising intersect with issues in health privacy and policy.

[1]https://github.com/UWCSESecurityLab/adscraper

2.1 Health Privacy and Online Tracking

The HIPAA Privacy Rule protects patients' health information from unauthorized disclosure from healthcare providers and covered entities like health insurance companies [59]; it does not prevent non-covered entities from learning and disclosing that information. The ubiquity of online tracking raises the possibility that non-covered entities like search engines, online advertising platforms, and social media companies could learn about users' health status through their online behaviors.

When users visit web pages related to their health, such as a hospital website's page on their cancer treatment facilities, or a WebMD page about a medication they take, third-party web trackers like analytics scripts, tracking pixels, and online advertisements [62], may record visits to those pages, providing technology companies with potential insights into users' health status. We previously conducted measurement studies showing that trackers are widespread on health-related websites: we found that 98.6% of hospital websites contained third-party web trackers [29], and that trackers were similarly prevalent on medical journal websites [33], COVID-19-related web pages [51] and reproductive healthcare clinic websites [28].

Some reports suggest that ad platforms and data brokers attempt to identify users' health conditions. The Markup obtained a large database of audience segments (user attributes that advertisers can use to target ads) from Xandr, Microsoft's ad platform [43]. They found that data suppliers on Xandr created segments for medical diagnoses such as diabetes, depression, and liver disease, and purchases of products like contraceptives and pregnancy tests. However, it remains unclear how ad platforms and data brokers make these classifications, and whether they are able to accurately determine if users belong to these segments.

2.2 Risks Posed by Health-Related Advertising

Online tracking of health status could enable advertisers to better target health-related advertising. Health-related advertising already poses unique risks to consumers: the American Medical Association argues that direct-to-consumer marketing of medications increases consumer demand for inaccessibly priced or non-indicated medical products, reducing access and increasing costs [8]. Furthermore, some advertisements promote fraudulent, ineffective, or even dangerous medical products: common health scams include fake addiction treatments, cures for dementia, cancer, and arthritis, anti-aging products, and treatments for chronic pain [19]. Advertising targeting certain medical conditions, such as weight loss medication ads that target people with obesity, can harm users by reinforcing low self-esteem and deepening anxieties about health [30].

With more effective targeting, advertisers could more easily and efficiently reach consumers with health conditions; potentially exposing them to more fraudulent health-related advertising, or contributing to the overutilization of certain medications [56]. There is some indirect evidence that targeting of individuals already occurs; pharmaceutical marketing companies claim to have techniques to identify and target both patients and healthcare providers [39].

3 Related Work

In this section, we discuss related work on auditing online tracking and targeted advertising, and how our work builds on this literature.

3.1 Online Tracking and Targeting of Sensitive User Characteristics

A growing body of work is investigating how vulnerable or marginalized people may be harmed by tracking and targeted advertising, by inferring sensitive characteristics like gender, ethnicity, sexual orientation, and age.

People generally find online tracking and targeted advertising to be creepy, and desire greater transparency [61, 67, 70]. People perceive these harms to be particularly severe when advertisers utilize sensitive characteristics. For example, in a case study of queer people, participants targeted by ads about their identities felt unsettled and tokenized [63]. Plane et al. found that most people perceive ad targeting on the basis of race to be problematic [58].

Empirical studies have found evidence that ad targeting harms marginalized and vulnerable users. Ali et al. found that Facebook's ad delivery algorithm unequally served housing and employment ads to people of different genders and ethnicities [3], and problematic advertising on Facebook was more likely to be shown to older people and minority groups [2]. Moti et al. found that children's websites served ads with topics that were inappropriate for children, such as mental health, dating, weight loss, and racy content [52].

Users have little control or recourse in preventing targeted advertising that they might find distressing: a study of Facebook's ad controls showed that disabling health-related targeted ads only resulted in a temporary reduction [27].

Our work adds to this literature by empirically measuring the extent to which people with health conditions are targeted by health-related ads.

3.2 Measurements of Third-Party Web Tracking

Web trackers are third-party resources embedded in web pages (such as images, scripts, and iframes) that third-parties use to track users' visits to pages across the web. Websites embed trackers because they provide functionality or data for the site owner: e.g. analytics scripts, advertising tools, or social media widgets.

There is an extensive literature measuring the overall prevalence of web trackers [16, 26, 42, 49, 62, 64]. These studies typically measure tracking on a large scale by crawling lists of top websites such as Tranco [46], or by analyzing large datasets like Common Crawl or aggregated anonymous data collected from users.

Our measurements of third-party tracking differs in scope; rather than measuring tracking across the whole web, we measure individual users' exposure to tracking in their browsing history, to understand how much of a person's health-related browsing can be observed by trackers in practice. A few other studies also examine tracking in the context of users' histories: Olejnik et al. collected users' web histories and found trackers could uniquely identify 70% of users after observing 500 pages in their history [54]. Dambra et al. measured third-party trackers observed by 250K real users, finding that only 20% of users' histories are not tracked [22].

3.3 Measurements of Targeted Advertising on the Web

Some studies have sought to empirically measure targeted advertising directly. This is a challenging task because there is little transparency into how ads are targeted, and targeting often must be inferred through statistical methods.

One method for detecting targeted advertising is to create artificial experiments and profiles to test specific hypotheses about targeting. Barford et al. constructed artificial browsing profiles based on advertising topics, collected ads through web crawls, and looked for correlations between profile topics and ad topics [11]. Lecuyer et al. developed an automatic method for statistical hypothesis generation for targeting detection, to determine if individual emails, searches, or website visits caused targeting [47]. Iqbal et al. provided evidence for cross-device targeted advertising, experimentally showing that interacting with skills on an Amazon Echo affected the ad auctions for web ads served to the same user [38].

Another method for detecting targeted advertising is to collect data directly from users in a field study, and look for differences across individuals or characteristics. Iordanou et al. developed a privacy-preserving, crowdsourced method for detecting targeted ads, and observed correlations with socio-economic factors [37]. Zeng et al. conducted a field study of behavioral targeting, using an experimental design that controlled for contextual targeting, and found evidence for demographic-based targeting [78].

Targeting can also be identified using targeting explanations provided by advertising platforms. Major platforms like Google and Meta provide users with an interface showing why they were targeted by ads. Prior work has collected data from these explanations to identify targeted ads [48, 69], but has also found that targeting explanations are often incomplete [5, 25].

Our work combines field and experimental methodologies: we conducted a controlled experiment to detect differences in ads between profiles. In contrast to other experiments [11], our profiles are based on real users' browsing histories. And while other studies analyze targeting on broad demographic factors [37, 78], we focus specifically on targeting of people with health conditions.

3.4 Deceptive Online Advertising

Online advertisements can use deceptive advertising techniques to mislead and defraud users. Techniques like phishing, overstating claims and threats, and clickbait have been observed in influencer VPN ads [1], software download ads [53], and political advertising [79]. Prior work has found that native advertising networks are disproportionately responsible for serving low quality, misleading ads on the web [12, 76]. In addition to exposing people to material harms, deceptive advertising also negatively impacts user experience [77].

Deceptive advertising for medications and health products are a longstanding issue. Some prior work has investigated the types of deceptive health-related online advertising, such as analyses of FDA warning letters to online advertisers [44], and case studies of misleading ketamine [21] and cardiovascular dietary supplement ads [13]. However, no work has measured deceptive health-related online advertising at scale.

Our work contributes to this literature by providing both a qualitative view of the types of products and deceptive techniques used in health-related ads on the web, and a quantitative view of the prevalence of deceptive health-related advertising.

4 Methodology

In this section, we describe our methodology for measuring users' health-related browsing, their exposure to web tracking, and for detecting if users were targeted by health-related advertising. Figure 1 shows an overview of our study design. In summary:

(1) We collected the web histories of 107 participants, 73 of whom had health conditions, and 34 who did not (Section 4.1).
(2) We crawled each participant's web history to create a simulation of their browsing profiles. During crawling, we collected the content of each page and third-party trackers (Section 4.2.1).
(3) We scraped online ads and landing pages using crawler instances initialized with each of the 107 browsing profiles. We collected ads from a set of 400 URLs, held constant across profiles to isolate the effect of behavioral targeting (Section 4.2.2-4.2.3).
(4) We automatically labeled health-related web pages in users' browsing histories, and health-related ads based on their landing pages (Section 4.4). We qualitatively coded the subset of health-related ads with attributes like the health condition addressed by the ad and deceptive advertising techniques used (Section 4.5).

Using this data, we tested the hypothesis that participants with health conditions do more health-related web browsing, and as a result their profiles will be served more health-related ads.

To enable this experiment, we built a measurement platform for conducting large-scale, stateful web crawls. We describe the challenges and implementation details in Section 4.3.

4.1 Collecting Web Histories

We contracted YouGov, a private survey firm, to collect a deidentified dataset of real users' web histories and health conditions. We asked YouGov to recruit a mix of participants with and without health conditions. In Table 1, we summarize the demographics of our participants.

For each participant, YouGov collected their demographics (age, gender, ethnicity) as well as their health status. Participants self-reported whether they had been diagnosed with any health conditions from a list of 35 conditions that we provided, such as heart attack, dementia, diabetes, and cancer. The list of health conditions was derived from the Charlson Comorbidity Index, a metric of the burden of diseases [18]. We selected the conditions that most impacted mortality risk, such as heart attacks and dementia. We augmented the list with other conditions that are known to be privacy-sensitive or stigmatized, such as COVID [74], sexually transmitted infections [36], and psoriasis [50]. The list of conditions is shown in Table 1.

YouGov also collected each participant's browsing history over the span of up to 90 days. Browsing histories consisted of URLs collected from participants' desktop web browsers via browser extension, and participants' cross-device YouTube watch history.

The dataset was collected in July 2022. It contained the histories of 139 participants, for a total of 1,489,994 URLs. We filtered out participants with fewer than 100 URLs in their dataset, leaving 107 participants — we refer to this subset for the remainder of the paper.

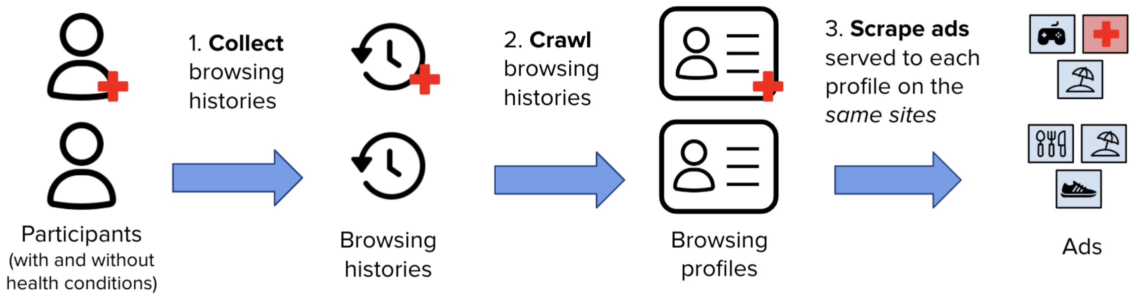

Figure 1: Overview of our experiment design and data collection methodology.

We also filtered out the frequently occurring URLs that require logins, as these cause web crawlers to perform a redirect. To do this, we manually examined the top domains in our dataset, and identified domains, subdomains, or URL path patterns that required authentication. For example, on social media platforms, we removed URLs for news feeds and settings pages, but preserved URLs to user profiles and posts. 29.9% of URLs the in initial dataset were removed. The top domains of removed URLs were: mail.google.com (15% of the removed URLs), outlook.live.com (14%), mail.yahoo.com (9%), facebook.com (8%), and twitter.com (6%).

4.2 Crawling Methodology

Next, we conducted a series of web crawls to build simulated browsing profiles and collect the ads served to those profiles. In this section, we describe the high-level goals and methodology for the three crawl phases (Sections 4.2.1 to 4.2.3). In the next section (Section 4.3) we will describe the implementation of the measurement platform we built to enable these web crawls.

4.2.1 Web History Crawl. First, we crawled the web histories of each participant to create *browsing profiles*, collect the content of the pages in participants' histories, and measure how much of participants' health-related browsing was exposed to third-party web trackers. For each participant's history, we crawled each URL in the order in which the participant browsed them. For each webpage, we scraped the HTML content, which we used later to classify whether the page was health-related. We also collected all network requests made by each page, which we used to detect third-party tracking. In total, we crawled 709,818 pages.

Each crawl of a participant's history was conducted using a separate browsing profile. Profiles are the persistent state stored by the browser (e.g. cookies, local storage). They can be used by third-party trackers to uniquely identify users across sites [62]. By crawling each site in a participant's web history, the browsing profile of the crawler should resemble the participant's actual browsing profile, meaning that types of ads served to the crawler should be similar to ads served to the participant. We reuse the profiles in subsequent crawls to measure whether ads shown to participants' profiles may differ depending on the pages that participants previously visited.

4.2.2 Targeted Ad Experiment Crawl. Second, we conducted a crawl to collect the ads served to each browsing profile, allowing us to compare the ads served to profiles of participants with and without health conditions. Using crawlers initialized with each of the browsing profiles created in the Web History Crawl, we crawled a sample of 400 URLs from the top websites on the web (described below). We call these pages "target pages" for the remainder of this section. On the target pages, we took screenshots of ads that appeared, and collected the URL of the landing pages.

We collected ads for each profile on the same set of 400 URLs. Holding these target pages constant across profiles allowed us to compare the ads served to each participant's profile, and attribute differences to behavioral targeting (targeting based on users' previous browsing behavior and inferred characteristics), rather than contextual targeting (targeting based on the web page that the ads appear on).

Additionally, to mitigate the effect of visiting target pages on the browsing profile, after visiting each target URL we discarded the changes that visiting that URL may have made to the profile, by reverting the version of the browsing profile prior to the target URL visit. In other words, the profiles created by the Web History Crawl were used unmodified for each target page visit.

Due to a discrepancy in crawler configuration, we were only able to collect ads using 104 of the 107 browsing profiles.

Constructing the List of Target Pages. We now describe how we created the sample of 400 URLs used in the targeted ad experiment crawl. Our goal was to pick a sample of pages from top websites that served ads, and for a subset of those pages to be health-related, to increase the chances we would observe health-related advertising due to contextual targeting. To identify these pages, we ran a pre-crawl of the top 15,000 domains in the Tranco top sites list [46]. In addition to the home page, we also crawled five randomly selected pages per domain. We filtered out sites without ads on them and non-English language sites (using fastText [40]). Then, we used a web page topic classifier on each page to identify health-related pages (we describe this classifier in detail later in Section 4.4). To create the final list of 400 URLs, we first selected all domains whose home page was classified as health-related, and one subpage from those domains (64 domains, 128 pages). We then randomly sampled 136 other domains, and randomly selected two pages from each domain, to bring the total number of URLs to 400.

4.2.3 Ad Landing Page Crawl. Lastly, we conducted a crawl to collect the landing pages of the ads in the Targeted Ad Experiment Crawl. Landing pages are the pages that browsers navigate to when a user clicks on an ad. We collect landing pages because their

Table 1: Participants' demographics and health status.

Demographic	n	%
Gender		
Female	59	55%
Male	48	45%
Age		
18-24	4	4%
25-34	17	16%
35-44	6	6%
45-55	14	13%
55-64	25	23%
65+	41	38%
Ever had a...		
Heart attack	3	3%
Stroke	5	5%
Transient ischemic attack	10	9%
Diagnosed with...		
Congestive heart failure	4	4%
Peripheral vascular disease	2	2%
Connective tissue disease	2	2%
Peptic ulcer or gastrointestinal bleeding	1	1%
Chronic kidney disease	4	4%
Cancer	7	7%
Diabetes	18	17%
Tumor	1	1%
Leukemia	1	1%
Cirrhosis	1	1%
Sexually transmitted infection	1	1%
None of the above	79	74%
Have any of...		
Erectile dysfunction	10	9%
Excessive sweating	6	6%
Psoriasis or chronic rash	9	8%
Irritable bowel syndrome	14	13%
Chronic constipation	2	2%
None of the above	72	67%
Uses walker	8	7%
Own or prescribed a hearing aid	13	12%
In the past year...		
Broken a bone	4	4%
Visited an emergency department	20	19%
Visited an urgent care clinic	21	20%
None of the above	69	64%
Ever had COVID	32	30%
Any health condition(s)	73	68%

content provides more context for classifying and qualitatively understanding the advertisement, compared to the banner ad. We conducted this as a separate crawl from the Targeted Ad Experiment Crawl to prevent landing page navigations from affecting browser profiles. In total, we collected 86,141 ads and their landing pages.

4.3 The Adscraper Measurement Platform

Next, we describe how we extended the *Adscraper* web crawler [75] into a measurement platform to enable our large-scale, crawler-based study of targeted advertising.

4.3.1 Motivation and Challenges. Measuring targeted advertising with web crawls poses a number of technical challenges. First, it is challenging to run stateful web crawls at scale. We found in our prior work that targeting results in small differences in the topics of ads served to different experimental groups (changes of <10% per topic) [78], requiring large sample sizes of participants. The experimental design for this study required us to run over 100 crawl instances in parallel, necessitating automation. Second, collecting ad content in a format that enables automatic classification is challenging. Recent work analyzing the content of ads uses screenshots of display ads, which relies either on manual labeling [38, 76], which is difficult to scale, or optical character recognition (OCR) [72, 79], which is often inaccurate, and the length of the text is often too short for good results from classifiers and topic models.

Existing web measurement tools are often unsuitable for measuring ad targeting. Earlier iterations of Adscraper implement the core functionality for scraping ad content from web pages [76, 79], but did not support stateful crawls or management of parallel crawler instances. OpenWPM [26] is one of the most popular tools for measuring third-party tracking, but does not include built-in support for collecting ad content or running parallel instances. eyeWnder, a targeting detection tool [37], requires real-time participation from users, and is specifically designed to not collect ad content, limiting studies on specific ad topics. And Sunlight [47] only implements statistical methods for targeting detection and requires ad data to be obtained separately.

4.3.2 Platform Implementation. To address these challenges, we extended the original Adscraper crawler into a distributed crawling platform to enable measurements of targeted advertising at scale. Researchers create a *high-level crawl specification* that defines tasks for individual *crawl worker* instances: the crawl list, the browsing profile, and the types of data to collect. Our *distributed crawling infrastructure* takes this specification as input, and automatically runs multiple crawl workers in parallel using a Kubernetes cluster. This platform significantly reduces the engineering and operational effort for researchers to conduct parallel, stateful web crawls for targeted advertising measurement, by only requiring researchers to specify the inputs to the crawl.

Crawl Worker Implementation. First, we describe capabilities of the crawl worker, which is based on the Adscraper implementation used in our previous work [76, 77, 79]. The crawl worker is a Node.js-based web crawler, based on Puppeteer, a browser automation library for Chromium. The crawl worker runs in a Docker container environment.

The crawl worker accepts a list of URLs as input, and visits each URL in order. On each page, it can scrape the content of the web page, and intercept third-party network requests made by subresources like scripts, iframes, and images.

The crawl worker can also scrape ads from the page. It detects ads on the page using CSS selectors from the EasyList ad-blocker filter list [20]. After detecting ads, it scrolls from top to bottom, taking screenshots of each ad.

We extended Adscraper's core ad collection capabilities by implementing the ability to identify the URL the ad links to when clicked. Importantly, this feature identifies the URL without actually visiting the page, by blocking any network requests that originate from the click. This prevents the click from registering a conversion with the ad network, which could bias the crawler's browsing profile.

We also added the ability for Adscraper to save and load browsing profiles between crawls. When crawling, browser state that can uniquely identify users (cookies, HTML5 web storage, and cache) is accumulated in Chrome's profile directory. At the end of a crawl, this directory can be copied to persistent storage and can be loaded for future crawls that use the same profile.

High-Level Crawl Specifications. We created a crawl specification format, that allows a researcher to run a crawl job by declaratively specifying crawls in JSON format. The specification consists of a set of directives for each crawl worker: the list of URLs to crawl, whether to load a browsing profile from disk before the crawl, whether to save the modified browsing profile back to disk after the crawl, and which types of data it should collect for each URL (third-party web trackers, page content, and/or ad content).

Each of the types of crawl we conducted can be defined by this specification. For example, in the web history crawl (Section 4.2.1), each crawl worker was assigned to one participant's full browsing history, it created and saved a new profile, and it collected data on page content and third-party trackers. Whereas in the targeted ad experiment crawls (Section 4.2.2), each crawl worker was assigned a single URL, it loaded an existing browsing profile but did not save it, (to mitigate the effect of visiting target sites on the profile), and it scraped ads on the page.

Distributed Crawling Architecture. The infrastructure for running distributed crawls consists of three components: a Kubernetes cluster for scheduling and running crawl workers (which can consist of any number of servers), a PostgreSQL database for storing metadata from crawls, and a network storage drive for storing scraped pages, ads, and Chromium profiles.

A crawl job is started by running the job creation script, which accepts high-level crawl specifications as input, and translates the specifications into a Kubernetes Job. The Kubernetes job scheduler launches as many parallel crawl workers as the cluster can support: each crawl worker needs at least 1.5 CPU cores and 8GB of memory. The scheduler waits for each running worker to complete, and then launches new workers until the job is complete.

4.3.3 Server and Network Configuration. We describe the specific server and network configuration used for our study. Our crawls were conducted using Google Chrome version 119.0.6045.105. We deployed our crawlers on six servers on Carnegie Mellon University's network, each with 12 CPUs and 64GB of RAM. To control for differences created by location-based ad targeting, and to mitigate IP-based bot detection, we routed all crawler traffic through a proxy server located on a local network at Carnegie Mellon behind a Network Address Translation service. All of our crawlers shared the same IP address, along with 221 other real user devices on the network at the time of the crawl (according to CMU Computing Services), masking our traffic in real user browsing. The crawls were conducted from May-June 2024, with web history crawls lasting 4 weeks and the ad targeting experiment and landing page crawls lasting 2 weeks.

4.4 Automated Classification of Topics in Browsing Histories and Ad Landing Pages

To enable content analysis of our crawled data at scale, we used a text classifier to automatically label the topics of websites and ads in our dataset. We used an off-the-shelf text classifier from uClassify [66], which classifies documents using topics from the Interactive Advertising Bureau's Content Taxonomy 2.0 [14]—a hierarchical list of website topics used by advertisers to identify the content of sites for contextual targeting. The underlying implementation is a multinomial Naive Bayesian classifier, which outputs a probability for whether a document belongs to each of the possible classes in the Content Taxonomy. Each class consists of a *main topic* (28 total), and a *sub-topic* (557 total).

This classifier differs from other commercial classifiers (e.g. CloudFlare Domain Intelligence, Google Chrome's Topics API) because it classifies the content of the page, rather than the domain or URL, which results in better classifications when the content of the page differs from the topic for the overall domain. For example, a Google search result page for a disease may be classified by uClassify with the topic "medical health", but other services may classify it as a technology or search engine site based on its domain (google.com).

We used the classifier to label topics for all 709,818 pages in the Web History Crawl (Section 4.2.1). For each page, we took the main topic and subtopic with the highest probability in the classifier's output as the labels for the page. We marked pages as *health-related* if the main topic or subtopic were health-related (e.g. "medical health", "health insurance"). In total, we classified 16,667 pages from participants' histories as health-related.

For the 86,141 ad landing pages (Section 4.2.3), we used the classifier as an initial filter to identify health-related ads, and we performed additional manual validation (see Section 4.5). Because we planned to perform manual validation afterwards, we used different threshold for the classifier that ensured high recall: we labeled ads as health-related if a health-related main topic or subtopic was in the top-n most probable classes.

To determine this top-n threshold, we created ground truth labels for a separate validation dataset ad landing pages and used these labels to evaluate the classifier. We collected a dataset of 1,367 ad landing pages by scraping ads from a random sample of top sites. Two researchers independently coded each ad, determining whether the ad promoted a product, service, information, or healthcare provider that addressed a health condition. Then, the researchers met to resolve all disagreements to reach 100% agreement. Using these ground-truth labels, we determined that the threshold that maximized recall with adequate precision was to classify ads as health-related if a health-related subtopic was in the top-5 ranked subtopics, ordered by probability. At the top-5 threshold, the recall was 0.91, precision was 0.71, and F1 score was 0.80.

Using this classifier, we provisionally classified 24,114 of 91,123 ad landing pages as health-related, prior to manual validation.

Ad Deduplication. We observed a high frequency of duplicate ads in our dataset. To reduce manual labeling, we automatically deduplicated ads with similar landing pages. We used a MinHash-based Locality Sensitive Hashing algorithm to identify ad landing pages with a Jaccard similarity > 0.75. To avoid false positives, we only ran our deduplication algorithm on groups of ad landing pages that shared a domain. This reduced the number of health-related ads we needed to qualitatively code from 24,114 to 1,419.

4.5 Qualitative Coding of Health-Related Ads

We qualitatively coded the subset of 1,419 unique health-related ads to enable deeper analysis of the content of health-related advertising. In this section we describe our qualitative coding methodology and the codebook.

4.5.1 Qualitative Coding Methodology. We used a deductive coding approach. First, we generated an initial codebook with codes in four categories: (i) whether the ad is health-related, (ii) the types of products being advertised (e.g. medications vs. hospitals), (iii) what specific health conditions were addressed by the product, and (iv) deceptive advertising practices used in the ad.

Then, we refined the codes through an iterative process. Five coders (four with medical training) independently labeled 200 ads. Then, all coders met to discuss disagreements and update the codebook to resolve ambiguities. We repeated this process until we reached a consensus among coders on the code definitions.

To assess consistency, we calculated agreement between all coders on the initial 200 ads using Krippendorff's alpha, with the MASI distance metric for categories of codes where multiple options could be selected [57]. We achieved an agreement level of $\alpha >= 0.86$, $\alpha >= 0.79$, $\alpha >= 0.72$, and $\alpha >= 0.71$ on the four categories of codes respectively.[2] Although the levels of agreement were acceptable, we decided to have two coders label each ad to ensure consistency.

For the remaining 1,219 ads, each ad was coded independently by two coders with medical training, who met to resolve disagreements. The full list of codes can be seen in Table 6, and full definitions for each code can be found in Appendix A.

4.5.2 Codes. Next, we describe the four categories of codes applied to each ad.

Health-Related. We labeled whether ads were health-related, to validate the provisional classifier label. We defined health-related to encompass ads that advertise a product, service, or information that is intended to address a person's health conditions or general wellness. If ads were not health-related, we stopped labeling here, as those ads were not relevant to our analysis. We also encountered ads that targeted healthcare professionals, such as job postings for nurses, or medical journal articles. Although these ads were related to health, they did not target consumers or patients (who our participants represent), so we excluded them from further analysis.

Product Type. We labeled the type of product being advertised. We iteratively generated codes through open coding of the first 200 ads. Our codes included FDA-approved drugs or medications, health information, medical devices, healthcare providers or facilities, skin care, dietary supplements, diet plans, health insurance, and charities. An ad could be coded with multiple product types, if multiple distinct products were advertised.

Health Condition Addressed by Ad. To enable analysis of whether an ad was targeted based on a participant's health condition, we coded the health condition(s) that each ad addressed. We created an initial set of codes based on the health conditions reported by participants in the web history dataset. Through open coding of the first 200 ads, we consolidated similar health conditions in the initial list, and added new health conditions observed in the dataset. An ad could be coded with multiple health conditions. The final list of health conditions included 35 codes.

Additionally, we merged similar health conditions into broader categories to simplify statistical analyses. We grouped health conditions into 16 body systems (e.g. respiratory, endocrine) derived from the top-level ICD-10-CM codes from the World Health Organization's medical classification list for diseases [17].

Deceptive Techniques. Lastly, to allow us to identify potential harms to users, we coded deceptive advertising techniques in health-related ads. We generated our initial codebook based on consumer advice from the Federal Trade Commission on identifying common health scams [19]. Based on this advice, we created the following codes: overstating benefits or understating costs/risks, undocumented testimonials, money-back guarantees, time-limited offers, and pseudo-science and prestigious prizes. Additionally, during the initial coding of 200 ads, we created two codes for clickbait and affiliate marketing, which are techniques identified in prior work on deceptive online advertising [12, 24, 76], but were not mentioned in the FTC's advice for health scams. Detailed descriptions of these deceptive techniques are available in Appendix A.4.

4.6 Ethics

We obtained approval from our institutional review boards (IRBs) to collect and analyze users' web history data. Participants' web histories were collected by YouGov, a private survey firm. YouGov's research and recruitment protocol was approved by Western IRB, an external commercial IRB. The IRBs at both of the researchers' institutions also approved our use of this dataset: one institution determined the research to be Exempt Category 2, and the other determined the research to be Not Human Subjects Research.

The dataset was de-identified: no identifiers like names, locations, or IP addresses were provided to us. Additionally, YouGov removed identifying or security-sensitive information from the URLs in the dataset by stripping the query parameters, such as parameters for a one-time password reset link. An allow-list of query parameters that are known to be search queries (e.g. `?q=search%20term` in a Google search) were maintained. We further protected the data by storing it in password protected, institutionally managed servers, storage, and devices, and by implementing access controls to restrict access to only members of the research team.

When conducting our web crawls we considered the impacts of our crawls on the sites we visited, and whether the impacts outweighed risks or harms. We believe our crawls have minimal impact on web servers; given the composition of the dataset and

[2] $\alpha >= 0.67$ is considered acceptable agreement for cautious conclusions, and $\alpha >= 0.8$ is considered good agreement [34].

our crawling methodology, the traffic load and volume is similar to that of 107 real users, a negligible count for most websites.

Our crawlers clicked ads and visited ad landing pages. This could potentially impact websites and advertisers, because clicking ads may cause the advertiser to be charged for the engagement, resulting in money being paid to the website and ad networks. We determined that the impacts of clicking ads were outweighed by the benefits. Accessing ad landing pages was critical for our analysis, as we needed the context from the content of landing pages to accurate classify ads into topics and qualitatively code attributes like deceptive advertising techniques.

We believe that the monetary impact to advertisers was within acceptable bounds. Based on a cost-per-click for Google Display Ads of $0.63, we estimate that the median advertiser may have been incurred a cost of $3.78. However, even this is an upper-bound estimate, as it is possible advertisers use other cost metrics like cost-per-mille that do not pay for each click, and that our clicks may have been classified as bot-driven and not counted as a real engagement.

Furthermore, we highlight that the online advertising ecosystem has long suffered from a lack of transparency, making it difficult for outside researchers to audit practices that harm consumer welfare, security, and privacy. We believe that the relatively small impact of this method to advertisers is justified by providing transparency to how online advertising may harm users. Our work is not alone in this space; other studies of malvertising [60, 71] and deceptive political advertising [79] have employed similar scraping strategies.

5 Results

We report our analyses of health-related browsing in users' web histories, web trackers observed on health-related sites that users visited, targeting of browsing profiles with health-related advertising, and the characteristics and prevalence of deceptive health-related advertising.

5.1 Health-Related Browsing in Users' Web Histories (RQ1)

We start with a description of participants' web histories, followed by an analysis of whether participants' health conditions influence the amount of health-related browsing in their histories.

5.1.1 Web history dataset overview. We crawled and scraped a total of 709,818 pages across 107 participants' web histories. The median participant's history contained 4,233 URLs (IQR 1,636–10,433), and the highest outlier contained 45,665 URLs.

The top main topics in the overall dataset were "technology and computing" (9.7%), "sports" (9.2%), "music and audio" (7.9%), "business and finance" (5.8%), and "fine art" (5.5%). Pages with health-related main topics ("medical health" and "healthy living") comprised 2.3% (n=16,667) of all pages that we crawled.

Of the pages with health-related main topics, the five most common sub topics were "weight loss" (8.1%), "vaccines" (5.9%), "medical tests" (5.7%), "health insurance" (5.5%), and "cold and flu" (5.0%). The domains with the most health-related pages were google.com (2,228 pages), youtube.com (654 pages), aol.com (515 pages), bing.com (331 pages), and healthline.com (328 pages).

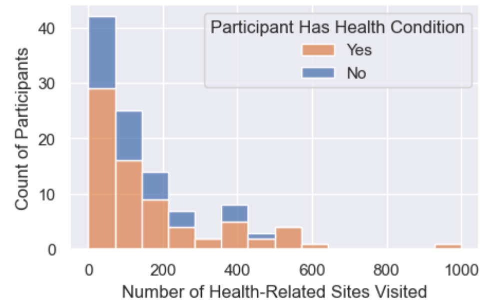

Figure 2: Histogram of the number of health-related pages visited by participants.

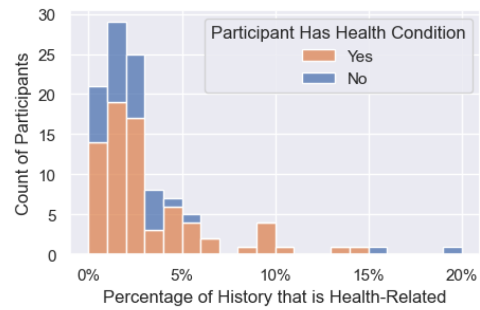

Figure 3: Histogram of the percentage of participants' web histories that were health-related.

On Google and Bing, most of the pages classified as health-related were search queries. The top topics of these queries were menopause (232), cold and flu (186), vaccines (170), weight loss (167), and foot health (153). Anecdotally, we found examples of participants performing search queries relevant to their health condition. For example, a participant with irritable bowel syndrome searched "can irritable bowel syndrome cause hemorrhoids"; a participant with diabetes searched "is soybean oil ok for diabetes type ii?", and a participant with psoriasis searched for "biologics for psoriasis".

5.1.2 Health-related browsing history. The quantity of health related browsing varied widely between participants. Figures 2 and 3 show the number and percentage of health-related pages in participants' web histories. The mean number of health-related sites in each participant's history was 155.75 (SD=171.63), and the mean percentage of health-related sites in each participant's history was 3.11% (SD=3.31%). Certain participants visited substantially more health-related pages than average. For the bottom 84% of participants, health-related pages made up less than 5% of histories. For the top 16% of participants, the percentage of their history that was health-related ranged from 5% to 19.3%. The maximum number of health-related pages visited was 999.

5.1.3 Health conditions and health-related browsing history. Table 2 shows the number of pages participants visited that were relevant to their self-reported health conditions. The sample size of health-related sites and participants with specific conditions was lower

Table 2: Mean number of pages participants visited relating to a health condition (grouped by ICD-10-CM condition codes), comparing participants with and without the condition. No differences were statistically significant, due to sample size constraints.

	Participants with Condition			Participants without Condition		
ICD-10-CM Condition	n	Pages (mean)	95% CI	n	Pages (mean)	95% CI
Circulatory	15	22.00	[-7.81, 51.81]	92	4.28	[2.29, 6.27]
Digestive	16	3.50	[0.31, 6.69]	91	1.76	[0.14, 3.37]
Ear	13	8.23	[-0.14, 16.60]	94	4.55	[1.66, 7.45]
Endocrine	23	17.43	[10.10, 24.77]	84	25.32	[16.07, 34.57]
Genitourinary	15	10.87	[-6.47, 28.21]	92	4.67	[2.93, 6.42]
Health Services	36	12.58	[4.79, 20.38]	71	25.24	[3.87, 46.61]
Musculoskeletal	12	16.08	[-12.97, 45.13]	95	7.34	[4.07, 10.60]
Neoplasms	9	2.56	[-0.73, 5.84]	98	2.57	[1.00, 4.15]
Respiratory	32	15.34	[7.92, 22.77]	75	23.59	[16.22, 30.95]
Skin	9	5.78	[-4.77, 16.32]	98	2.46	[1.28, 3.64]
Any Condition	73	166.39	[122.48, 210.32]	34	132.88	[88.04, 177.72]

than anticipated, so we did not have the statistical power to conclusively determine whether participants with health conditions visited more pages about those health conditions.

For participants with any health condition, 3.2% of pages visited were health-related (95% CI [2.5%, 3.9%]), while for people with no health conditions, 2.9% of pages visited were health-related (95% CI [1.5%, 4.2%]). This difference was not significant according to a Mann-Whitney U test. A sensitivity power analysis showed that our test was underpowered (see Appendix B.1). This suggests that the differences were either too small for us to detect at this sample size, or that there is not a meaningful difference.

We could not conclusively determine whether participants who had a specific health condition visited more pages related to that condition than other participants. For each health condition, we conducted a Mann-Whitney U test comparing the number of sites about the condition visited by participants who have that condition to participants who do not have that condition. After performing a Holm-Bonferroni correction for multiple comparisons, we did not find significant differences for any of the health conditions. Sensitivity power analyses found that these tests were underpowered (see Appendix B.1).

Additionally, we did not find a correlation between the number of health conditions a participant reported (i.e., comorbidities) and the number of health-related pages visited. Participants reported between zero and six health conditions. An analysis of variance based on negative binomial regression indicated no statistically significant effect of number of comorbidities on the number of pages visited that were health-related (χ^2(1, N=104)=117.53,*n.s.*). We report the regression outputs in Appendix B.1.

We speculate that the lack of observed differences between participants with and without health conditions reflects the limited amount of health-related browsing in the dataset. Participants' histories were only collected during a 90-day time period. However, online health information seeking may be episodic; browsing may be more common during the onset of symptoms or new developments. It is possible that the period during which data was collected did not overlap with when participants sought health information.

Finding 1: Approximately 3% of web pages visited by participants are about health topics. We could not conclusively determine whether participants with health conditions visited more pages related to their conditions than participants without conditions.

5.2 Presence of Web Trackers in Participants' Health-Related Browsing (RQ2)

In this section, we investigate how much of participants' browsing histories were observed by third-party web trackers, how much of their health-related browsing specifically was tracked, and what proportion of a participant's health-related browsing was observed by each individual tracking entity.

5.2.1 Overview of third-party web tracker data. During the web history crawl, we collected all network requests made by each page in participants' histories. We used the Ghostery Tracker Database [31] to determine which requests were initiated by third-party trackers and which business entities the trackers were associated with.

Across all participants' browsing histories, we identified 15,407,739 unique third-party trackers by the full URL of the trackers, collected from 709,818 web pages. These trackers belong to 1,425 unique domains (e.g., googlesyndication.com) and 1,322 unique tracking entities (e.g., Google, Meta, Criteo).

5.2.2 Third-party web tracking is highly prevalent in users' web histories. Third-party trackers were present in all of 107 participants' histories, and almost 70% (474,768) of the pages contain at least one third-party tracker. Table 3 shows the top owners of web trackers observed while crawling participants' web histories, ranked by number of pages on which the trackers were present. Trackers associated with Google, Microsoft, and Meta appeared in every participant's history. Trackers from Google alone appeared on over 50% of pages.

Table 3: Top web tracker entities observed in participants' histories, by number of profiles and pages in which they were observed.

Web Tracker Entity	Num. of Profiles	% of Profiles	Num. of Pages	% of Pages
Google	107	100.00	403,580	56.86
Microsoft	107	100.00	168,425	23.73
Meta	107	100.00	127,352	17.94
Amazon	106	99.07	114,783	16.17
Adobe	106	99.07	104,649	14.74
Criteo	106	99.07	101,895	14.36
The Trade Desk	105	98.13	90,936	12.81
The Rubicon Project, Ltd.	106	99.07	90,062	12.69
Index Exchange, Inc.	106	99.07	88,240	12.43
PubMatic, Inc.	106	99.07	87,882	12.38
Verizon	106	99.07	81,309	11.45
comScore, Inc.	105	98.13	80,206	11.30
OpenX Software Ltd.	106	99.07	79,560	11.21
LiveRamp	106	99.07	74,325	10.47
LiveIntent	105	98.13	74,010	10.43

5.2.3 Third party web tracking is prevalent in users' health-related browsing. We found at least one third-party tracker on 70% of health-related pages, 3 percentage points more than on non-health-related pages (66.8%). On average, health-related pages initiated 84.4 unique tracking requests, 28% more than the overall average per page (65.7).

Participants with and without health conditions were exposed to the same amount of tracking. Figure 4 shows the percentage of health-related pages in each participant's history that had a third-party tracker. On average, at least one tracker appeared on 70.7% of pages from participants with health conditions and 69.3% from those without any health conditions.

5.2.4 Health-related pages have as many trackers as non-health-related pages. We counted the number of unique tracking entities on each page from participants' histories. Figure 5 shows that a similar number of unique tracking entities show up on a given proportion of pages regardless of the page's topic. For example, about 80% of both health-related and non-health-related pages have 20 or fewer unique tracking entities on them. The average number of unique tracking entities are almost the same on health-related pages (12.7) compared to non-health-related pages (11.8).

5.2.5 Top third-party trackers can observe the majority of users' health-related browsing. We shift our focus from trackers at-large to individual trackers with high coverage. If one tracker has coverage over a substantial portion of a person's health-related browsing, then it could have the ability to draw more complex inferences over a larger dataset of browsing. To investigate these trackers, we queried for tracking domains that were able to observe more than 50% of an individual's health-related browsing history. Table 4 shows the top 15 owners of tracking domains by coverage of health-related pages in participants' browsing history.

Google appeared on more than half of the health-related pages in a participant's history for 75 of 107 participants, followed by Microsoft (17) and Adobe (10). In addition to Google, other advertising platforms such as Criteo and Adform also tracked over 50%

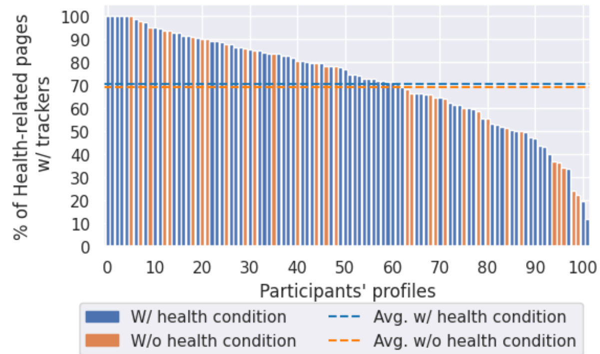

Figure 4: Percentage of health-related pages in a participant's browsing history where at least one tracker was present. Each vertical bar represents one participant.

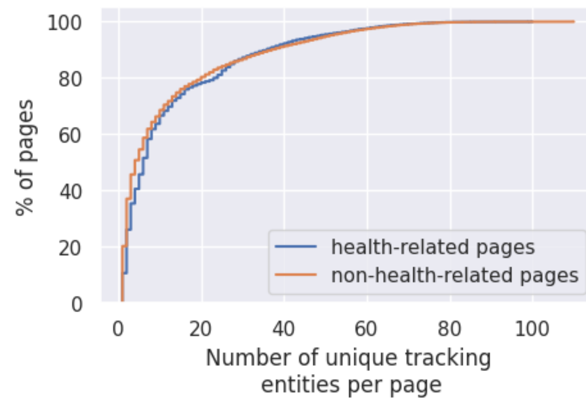

Figure 5: Cumulative distribution function of the proportion of pages having the given number of unique tracking entities, comparing health-related vs non-health-related pages. The distributions are approximately equal.

of health-related pages for 5 and 3 participants, respectively. This suggests that multiple companies, including Google, Meta, Adobe, and Microsoft, are able to collect data on users' browsing histories that would enable complex inferences about users' health status.

> **Finding 2**: Web trackers are as common on health-related sites visited by participants as other sites. Google's trackers are present on the majority of health-related sites visited by participants.

5.3 Targeting of Health-Related Ads (RQ3)

We investigate whether the simulated browsing profiles of participants with health conditions were targeted with ads relevant to their health conditions. Using data from the targeted ad experiment crawl (Section 4.2.2), we investigated behavioral targeting of health-related ads by comparing the frequency of health-related ads served to each participant's simulated browsing profile, on the

Table 4: Number of participants (out of 107) where a tracking entity is present on over 50% of the health-related pages they visited.

		Mean % of Pages Tracked	
Tracker Entity	Num. Profiles Tracked	Profiles w/Trackers on >50% of Health Sites	All Profiles
Google	75	74.39	64.23
Microsoft	17	69.80	30.46
Adobe	10	68.53	25.21
Meta	10	59.26	23.44
The Trade Desk	6	68.42	16.64
Amazon	5	55.38	17.10
Criteo	5	64.07	16.99
Index Exchange, Inc.	5	62.39	15.43
The Rubicon Project, Ltd.	5	61.77	15.53
Verizon	5	66.12	15.62
OpenX Software Ltd.	4	63.49	14.58
LiveRamp	3	58.48	13.69
Adform	3	60.51	13.42
Comscore, Inc.	3	72.23	14.76
Exponential Interactive, Inc.	3	60.59	12.04

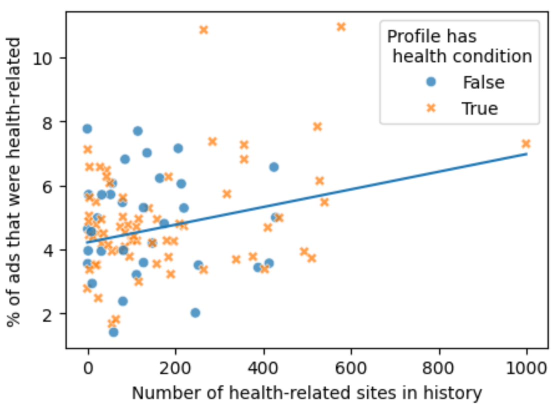

Figure 6: Scatter plot of the number of health-related sites browsed, and number of health-related ads collected for each profile. Each 100 health-related sites increases the expected number of health-related ads by 2.3.

same set of 400 web pages. We analyzed the effects of factors including participants' health conditions, the amount of health-related browsing in their history, and their demographics.

5.3.1 Overview of ads dataset. In total, we collected 86,141 ads from 104 crawls (one for each participant's browsing profile) of 400 web pages. Due to a discrepancy in crawler configuration, ads were not collected for 3 profiles. Of these ads, 4,187 ads were health-related (4.9%). Some ads appeared multiple times; after deduplicating these ads, 7,362 ads were unique, and 612 unique ads were health-related (8.3%). The mean number of ads observed per profile was 828.28 (SD=42.0), the mean number of health-related ads observed per profile was 40.3 (SD=13.5), and the mean percentage of observed ads that were health-related was 4.9% (SD=1.6%).

5.3.2 Health-related browsing increases the overall number of health-related ads seen. First, we investigate whether the overall number of health-related ads served to a profile is affected by participants' health conditions, demographics, and history. We conducted a multiple linear regression that used age, gender, the number of health-related pages in each participant's history, and the health conditions of the profile as independent variables, and the percentage of ads served that were health-related as the dependent variable. To reduce the number of variables in the regression model, we grouped participants' health conditions by body system, using the top-level ICD-10-CM codes (the World Health Organization's system for coding diseases [17]).

We found a statistically significant effect of the number of health-related sites in a participants' history on the percentage of health-related ads collected (β=0.0028, 95% CI [0.0009, 0.0046], p=0.015, adjusted R^2=0.080): every 100 health-related pages in the profile's history increased the percentage of health-related ads by 0.28%, or 2.3 ads across 400 page visits. Profiles from participants with digestive system-related conditions were served 0.91% fewer health-related ads (β =-0.919, 95% CI [-1.830,-0.008], p=0.046), but no significant effect was observed for other health conditions. Neither age nor gender had a significant effect. We report regression outputs and model selection details in Appendix B.2. Figure 6 visualizes the correlation between number of health-related sites in a participant's history and the percentage of ads received by their profile that were health-related.

5.3.3 Specific health conditions and health-related advertising. We could not conclusively determine whether profiles of participants with health conditions received more ads related to their condition, nor could we determine whether visits to pages about a health condition resulted in receiving relevant ads. Similar to our analyses in Section 5.1.3, the sample size of participants and ads for specific health conditions was lower than anticipated, resulting in low statistical power.

Table 5 shows the mean number of ads that participants' profiles received that were related to a condition the participant had. Conditions were grouped by body system based on ICD-10-CM codes. For each condition in Table 5, we conducted Mann Whitney U tests to compare differences in condition-related ads between profiles of participants with and without the condition. After applying the Holm-Bonferroni procedure, we found no significant differences for any condition (see Appendix B.2 for details).

Visiting pages about specific health conditions also did not correlate with the number of ads relating to the condition. For each health condition, we conducted a linear regression, comparing the number of pages a participant visited related to the condition with the percentage of ads their profile received related to that condition. Figure 7 visualizes the regressions on a scatter plot. We found no significant association between these variables for all health conditions. (see Appendix B.2 for details).

5.3.4 Deceptive health-related advertising. Lastly, we investigate whether the amount of deceptive health-related ads received by a profile is affected by a participant's health conditions, browsing history, or demographics. We discuss the content of deceptive ads in greater detail in Section 5.4.1.

The mean percentage of ads served to each profile that were health-related and deceptive was 2.41% (SD=1.15%). We conducted a negative binomial regression using age, gender, the number of

Table 5: Number of ads relating to a specific health condition, shown to participants who had that health condition. For each condition, we did not detect significant differences in ad count between participants with and without the condition.

	Participants with Condition			Participants without Condition		
ICD-10-CM Condition	*n*	Ads (mean)	95% CI	*n*	Ads (mean)	95% CI
Circulatory	14	4.71	[3.55, 5.88]	90	3.46	[2.99, 3.92]
Digestive	15	1.60	[1.05, 2.15]	89	2.73	[2.05, 3.41]
Ear	13	1.77	[0.06, 3.48]	91	1.09	[0.72, 1.46]
Endocrine	23	7.43	[6.06, 8.81]	81	8.43	[7.40, 9.46]
Genitourinary	14	3.43	[1.55, 5.31]	90	3.24	[2.64, 3.85]
Health Services	35	8.31	[6.23, 10.40]	69	6.87	[5.98, 7.76]
Musculoskeletal	11	2.36	[1.74, 2.98]	93	3.35	[2.78, 3.93]
Neoplasms	9	1.44	[0.42, 2.47]	95	1.86	[1.48, 2.24]
Respiratory	32	3.28	[2.02, 4.54]	72	3.46	[2.76, 4.16]
Skin	9	14.22	[-3.44, 31.88]	95	5.45	[4.46, 6.44]
Any Condition	70	40.46	[37.21, 43.70]	34	39.85	[35.20, 44.50]

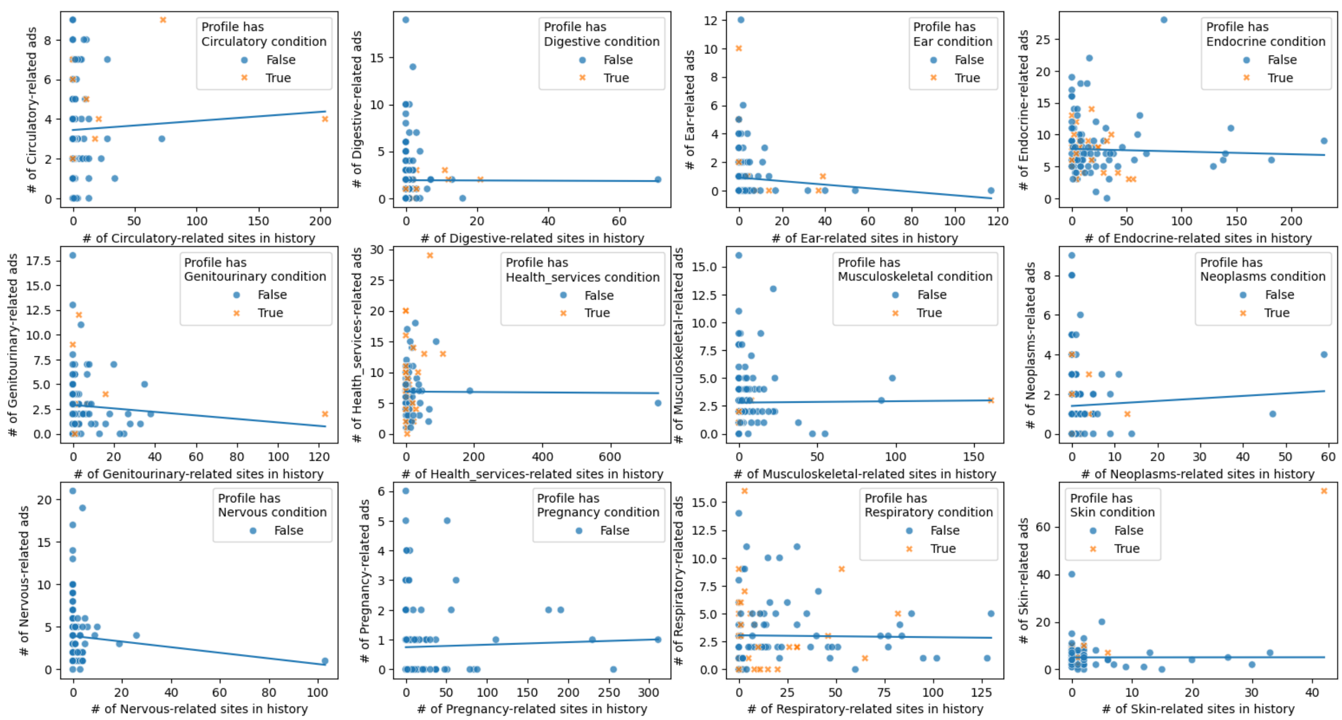

Figure 7: Scatterplots for each category of health condition, showing the number of sites a participant browsed about the health condition versus the number of ads about that condition served to their profile. We did not detect any significant associations.

health-related pages in the participant's history, and the participant's health conditions as independent variables, and the number of health-related ads labeled as deceptive as the dependent variable. We found that older participants' profiles were slightly more likely to receive deceptive ads: each one-year increase in age increased the rate of deceptive ads shown to the profile by 0.5% (p=0.043), or an absolute increase of 0.012 percentage points per year. The profiles of participants with cancer-related conditions had 42% higher rates of deceptive ads (p=0.025), an increase of 1.01 percentage points, but no other conditions had a significant effect.

Finding 3: Profiles of participants who visited more health-related pages received more health-related advertising. But we could not conclusively determine if profiles of participants with health conditions received more health-related advertising.

5.4 Deceptive Advertising Techniques in Health-Related Ads (RQ4)

In this section, we qualitatively and quantitatively describe the products, health conditions, and deceptive techniques present in health-related ads that we collected (Section 4.2.2). We find that 49.5% of health-related ads in our dataset use deceptive advertising techniques, and are most common in ads for supplements, and ads served by native advertising platforms.

5.4.1 Content of Health-Related Ads. First, we describe the types of products and health conditions addressed by ads in our dataset. Table 6 shows the products, health conditions, and deceptive techniques present in our dataset.

Types of health products. The most common products in health-related ads were FDA approved drugs and medications (24.3% of health-related ads), such as Rybelsus, a semaglutide medication to treat diabetes and obseity (103 ads); Allegra, a prescription allergy medication (70 ads), and Auvelity, an antidepressant (61 ads). Ads for health information were also common (22.0%). These included ads from hospital systems or pharmaceutical companies that seek to inform readers about how to manage health conditions like heart disease or allergies, health news aggregator sites, and advertorials that are framed as informational articles about a condition. Other common categories included medical devices (23.0%), such as dental implants (205 ads) and catheters (70 ads); healthcare providers (14. 3%) such as hospitals, rehabilitation facilities and mental health therapists; and skincare products (10.7%).

Health conditions addressed by ads. The most common health conditions addressed were general aging-related issues (17.8%), obesity and weight management (16.8%), general wellness (encompassing fatigue, energy, cognition, and strength – 15.7%), skin conditions (15.4%), dental issues (10.4%). Many of the chronic and sensitive health conditions that we surveyed participants on, such as cancer, cerebrovascular issues, dementia, and gastrointestinal issues were less common, with less than 9% of ads addressing these conditions. These findings suggest that online advertising is more common for health conditions that are less severe and have more over-the-counter or unregulated treatments.

Deceptive advertising techniques. 49.5% of health-related ads used one or more deceptive advertising techniques. The most common was undocumented testimonials (29.3%). Many products, including legitimate prescription medications, showed customer testimonials or reviews that could not be verified, and fabricated social media indicators like fake Facebook comment feeds full of positive reviews. 25% of ads overstated benefits or understated costs, typically using sensationalist language when describing the efficacy of their product. 22.5% of ads made pseudoscientific claims, often using jargon with no scientific basis such as "nootropic", "fat melting", or "science-backed doses". 21.2% of ads used clickbait: these ads would claim to have easy solutions to conditions like cognitive decline, obesity, or lack of energy, but did not disclose what the actual product was - including the name. Instead, users were prompted to watch a video or provide personal information to proceed.

Table 6: Summary of manually-labeled attributes of health-related ads: types of products promoted, health conditions addressed, and deceptive techniques in health-related ads. Each ad may have more than one label in each category.

Category	Count	Percent
Product Type		
Drugs and Medications	1016	24.3%
Medical Devices	965	23.0%
Health Information	921	22.0%
Healthcare Provider or Facility	598	14.3%
Dietary Supplements	453	10.8%
Skin care	446	10.7%
Diet Nutrition Weight Loss Exercise Plan	400	9.6%
Health Insurance or other Financial Services	121	2.9%
Other	116	2.8%
Health-related Charity	58	1.4%
Health Conditions Addressed		
General aging-related issues	746	17.8%
Obesity and weight management	702	16.8%
General wellness (e.g. muscle building, improving cognition)	659	15.7%
Skin care and disease (e.g. acne, skin cancer, psoriasis, rash)	646	15.4%
Dental problems	437	10.4%
Diabetes	375	9.0%
Vision problems	373	8.9%
Cardiovascular problems (e.g. heart attack)	338	8.1%
Mental health	315	7.5%
Erectile dysfunction BPH or other urogenital problems	309	7.4%
Other	303	7.2%
Mobility problems joint pain or arthritis	286	6.8%
Chronic pain, fibromyalgia, neuropathy, functional disorders	275	6.6%
Dementia, Alzheimer's, or neurological impairment	255	6.1%
Gastrointestinal Problems (e.g. irritable bowel syndrome)	219	5.2%
Addiction, substance abuse	203	4.8%
COVID-19 or other infection	199	4.8%
Cancer tumor leukemia or lymphoma	190	4.5%
Sleep disorders (e.g. insomnia obstructive sleep apnea)	135	3.2%
Allergies	134	3.2%
None of the Above	131	3.1%
Hearing problems	122	2.9%
Asthma, COPD, or other pulmonary problems	119	2.8%
Pregnancy, abortion	98	2.3%
Hypertension	94	2.2%
Headaches, migraines	94	2.2%
Autoimmune disease or connective tissue disease	78	1.9%
Sexually transmitted infection (e.g. genital warts)	78	1.9%
Cerebrovascular problems (e.g. stroke)	74	1.8%
Preventative care	70	1.7%
Osteoporosis bone problems	60	1.4%
Liver disease (e.g. cirrhosis of the liver)	53	1.3%
Visited an emergency department or urgent care clinic	51	1.2%
Endocrine disorders (excluding diabetes)	47	1.1%
Chronic kidney disease	25	0.6%
AIDS or other immunodeficiency	20	0.5%
Blood disorder	2	0.0%
Deceptive Techniques		
None of the Above	2114	50.5%
Undocumented testimonials	1225	29.3%
Overstating benefits or understating costs/risks	1046	25.0%
Pseudo-science and prestigious prizes	942	22.5%
Clickbait	886	21.2%
Time-limited offer	502	12.0%
Affiliate marketing	356	8.5%
Money-back guarantee	324	7.7%
Other	24	0.6%

5.4.2 Advertising platforms serving deceptive health-related ads. Next, we analyze the advertising platforms responsible for serving health-related ads in our dataset, and identify the platforms that served a high proportion of deceptive advertising.

Table 7: Top 10 ad platforms that served health-related ads in our dataset, and the percent of their health-related ads that were deceptive.

Ad Platform	Count	Percent	% Deceptive
Google Ads	2215	52.9%	31.8%
Taboola	922	22.0%	90.1%
Criteo	173	4.1%	55.5%
Outbrain	128	3.1%	79.7%
Celtra	86	2.1%	91.9%
The Trade Desk	71	1.7%	14.1%
Viant	60	1.4%	0.0%
BidSwitch	56	1.3%	98.2%
AdRoll	36	0.9%	77.8%
Nativo	32	0.8%	0.0%

We identified the advertising platform involved in serving each health-related ad using the "ad URL" – the initial URL the browser navigates to when clicking on an ad. This URL is often a redirect owned by an ad platform, allowing the platform to measure ad performance. For each ad URL in the dataset, we identified the owner of the domain name, and created a mapping to the matching business entity. We identified an ad platform for 3,949 of 4,254 ads. The remainder were ads whose landing page was directly linked with no redirect URLs.

Table 7 shows the most common ad platforms in our dataset, and the percentage of health-related ads from each platform that used one or more deceptive techniques. Google Ads was the most common platform, accounting for 53% of health-related ads in the dataset, and a moderate amount of those ads used deceptive techniques (32%). By contrast, several smaller platforms served predominantly deceptive health-related ads: Taboola (90% deceptive), Outbrain (79% deceptive), and Celtra (91% deceptive). Taboola and Outbrain are native advertising platforms, which serve ads that imitate the look and feel of first-party content, and have been found to serve deceptive advertising in prior work [12, 76].

5.4.3 Case studies of deceptive health-related advertising. In this section, we examine how deceptive advertising techniques are used to promote products that address the six most common health conditions addressed by ads in our dataset: obesity, skin conditions, general aging, dental issues, general wellness, and diabetes.

Table 8 shows for each of these categories of health conditions, the types of products promoted, and the percentage of those products that used more than two deceptive advertising techniques (which we call "highly deceptive" in this section). Across all health conditions, ads for dietary supplements contained the most highly deceptive ads, at 63.6%, followed by medical devices at 49.5%.

Ads for obesity and weight loss supplements are almost universally deceptive. Ads that addressed obesity, weight management and diabetes were the most common in our dataset and also contained a high proportion of deceptive advertising, particularly in the dietary supplement category.

99.4% of obesity-related supplements and 96.2% of diabetes-related supplements used at least two deceptive techniques, such as the ad shown in Figure 8. The most common deceptive technique used was overstating the benefits of the supplement. In order to capture viewers' attention, many products were marketed as having "fat dissolving" and "fat melting" properties that spare consumers from the hassles of dieting and exercising. Another common technique was undocumented testimonials, in the form of "before and after" pictures from presumed product users, claiming dramatic

Figure 8: Weight loss supplement ad that showcases several deceptive techniques: clickbait (requiring the viewer to watch the video to learn more), undocumented testimonials ("top scientists"), and overstating claims (that just drinking coffee can help you lose weight).

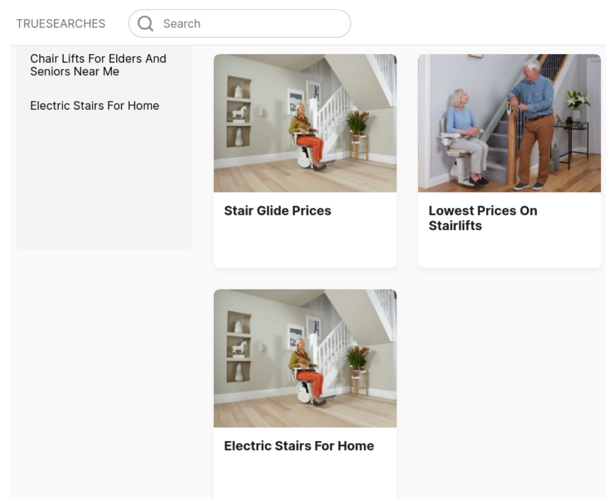

Figure 9: An affiliate marketing ad targeting seniors, promoting stairlifts. Note that no specific product or brand is mentioned. Clicking on these links leads to a Yahoo Search result page. Yahoo likely pays affiliates like TrueSearches to drive traffic to their site for search arbitrage [6].

Table 8: Number of ads by health condition and product type, and the percentage of those ads that are highly deceptive (labeled with more than one deceptive advertising technique). Groups that are > 50% highly deceptive are in bold. Some ads promoted multiple product types or addressed multiple conditions, and are counted multiple times in this table.

Health Condition Addressed	Obesity		Skin		Aging		Dental		Wellness		Diabetes		All Conditions	
Product Type Advertised	n	% Decept.	n	% Decept.	n	% Decept.	n	% Decept.	n	% Decept.	n	% Decept.	n	% Decept.
Diet/Exercise Plan	338	**51.8%**	92	**100.0%**	51	11.8%	70	**100.0%**	187	47.1%	74	**94.6%**	400	44.8%
Dietary Supplements	165	**99.4%**	104	**100.0%**	64	**76.6%**	78	**100.0%**	340	**57.6%**	78	**96.2%**	453	**63.6%**
Drugs and Medications	223	32.3%	143	**53.1%**	36	25.0%	76	**94.7%**	131	**93.1%**	182	39.6%	1016	14.7%
Health Information	224	**52.2%**	133	**75.2%**	121	44.6%	83	**86.7%**	233	36.1%	166	**62.7%**	921	30.9%
Healthcare Provider	77	31.2%	47	**53.2%**	149	18.8%	72	**93.1%**	33	0.0%	57	0.0%	598	16.7%
Medical Devices	95	**82.1%**	85	**100.0%**	128	14.1%	421	**84.3%**	91	**86.8%**	153	**75.8%**	965	49.5%
Other	5	0.0%	2	0.0%	0	0.0%	4	0.0%	8	0.0%	2	0.0%	116	3.4%
Skincare	28	**100.0%**	431	42.7%	292	**53.8%**	4	50.0%	6	50.0%	5	40.0%	446	44.6%
Total	702	43.7%	646	48.0%	746	36.3%	437	**81.2%**	659	34.3%	375	40.3%	4187	28.3%

weight changes in short timeframes. Some ads also used pseudoscientific terms and appeals to authority to promote remedies; for example, an ad claiming to treat diabetic neuropathy stated: "Top U.S. Neurologists: Painful Neuropathy? You Don't Have To Suffer, Try This Immediately…".

Medications that addressed obesity and diabetes primarily promoted Ozempic and other GLP-1 receptor agonists. This is reflective of the rise in popularity of this medication class in recent years. These medications had a far lower percentage of deceptive techniques (32.3% of obesity medications, 39.6% of diabetes medications), though they did tend to use patient testimonials that could not be easily verified.

Supplements for general wellness lean on pseudoscientific claims. Ads that addressed general wellness (unspecific health conditions or problems) were mostly dietary supplements (51.5%). These supplement ads also engaged in deceptive techniques, though at a lower rate than obesity-related ads (57.6%). They often used pseudoscientific terms that overstate a supplement's benefits, such as promising to "nourish, energize, and heal the immune system" or to "work in synergy with the innate wisdom of your body". Others use testimonials and appeals to doctors' authority, such as a supposed news article that mentions a urologist's "no-nonsense daily ritual which ignites explosive muscle growth" and a screenshot of social media comments that supposedly lend truth to how this urologist's health practices have led to "astonishing transformations".

Skincare products frequently use unverifiable claims and testimonials. Ads that addressed skin conditions predominantly promoted creams and serums (66.7%), while fewer advertised dietary supplements, at-home remedies, home-use medical devices, or in-office treatments. Regardless of the modality, the overarching focus of these skincare ads was the elimination of signs of aging: fine lines, wrinkles, dark circles, or dark spots.

Many ads claimed to yield immediate, dramatic results and boasted formulations "based on advanced dermatological research" and "supported by clinical science". Frequently, these deceptive ads attempted to gain credibility by citing that their product was developed or approved by "experts" or "world-renowned" doctors. However, for at least 3 of these cases, we were unable to validate the physicians' identities or association with the product. Ads also contained pseudoscientific jargon such as "maintains dermal hydration" and unscientific claims like "increases the density and cohesion of the skin by up to 87% after just one application". These ads relied heavily on undocumented testimonials from "real," or "verified" customers who claimed that the product made them look 20 years younger, or caused their wrinkles to suddenly disappear.

Many products that target seniors use deceptive affiliate marketing. Aging-related ads overlapped significantly with other conditions, but specifically targeted seniors. A deceptive technique that we saw many unique examples of in this area was affiliate marketing: ads for senior housing, hearing aids, stairlifts (Figure 9), and dental implants were predominant. These ads do not mention brand names or specific products and typically link to a Yahoo search results page—likely as part of a search arbitrage scheme [6]. Beyond not advertising products directly, they also emphasized the "low cost" of the product and the convenience of the product or service being "near me". Supplements targeting aging populations utilize words and phrases such as "longevity and vitality", "reverse aging", and "scientifically proven" to promote their product. Other ads used celebrity endorsements (e.g. from Chuck Norris) and clickbait to promote diet plans to slow aging.

Widespread affiliate marketing campaign promotes dental implants. The majority of ads addressing dental issues were affiliate marketing campaigns that promoted dental implants. The language used in these ads suggests affordability and convenience (e.g. "near me", "low cost", "same day"). Claims of "beautiful new permanent teeth" in just 24 hours likely understate the health risks of immediate implant placement. Non-deceptive ads often targeted dental hygiene or orthodontics services, rather than dental implants.

Finding 4: 49.5% of health-related advertising in our dataset used deceptive advertising techniques. Deceptive techniques are mostly commonly used to promote products like supplements and medical devices, and were observed in more than 75% in the ads served by native advertising platforms.

6 Discussion

We end with a discussion of the policy implications of our results, methodological contributions, and directions for future work on web tracking and targeted advertising.

6.1 Results Summary

6.1.1 Users who browse health-related sites may be targeted with health-related ads. Our results indicate a correlation between the amount of health-related sites a user browses and the amount of health-related ads they receive. We observed that 3.1% of participants' web browsing was health-related (Section 5.1). Participants' visits to health-related pages are almost always tracked: on average, trackers were present on 70% of health-related pages in a participant's browsing history, and trackers associated with Google appeared on 57% of health-related pages (Section 5.2). Health-related browsing then leads to a statistically significant increase in behaviorally targeted health-related ads—every 100 health-related pages a participant browsed increased the percentage of health-related ads served to their profile by 0.28 percentage points (Section 5.3). These results suggest that advertising platforms incorporate users' visits to health-related websites into their advertising profiles, enabling targeting or delivery of ads generally relating to health topics.

However, we did not find conclusive evidence that users are targeted based on specific health conditions: we did not find a correlation between participants' self-reported health conditions and the quantity or content of health-related advertising. Our ability to observe this kind of fine-grained targeted may have been limited by challenges with reliably measuring user and advertiser behavior.

One possible explanation is that our web history dataset contained a low density of health-related browsing. We speculate that people are more likely to perform health information seeking online for a short period of time following a health incident or new diagnosis, and the 90-day window of data collected may not have captured this activity for our participants with health conditions. Another possible explanation is that some ad networks in our sample may prohibit targeting by health conditions. More reputable networks like Google Ads have policies to prevent advertisers of products that treat health conditions from using advanced targeting parameters [32] (though this would not affect targeting outcomes resulting from Google's ad delivery algorithm). Furthermore, ad campaigns are constantly changing, and it is possible that ad campaigns relevant to our participants' health conditions were not running at the time of our crawls. We also discuss other methodological limitations that may have affected our ability to observe targeting in Section 6.4.

While our measurements do not conclusively indicate that online advertisers and platforms infer users' health conditions, they provide a lower bound estimate of the incidence of targeted health-related advertising. Future work could measure such targeting more precisely by conducting measurements from people at the onset of their health conditions and/or conducting measurements directly in users' browsers.

Our results align with prior work measuring tracking and targeted advertising. We find a similar incidence of web tracking, and similar top web trackers as other recent measurement studies [16, 52]. Our work also more precisely estimates the extent of targeted health-related advertising. In a study of ads seen by real users, Zeng et al. did not observe differences in medication ads across demographics, but observed an unequal distribution with the top 20% of users seeing 5% of medication ads [78]. Barford et al. found that with artificial user profiles, health profiles had 14% health ads whereas empty profiles were served 3% health ads [11].

6.1.2 Deceptive advertising techniques are common in health-related advertising. A surprisingly large proportion (49.5%) of health-related advertising in our dataset used deceptive advertising techniques. Dietary supplements were the most deceptive category of product, with 63.6% of supplement ads using at least two deceptive techniques. For certain conditions, deceptive techniques were ubiquitous: 99.4% of weight loss supplements used at least two deceptive techniques. We observed deceptive health-related ads on multiple ad platforms, but we found that native advertising platforms like Taboola had a particularly high concentration of deceptive health ads (90.1%), consistent with prior measurement studies that found that deceptive techniques were frequent in native advertising [12, 76].

We speculate that deceptive health-related advertising is common because of weaknesses in the regulatory environment for online advertising. Many ads we considered deceptive exploited gray areas in the U.S. Federal Trade Commission's regulations on health advertising, such as making claims that supplements "may help" or "promote" body functions [10]. Regulators' resources are limited, and they cannot scrutinize all ads on the internet [23]. Thus, ad platforms have significant discretion in determining what kinds of health-related advertising are permitted. This provides opportunities for deceptive advertisers, who can exploit the gray areas in content policies of platforms like Google, or run ads on platforms with low standards for deceptive content, like Taboola and Outbrain.

6.2 Implications for Policy

Our empirical measurements show that online tracking of browsing behavior and deceptive health-related advertising poses numerous risks to health privacy and consumer welfare. To address these issues, we call on legislators and regulators to implement new regulations and enforcement strategies to limit the harms created by the online tracking and advertising ecosystem.

6.2.1 Health Privacy. We find that users' health-related browsing is frequently exposed to online tracking: top trackers like Meta and Google were present on 23-64% of health-related sites in individuals' web histories. Advertising platforms and data brokers are not covered entities under the HIPAA Privacy Rule, and have no obligation to protect information they learned about users' health through online tracking.

Consequently, we reiterate calls for organizations with obligations to the privacy of their patients or users to remove third-party trackers from their websites, to prevent possible leakage of their users' sensitive health information [28, 29, 33, 51]. Patients usually have little choice but to use the websites of their healthcare providers. We also suggest that new data privacy legislation be enacted to prevent any entity from collecting identifiable data on users' visits to health-related sites without explicit user consent.

Furthermore, privacy regulations should prohibit advertisers and ad platforms from targeting users based on users' health status, even if that information was obtained from sources not considered protected health information (like third-party trackers). Existing regulations in the U.S. do not outright ban targeting of health-related advertising, and ad platforms have inconsistent policies. While Google prohibits advertisers from using sophisticated targeting parameters when promoting health-related products [32], Facebook does not—suggesting that a self-regulatory approach is insufficient. Even if advertisers are not permitted to use health-related targeting parameters, ad platforms' delivery optimization algorithms may result in users being targeted [3]. Regulations should be enacted to prevent biased outcomes for people with health conditions, to reduce the incentives for advertisers and platforms to violate users' health privacy.

6.2.2 Consumer Protection. Deceptive health-related advertising on the web is highly prevalent, suggesting that a new regulatory approach is needed. Prior work shows that the fragmented nature of the online advertising ecosystem means that even when major ad platforms like Google implement stronger policies against deceptive health-related advertising, other ad networks such as native advertising networks may have lower standards, allowing deceptive ads to proliferate [12, 76]. Our results support these findings: native advertising platforms were disproportionately responsible for serving deceptive health-related ads (Section 5.4.2). We suggest that regulators like the U.S. Federal Trade Commission increase the reach of their enforcement actions by shifting from actions against individual advertisers to the ad platforms responsible for hosting a particularly high proportion of deceptive health-related advertising.

6.3 Tools and Directions for Future Work

We extended Adscraper into a scalable measurement platform (Section 4.3) that enables future work on measuring online tracking and targeted advertising. With Adscraper, researchers can now easily create hundreds of browsing profiles, using either real users' histories or artificial crawl lists, and collect ads served to those profiles in a controlled experimental environment. This reduces the implementation burden for testing hypotheses about whether browsing activities lead to targeted advertising.

Adscraper and our experimental methodology could be used in future work to investigate whether advertisers are targeting users on other sensitive attributes. For example, are advertisers targeting users based on their political affiliation or sexual orientation? Are advertisers able to identify and target teens and children? Such studies could use real web histories, as in our study, or artificially constructed histories if such data is difficult to obtain.

Future work could also extend Adscraper beyond display advertising on the web. For example, one could investigate whether web browsing activities are linked to social media advertising via the Meta Pixel, by logging into a Facebook or Instagram account at the beginning of the web history crawl, and then collecting ads from the News Feed or Instagram feed.

To facilitate future research, the measurement platform code has been merged into the Adscraper open-source project and is available on GitHub[3]. We also provide a qualitative framework for labeling the harms of health-related advertising. We taxonomize the deceptive practices unique to this domain and the types of health conditions addressed in health-related advertising. Our codebook, which includes detailed definitions, is available in Appendix A, and our dataset of health-related advertising is available online[4].

6.4 Limitations

Our data collection methodology was based on using web crawlers seeded with users' browsing histories (i.e. sockpuppets), which likely affected the ecological validity of the results. For example, the browsing profiles of the crawlers differed from participants because our crawlers could not access logged-in sites like social media platforms or other sites that require accounts. We filtered out 30% of URLs from the initial history dataset because they required logins. Additionally, the ads our crawlers collected may not reflect what real users received because our crawlers may have been identified as bots by advertising platforms, resulting in them receiving different kinds of ads, or fewer ads.

We only collected data from a single vantage point: an IP address at our institution shared by other real user devices via NAT. The goal was to control for geographic differences between profiles, so that differences could be attributed to differences in participants' browsing histories. Consequently, our results only reflect advertising observed at this location, rather than Internet-wide; and may introduce measurement bias due to being a cloud IP rather than a residential IP [41].

Our classification of health-related websites in users' web histories is less reliable than for ads. We used a machine learning classifier as an initial filter for health-related content in participants' web histories and ad landing pages. However, we only had the resources to manually validate the classifier's labels for ad landing pages (1419 unique pages), and not for web histories (16,667 pages). This may have introduced more uncertainty into the web history and tracking results.

We observed fewer health-related sites and health-related ads in our dataset than expected. Users' web histories were only collected from users in a 90-day time window; users varied in their level of engagement, resulting in some profiles having very little data. In many cases, we did not have the statistical power to conclusively detect differences (or the lack of differences) between users with and without health conditions. However, our work provides a baseline for which future studies can design experiments with appropriate statistical power.

Our dataset of health-related ads was collected from a sample of 400 top sites, and may not generalize to the web as a whole. Ads on social media platforms, ad platforms not present on in the sample, and less frequent advertisers may not be represented in the data.

7 Conclusion

The online tracking and advertising ecosystem poses significant risks to peoples' health privacy. Using the web histories of 107 real users, we conducted a web crawler-based experiment to investigate whether online advertisers target users based on their health status.

[3]https://github.com/UWCSESecurityLab/adscraper
[4]https://osf.io/gdx8k/

We found that the profiles of participants who visited more health-related websites received more health-related web advertising. A substantial portion of participants' health-related browsing histories were observed by third-party web trackers. Furthermore, 49.5% of the health-related ads in our dataset engaged in deceptive advertising practices. To enable this research, we extended the Adscraper web crawler into a measurement platform for auditing targeted advertising at scale, and released the code to enable further research in this area. Our work highlights the need for privacy regulations to protect users' health status from being exposed to online tracking, and stronger enforcement of regulations against deceptive online health advertising.

Acknowledgments

The work described in this paper was supported in part by the Penn Medical Communications Research Institute and the U.S. Army Research Office under MURI grant W911NF-21-1-0317. We thank Carnegie Mellon University Computing Services and Michael Stroucken for setting up the networking infrastructure for our measurement platform. We also thank Jon Maiga for providing an academic license for the uClassify service. Andy Gallardo, Trevor Kann, and McKenna McCall provided feedback on a draft of this paper, and we are grateful for their help. Lastly, we thank the study participants, who provided the web history data that enabled this work.

References

[1] Omer Akgul, Richard Roberts, Moses Namara, Dave Levin, and Michelle L. Mazurek. 2022. Investigating Influencer VPN Ads on YouTube. In *2022 IEEE Symposium on Security and Privacy (SP)*. IEEE Computer Society, San Francisco, CA, USA, 876–892. https://doi.org/10.1109/SP46214.2022.9833633

[2] Muhammad Ali, Angelica Goetzen, Alan Mislove, Elissa M. Redmiles, and Piotr Sapiezynski. 2023. Problematic Advertising and Its Disparate Exposure on Facebook. In *32nd USENIX Security Symposium (USENIX Security 23)*. USENIX Association, Anaheim, CA, USA, 5665–5682. https://www.usenix.org/conference/usenixsecurity23/presentation/ali.

[3] Muhammad Ali, Piotr Sapiezynski, Miranda Bogen, Aleksandra Korolova, Alan Mislove, and Aaron Rieke. 2019. Discrimination through Optimization: How Facebook's Ad Delivery Can Lead to Biased Outcomes. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–30. https://doi.org/10.1145/3359301 arXiv:1904.02095

[4] Marshall Allen. 2018. Health Insurers Are Vacuuming Up Details About You — And It Could Raise Your Rates. https://www.propublica.org/article/health-insurers-are-vacuuming-up-details-about-you-and-it-could-raise-your-rates.

[5] Athanasios Andreou, Giridhari Venkatadri, Oana Goga, Krishna P Gummadi, Patrick Loiseau, and Alan Mislove. 2018. Investigating Ad Transparency Mechanisms in Social Media: A Case Study of Facebook's Explanations. In *NDSS 2018 - Network and Distributed System Security Symposium*. Internet Society, San Diego, CA, 1–15. https://doi.org/10.14722/ndss.2018.23204

[6] Anura. 2024. Understanding Search Arbitrage: What You Need to Know. https://web.archive.org/web/20240416172800/https://www.anura.io/fraud-tidbits/what-is-search-arbitrage.

[7] Noah Ashman. 2020/2021. Outed by Advertisements: How LGBTQ Internet Users Present a Case for Federal Data Privacy Legislation Comment. *Oregon Law Review* 99, 2 (2020/2021), 523–556. https://heinonline.org/HOL/P?h=hein.journals/orglr99&i=525.

[8] American Medical Association. 2015. AMA Calls for Ban on DTC Ads of Prescription Drugs and Medical Devices. https://www.ama-assn.org/press-center/press-releases/ama-calls-ban-dtc-ads-prescription-drugs-and-medical-devices.

[9] American Medical Association. 2024. Privacy, Confidentiality & Medical Records. https://code-medical-ethics.ama-assn.org/chapters/privacy-confidentiality-medical-records.

[10] Rosemary J. Avery, Matthew D. Eisenberg, and Jonathan H. Cantor. 2017. An Examination of Structure-Function Claims in Dietary Supplement Advertising in the U.S.: 2003–2009. *Preventive Medicine* 97 (April 2017), 86–92. https://doi.org/10.1016/j.ypmed.2017.01.008

[11] Paul Barford, Igor Canadi, Darja Krushevskaja, Qiang Ma, and S. Muthukrishnan. 2014. Adscape: Harvesting and Analyzing Online Display Ads. http://arxiv.org/abs/1407.0788. arXiv:1407.0788 [cs]

[12] Muhammad Ahmad Bashir, Sajjad Arshad, and Christo Wilson. 2016. "Recommended For You": A First Look at Content Recommendation Networks. In *Proceedings of the 2016 Internet Measurement Conference (IMC '16)*. Association for Computing Machinery, New York, NY, USA, 17–24. https://doi.org/10.1145/2987443.2987469

[13] Artur Bjelica, Sanja Aleksić, Svetlana Goločorbin-Kon, Darija Sazdanić, Ljilja Torović, and Jelena Cvejić. 2020. Internet Marketing of Cardioprotective Dietary Supplements. *The Journal of Alternative and Complementary Medicine* 26, 3 (March 2020), 204–211. https://doi.org/10.1089/acm.2019.0128

[14] Interactive Advertising Bureau. 2017. Content Taxonomy. https://iabtechlab.com/standards/content-taxonomy/.

[15] M Ryan Calo. 2011. The Boundaries of Privacy Harm. *INDIANA LAW JOURNAL* 86, 3 (July 2011), 1131–1162. https://www.repository.law.indiana.edu/ilj/vol86/iss3/8.

[16] Darion Cassel, Su-Chin Lin, Alessio Buraggina, William Wang, Andrew Zhang, Lujo Bauer, Hsu-Chun Hsiao, Limin Jia, and Timothy Libert. 2022. OmniCrawl: Comprehensive Measurement of Web Tracking With Real Desktop and Mobile Browsers. *Proceedings on Privacy Enhancing Technologies* 2022, 1 (Jan. 2022), 227–252. https://doi.org/10.2478/popets-2022-0012

[17] Centers for Disease Control. 2024. ICD-10-CM. https://www.cdc.gov/nchs/icd/icd-10-cm/index.html.

[18] Mary E. Charlson, Peter Pompei, Kathy L. Ales, and C. Ronald MacKenzie. 1987. A New Method of Classifying Prognostic Comorbidity in Longitudinal Studies: Development and Validation. *Journal of Chronic Diseases* 40, 5 (Jan. 1987), 373–383. https://doi.org/10.1016/0021-9681(87)90171-8

[19] Federal Trade Commission. 2022. Common Health Scams. https://consumer.ftc.gov/articles/common-health-scams.

[20] EasyList Contributors. 2024. EasyList List. https://easylist.to/easylist/easylist.txt.

[21] Matthew A. Crane, Michael J. DiStefano, and Thomas J. Moore. 2023. False or Misleading Claims in Online Direct-to-Consumer Ketamine Advertising in Maryland. *JAMA Network Open* 6, 11 (Nov. 2023), e2342210. https://doi.org/10.1001/jamanetworkopen.2023.42210

[22] Savino Dambra, Iskander Sanchez-Rola, Leyla Bilge, and Davide Balzarotti. 2022. When Sally Met Trackers: Web Tracking From the Users' Perspective. In *31st USENIX Security Symposium (USENIX Security 22)*. USENIX Association, Boston, MA, USA, 2189–2206. https://www.usenix.org/conference/usenixsecurity22/presentation/dambra.

[23] Mary Denigan-Macauley. 2023. *Direct-to-Consumer Advertising of Medical Devices*. Technical Report GAO-23-106197. U.S. Government Accountability Office. https://www.gao.gov/assets/gao-23-106197.pdf.

[24] Benjamin Edelman and Wesley Brandi. 2015. Risk, Information, and Incentives in Online Affiliate Marketing. *Journal of Marketing Research* 52, 1 (Feb. 2015), 1–12. https://doi.org/10.1509/jmr.13.0472

[25] Laura Edelson, Shikhar Sakhuja, Ratan Dey, and Damon McCoy. 2019. An Analysis of United States Online Political Advertising Transparency. https://doi.org/10.48550/arXiv.1902.04385 arXiv:1902.04385 [cs]

[26] Steven Englehardt and Arvind Narayanan. 2016. Online Tracking: A 1-Million-Site Measurement and Analysis. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. ACM, Vienna Austria, 1388–1401. https://doi.org/10.1145/2976749.2978313

[27] Panoptykon Foundation. 21. Algorithms of Trauma: New Case Study Shows That Facebook Doesn't Give Users Real Control over Disturbing Surveillance Ads. https://edri.org/our-work/algorithms-of-trauma-new-case-study-shows-that-facebook-doesnt-give-users-real-control-over-disturbing-surveillance-ads/.

[28] Ari B. Friedman, Lujo Bauer, Rachel Gonzales, and Matthew S. McCoy. 2022. Prevalence of Third-Party Tracking on Abortion Clinic Web Pages. *JAMA Internal Medicine* 182, 11 (Nov. 2022), 1221–1222. https://doi.org/10.1001/jamainternmed.2022.4208

[29] Ari B. Friedman, Raina M. Merchant, Amey Maley, Karim Farhat, Kristen Smith, Jackson Felkins, Rachel E. Gonzales, Lujo Bauer, and Matthew S. McCoy. 2023. Widespread Third-Party Tracking On Hospital Websites Poses Privacy Risks For Patients And Legal Liability For Hospitals. *Health Affairs* 42, 4 (April 2023), 508–515. https://doi.org/10.1377/hlthaff.2022.01205

[30] Liza Gak, Seyi Olojo, and Niloufar Salehi. 2022. The Distressing Ads That Persist: Uncovering The Harms of Targeted Weight-Loss Ads Among Users with Histories of Disordered Eating. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (Nov. 2022), 1–23. https://doi.org/10.1145/3555102

[31] Ghostery. 2024. Ghostery Tracker Database. https://github.com/ghostery/trackerdb.

[32] Google. 2024. Personalised Advertising - Advertising Policies Help. https://web.archive.org/web/20240626005232/https://support.google.com/adspolicy/answer/143465?hl=en.

[33] Ravi Gupta, Ari B Friedman, and Matthew S McCoy. 2023. Medical Journals and Advertiser Tracking—Consequences for Patients, Clinicians, and Editors. *DIGITAL HEALTH* 9 (Jan. 2023), 20552076231176654. https://doi.org/10.1177/

20552076231176654
[34] Andrew F. Hayes and Klaus Krippendorff. 2007. Answering the Call for a Standard Reliability Measure for Coding Data. *Communication Methods and Measures* 1, 1 (April 2007), 77–89. https://doi.org/10.1080/19312450709336664
[35] Paul W. Holland and Roy E. Welsch. 1977. Robust Regression Using Iteratively Reweighted Least-Squares. *Communications in Statistics - Theory and Methods* 6, 9 (Jan. 1977), 813–827. https://doi.org/10.1080/03610927708827533
[36] Julia E. Hood and Allison L. Friedman. 2011. Unveiling the Hidden Epidemic: A Review of Stigma Associated with Sexually Transmissible Infections. *Sexual Health* 8, 2 (May 2011), 159–170. https://doi.org/10.1071/SH10070
[37] Costas Iordanou, Nicolas Kourtellis, Juan Miguel Carrascosa, Claudio Soriente, Ruben Cuevas, and Nikolaos Laoutaris. 2019. Beyond Content Analysis: Detecting Targeted Ads via Distributed Counting. In *Proceedings of the 15th International Conference on Emerging Networking Experiments And Technologies (CoNEXT '19)*. Association for Computing Machinery, New York, NY, USA, 110–122. https://doi.org/10.1145/3359989.3365428
[38] Umar Iqbal, Pouneh Nikkhah Bahrami, Rahmadi Trimananda, Hao Cui, Alexander Gamero-Garrido, Daniel J. Dubois, David Choffnes, Athina Markopoulou, Franziska Roesner, and Zubair Shafiq. 2023. Tracking, Profiling, and Ad Targeting in the Alexa Echo Smart Speaker Ecosystem. In *Proceedings of the 2023 ACM on Internet Measurement Conference (IMC '23)*. Association for Computing Machinery, New York, NY, USA, 569–583. https://doi.org/10.1145/3618257.3624803
[39] Iqvia. 2022. Improving Patient Profiling and Physician Targeting. https://web.archive.org/web/20230818115812/https://www.iqvia.com/library/case-studies/improving-patient-profiling-and-physician-targeting.
[40] Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2016. Bag of Tricks for Efficient Text Classification. https://arxiv.org/abs/1607.01759v3.
[41] Jordan Jueckstock, Shaown Sarker, Peter Snyder, Aidan Beggs, Panagiotis Papadopoulos, Matteo Varvello, Benjamin Livshits, and Alexandros Kapravelos. 2021. Towards Realistic and ReproducibleWeb Crawl Measurements. In *Proceedings of the Web Conference 2021*. ACM, Ljubljana Slovenia, 80–91. https://doi.org/10.1145/3442381.3450050
[42] Arjaldo Karaj, Sam Macbeth, Rémi Berson, and Josep M. Pujol. 2019. WhoTracks .Me: Shedding Light on the Opaque World of Online Tracking. https://doi.org/10.48550/arXiv.1804.08959 arXiv:1804.08959 [cs]
[43] Jon Keegan and Joel Eastwood. 2023. From "Heavy Purchasers" of Pregnancy Tests to the Depression-Prone: We Found 650,000 Ways Advertisers Label You – The Markup. https://themarkup.org/privacy/2023/06/08/from-heavy-purchasers-of-pregnancy-tests-to-the-depression-prone-we-found-650000-ways-advertisers-label-you.
[44] Hyosun Kim. 2015. Trouble Spots in Online Direct-to-Consumer Prescription Drug Promotion: A Content Analysis of FDA Warning Letters. *International Journal of Health Policy and Management* 4, 12 (Aug. 2015), 813–821. https://doi.org/10.15171/ijhpm.2015.157
[45] Michal Kosinski, David Stillwell, and Thore Graepel. 2013. Private Traits and Attributes Are Predictable from Digital Records of Human Behavior. *Proceedings of the National Academy of Sciences* 110, 15 (April 2013), 5802–5805. https://doi.org/10.1073/pnas.1218772110
[46] Victor Le Pochat, Tom Van Goethem, Samaneh Tajalizadehkhoob, Maciej Korczynski, and Wouter Joosen. 2019. Tranco: A Research-Oriented Top Sites Ranking Hardened Against Manipulation. In *Proceedings 2019 Network and Distributed System Security Symposium*. Internet Society, San Diego, CA. https://doi.org/10.14722/ndss.2019.23386
[47] Mathias Lecuyer, Riley Spahn, Yannis Spiliopolous, Augustin Chaintreau, Roxana Geambasu, and Daniel Hsu. 2015. Sunlight: Fine-grained Targeting Detection at Scale with Statistical Confidence. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*. ACM, Denver Colorado USA, 554–566. https://doi.org/10.1145/2810103.2813614
[48] Hao-Ping Hank Lee, Jacob Logas, Stephanie Yang, Zhouyu Li, Natã Barbosa, Yang Wang, and Sauvik Das. 2023. When and Why Do People Want Ad Targeting Explanations? Evidence from a Four-Week, Mixed-Methods Field Study. In *2023 IEEE Symposium on Security and Privacy (SP)*. IEEE Computer Society, San Francisco, CA, USA, 2903–2920. https://doi.org/10.1109/SP46215.2023.10179452
[49] Ada Lerner, Anna Kornfeld Simpson, Tadayoshi Kohno, and Franziska Roesner. 2016. Internet Jones and the Raiders of the Lost Trackers: An Archaeological Study of Web Tracking from 1996 to 2016. In *25th USENIX Security Symposium (USENIX Security 16)*. USENIX Association, Austin, TX, USA. https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/lerner.
[50] C. Luck-Sikorski, P. Roßmann, J. Topp, M. Augustin, R. Sommer, and N.a. Weinberger. 2022. Assessment of Stigma Related to Visible Skin Diseases: A Systematic Review and Evaluation of Patient-Reported Outcome Measures. *Journal of the European Academy of Dermatology and Venereology* 36, 4 (2022), 499–525. https://doi.org/10.1111/jdv.17833
[51] Matthew S. McCoy, Timothy Libert, David Buckler, David T. Grande, and Ari B. Friedman. 2020. Prevalence of Third-Party Tracking on COVID-19–Related Web Pages. *JAMA* 324, 14 (Oct. 2020), 1462–1464. https://doi.org/10.1001/jama.2020.16178
[52] Zahra Moti, Asuman Senol, Hamid Bostani, Frederik Zuiderveen Borgesius, Veelasha Moonsamy, Arunesh Mathur, and Gunes Acar. 2024. Targeted and Troublesome: Tracking and Advertising on Children's Websites. In *2024 IEEE Symposium on Security and Privacy (SP)*. IEEE Computer Society, San Francisco, CA, USA, 1517–1535. https://doi.org/10.1109/SP54263.2024.00118
[53] Terry Nelms, Roberto Perdisci, Manos Antonakakis, and Mustaque Ahamad. 2016. Towards Measuring and Mitigating Social Engineering Software Download Attacks. In *25th USENIX Security Symposium (USENIX Security 16)*. USENIX Association, Austin, TX, USA, 773–789. https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/nelms.
[54] Łukasz Olejnik, Claude Castelluccia, and Artur Janc. 2012. Why Johnny Can't Browse in Peace: On the Uniqueness of Web Browsing History Patterns. In *5th Workshop on Hot Topics in Privacy Enhancing Technologies (HotPETs 2012)*. Sciendo, Vigo, Spain.
[55] Charles Ornstein. 2015. Small Violations Of Medical Privacy Can Hurt Patients And Erode Trust. https://www.npr.org/sections/health-shots/2015/12/10/459091273/small-violations-of-medical-privacy-can-hurt-patients-and-corrode-trust.
[56] Natasha Parekh and William H. Shrank. 2018. Dangers and Opportunities of Direct-to-Consumer Advertising. *Journal of General Internal Medicine* 33, 5 (May 2018), 586–587. https://doi.org/10.1007/s11606-018-4342-9
[57] Rebecca Passonneau. 2006. Measuring Agreement on Set-valued Items (MASI) for Semantic and Pragmatic Annotation. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*. European Language Resources Association (ELRA), Genoa, Italy. http://www.lrec-conf.org/proceedings/lrec2006/pdf/636_pdf.pdf.
[58] Angelisa C. Plane, Elissa M. Redmiles, Michelle L. Mazurek, and Michael Carl Tschantz. 2017. Exploring User Perceptions of Discrimination in Online Targeted Advertising. In *26th USENIX Security Symposium (USENIX Security 17)*. USENIX Association, Vancouver, BC, Canada, 935–951. https://www.usenix.org/conference/usenixsecurity17/technical-sessions/presentation/plane.
[59] W. Nicholson Price and I. Glenn Cohen. 2019. Privacy in the Age of Medical Big Data. *Nature Medicine* 25, 1 (Jan. 2019), 37–43. https://doi.org/10.1038/s41591-018-0272-7
[60] Vaibhav Rastogi, Rui Shao, Yan Chen, Xiang Pan, Shihong Zou, and Ryan Riley. 2016. Are These Ads Safe: Detecting Hidden Attacks through the Mobile App-Web Interfaces. In *Proceedings 2016 Network and Distributed System Security Symposium*. Internet Society, San Diego, CA. https://doi.org/10.14722/ndss.2016.23234
[61] Nathan Reitinger, Bruce Wen, Michelle Mazurek, and Blase Ur. 2024. What Does It Mean to Be Creepy? Responses to Visualizations of Personal Browsing Activity, Online Tracking, and Targeted Ads. *Proceedings on Privacy Enhancing Technologies* 2024, 3 (2024). https://petsymposium.org/popets/2024/popets-2024-0101.php.
[62] Franziska Roesner, Tadayoshi Kohno, and David Wetherall. 2012. Detecting and Defending Against {Third-Party} Tracking on the Web. In *9th USENIX Symposium on Networked Systems Design and Implementation (NSDI 12)*. USENIX Association, San Jose, CA, USA, 155–168. https://www.usenix.org/conference/nsdi12/technical-sessions/presentation/roesner.
[63] Princess Sampson, Ro Encarnacion, and Danaë Metaxa. 2023. Representation, Self-Determination, and Refusal: Queer People's Experiences with Targeted Advertising. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 1711–1722. https://doi.org/10.1145/3593013.3594110
[64] Sebastian Schelter and Jérôme Kunegis. 2016. Tracking the Trackers: A Large-Scale Analysis of Embedded Web Trackers. *Proceedings of the International AAAI Conference on Web and Social Media* 10, 1 (2016), 679–682. https://doi.org/10.1609/icwsm.v10i1.14769
[65] Latanya Sweeney. 2013. Discrimination in Online Ad Delivery: Google Ads, Black Names and White Names, Racial Discrimination, and Click Advertising. *Queue* 11, 3 (March 2013), 10–29. https://doi.org/10.1145/2460276.2460278
[66] uClassify. 2024. IAB Taxonomy V2 Classifier. https://uclassify.com/browse/uclassify/iab-taxonomy-v2.
[67] Blase Ur, Pedro Giovanni Leon, Lorrie Faith Cranor, Richard Shay, and Yang Wang. 2012. Smart, Useful, Scary, Creepy: Perceptions of Online Behavioral Advertising. In *Proceedings of the Eighth Symposium on Usable Privacy and Security (SOUPS '12)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/2335356.2335362
[68] U.S. Department of Health and Human Services. 2008. Summary of the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html.
[69] Miranda Wei, Madison Stamos, Sophie Veys, Nathan Reitinger, Justin Goodman, Margot Herman, Dorota Filipczuk, Ben Weinshel, Michelle L Mazurek, and Blase Ur. 2020. What Twitter Knows: Characterizing Ad Targeting Practices, User Perceptions, and Ad Explanations Through Users' Own Twitter Data. In *29th USENIX Security Symposium (USENIX Security 20)*. USENIX Association, 145–162.
[70] Ben Weinshel, Miranda Wei, Mainack Mondal, Euirim Choi, Shawn Shan, Claire Dolin, Michelle L. Mazurek, and Blase Ur. 2019. Oh, the Places You've Been! User

Reactions to Longitudinal Transparency About Third-Party Web Tracking and Inferencing. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security (CCS '19)*. Association for Computing Machinery, New York, NY, USA, 149–166. https://doi.org/10.1145/3319535.3363200

[71] Xinyu Xing, Wei Meng, Byoungyoung Lee, Udi Weinsberg, Anmol Sheth, Roberto Perdisci, and Wenke Lee. 2015. Understanding Malvertising Through Ad-Injecting Browser Extensions. In *Proceedings of the 24th International Conference on World Wide Web (WWW '15)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 1286–1295. https://doi.org/10.1145/2736277.2741630

[72] Christina Yeung, Umar Iqbal, Yekaterina Tsipenyuk O'Neil, Tadayoshi Kohno, and Franziska Roesner. 2023. Online Advertising in Ukraine and Russia During the 2022 Russian Invasion. In *Proceedings of the ACM Web Conference 2023 (WWW '23)*. Association for Computing Machinery, New York, NY, USA, 2787–2796. https://doi.org/10.1145/3543507.3583484

[73] Elad Yom-Tov. 2020. Screening for Cancer Using a Learning Internet Advertising System. *ACM Transactions on Computing for Healthcare* 1, 2 (April 2020), 1–13. https://doi.org/10.1145/3373720

[74] Kai Yuan, Xiao-Lin Huang, Wei Yan, Yu-Xin Zhang, Yi-Miao Gong, Si-Zhen Su, Yue-Tong Huang, Yi Zhong, Yi-Jie Wang, Ze Yuan, Shan-Shan Tian, Yong-Bo Zheng, Teng-Teng Fan, Ying-Jian Zhang, Shi-Qiu Meng, Yan-Kun Sun, Xiao Lin, Tian-Ming Zhang, Mao-Sheng Ran, Samuel-Yeung-Shan Wong, Nicolas Rüsch, Le Shi, Yan-Ping Bao, and Lin Lu. 2022. A Systematic Review and Meta-Analysis on the Prevalence of Stigma in Infectious Diseases, Including COVID-19: A Call to Action. *Molecular Psychiatry* 27, 1 (Jan. 2022), 19–33. https://doi.org/10.1038/s41380-021-01295-8

[75] Eric Zeng. 2025. Adscraper: A Web Crawler for Measuring Online Ad Content. https://github.com/UWCSESecurityLab/adscraper.

[76] Eric Zeng, Tadayoshi Kohno, and Franziska Roesner. 2020. Bad News: Clickbait and Deceptive Ads on News and Misinformation Websites. In *Workshop on Technology and Consumer Protection*. IEEE Computer Society. https://www.ieee-security.org/TC/SPW2020/ConPro/papers/zeng-conpro20.pdf.

[77] Eric Zeng, Tadayoshi Kohno, and Franziska Roesner. 2021. What Makes a "Bad" Ad? User Perceptions of Problematic Online Advertising. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–24. https://doi.org/10.1145/3411764.3445459

[78] Eric Zeng, Rachel McAmis, Tadayoshi Kohno, and Franziska Roesner. 2022. What Factors Affect Targeting and Bids in Online Advertising?: A Field Measurement Study. In *Proceedings of the 22nd ACM Internet Measurement Conference*. ACM, Nice France, 210–229. https://doi.org/10.1145/3517745.3561460

[79] Eric Zeng, Miranda Wei, Theo Gregersen, Tadayoshi Kohno, and Franziska Roesner. 2021. Polls, Clickbait, and Commemorative $2 Bills: Problematic Political Advertising on News and Media Websites around the 2020 U.S. Elections. In *Proceedings of the 21st ACM Internet Measurement Conference*. ACM, Virtual Event, 507–525. https://doi.org/10.1145/3487552.3487850

A Qualitative Codebook for Health-Related Ads

In this appendix, we provide the codes used to label health-related ad landing pages, and the definitions and instructions provided to coders.

A.1 Health-Related

- *Health-Related*: An ad is health-related if it promotes: a drug, supplement, service, or medical device that treats a health condition; a pharmacy or other business that sells drugs, supplements, or other treatments; a facility/organization that provides healthcare, like a clinic, a doctor's office, or hospital; insurance or other financial services for people seeking medical care information about medications, supplements, or conditions
- *Targeted at Healthcare Professionals*: Ads that are not intended for patients, but are related to health. This includes journal articles, job postings for healthcare workers, the HCP-oriented page for prescription drugs, and professional organizations.
- *Not Health Related*: Ads that are not related to human health or wellness.

If the ad is not health-related, we stopped labeling at this point.

A.2 Product Type

The type of product being promoted in the ad. An ad may promote multiple types of products, and could have multiple labels.

- *Diet, Exercise, Nutrition, and Weight Loss Plans*
- *Dietary Supplements*
- *Drugs and Medications*
- *Health Information*
- *Health Insurance or other Financial Services*
- *Healthcare Provider or Facility*
- *Medical Devices*
- *Skincare*
- *Other*

A.3 Health Conditions

An ad is relevant to a health condition if the ad:

(1) Promotes a medication, supplement, or medical device that treats the condition,
(2) Promotes or provides information on how to manage a condition, or
(3) Promotes healthcare providers who could treat the specific condition.

The following is the list of health conditions. An ad may address more than one health condition, and could have multiple labels.

- Addiction/Substance Abuse
- Autoimmune disease or connective tissue disease
- Asthma/COPD/Pulmonary problems
- Blood Disorder
- Cardiovascular Problems (including Heart Attack, congestive heart failure, peripheral vascular disease)
- Cancer/Tumor/Leukemia/Lymphoma
- Cerebrovascular Problems/Stroke/TIA
- Chronic kidney disease
- Chronic Pain/Fibromyalgia/Neuropathy/Functional Disorders/Paralysis
- Headaches, migraines
- COVID-19 or other infection
- Dementia (including Alzheimer's disease) or other neurological impairment
- Dental problems
- Diabetes
- Other Endocrine Disorders (excluding Diabetes)
- Erectile Dysfunction/BPH/other urogenital
- General aging-related issues
- General Wellness /(including supplements for body, muscle building, cognition)
- GI Problems (e.g. peptic ulcer disease, gastrointestinal bleeding, irritable bowel syndrome, chronic constipation)
- Hearing Problems
- Hypertension
- Sleep disorders (e.g. Insomnia, Obstructive sleep apnea)
- Liver Disease (e.g. cirrhosis of the liver)
- Mental Health
- Mobility problems/Joint pain/arthritis

- Obesity/weight management
- Osteoporosis/bone problems
- Pregnancy/Abortion
- Preventative care
- Seasonal Allergies
- Sexually Transmitted Infection (including anal/genital warts)
- Skin care/disease (acne, skin cancer, psoriasis, other rash)
- Vision Problems
- Visited an emergency department or urgent care

A.4 Deceptive Techniques

For each ad, we label whether it engages in any deceptive advertising practices. We derived the following framework based on guidance from the Federal Trade Commission and Food and Drug Administration:

- *Overstating benefits or understating costs/risks*: This category includes exaggeration of benefits across disease areas (i.e., one product cures a range of conditions) and within a given disease (i.e., the product cures something like Alzheimer's for which there is now cure). This can also include ads that downplay costs or risks – for example, losing weight without needing to exercise or change diet.
- *Undocumented testimonials*: Testimonials that aren't linked to a verifiable person, such as "Top Scientists say...", "Cardiologist...", or "John, 36". We distinguish testimonials from regular product reviews using two heuristics: (1) if there's no interface to leave your own review, then it's an undocumented testimonial. If it looks like you can leave a review, and it will actually be added to the site, then it is not an undocumented testimonial. (2) If the site shows a list of reviews, there are no negative reviews in the list, and the reviews look cherry-picked to craft a certain message, it may be an undocumented testimonial.
- *Money-back guarantee*: Includes ads that promise a refund if the product doesn't work as advertised. This is because in practice, there is no way to verify that a refund request will be carried out.
- *Time-limited offer*: Includes ads that make time-limited offers or use other high pressure language to encourage the consumer to act quickly.
- *Pseudo-science and prestigious prizes*: Includes ads that reference fake scientific or medical terminology or real scientific prizes like the Nobel Prize. Also include claims like "FDA approved ingredients" here in that it references a scientific regulatory agency but in a misleading way. We do not categorize mainstream products that are a bit questionable scientifically as pseudo-science - for example, an ad for a skincare product from a mainstream skincare company that makes questionable claims about mechanism of action.
- *Clickbait*: Ads where the actual product or information is not explicitly described, and you must take an action (like watching a video, providing personal information) to find out.
- *Affiliate marketing*: Ads where the page does not promote a single product, but instead, links to multiple other sites that presumably actually advertise the products. Typically, these contain links to multiple products - e.g. ads within the ad - that aren't their own products. This is sometimes an arbitrage technique used to drive fraudulent traffic. For example, a search engine may have promised an advertiser a certain amount of traffic for certain keywords. To achieve their traffic goals, they pay affiliates to run ads on other websites that link to the desired search query — for a lower cost per impression than they are receiving from the advertiser. However, these searches are not truly "organic".

B Statistical Appendix

In this Appendix, we report details on statistical tests, including details on testing of assumptions, power analyses, and full regression tables.

B.1 Health-Related Browsing History

In Section 5.1.3, we analyzed whether there was a relationship between participants' health conditions and the number of health-related webpages they visited.

Health condition vs. webpages about health condition. For each of the health condition categories specified by ICD-10-CM, we compared how many pages related to that condition were browsed by participants with that condition, versus participants who did not have that condition. Because the data was not normally distributed, we conducted a Mann-Whitney U test for each condition. To correct for multiple comparisons, we conducted a Holm-Bonferroni correction. The results and corrected p-values are in Table 9. We did not detect significant differences for any health condition.

Because the sample size of participants per health condition was low, to help us interpret the non-significant results, we conducted a sensitivity power analysis. This helps determine at what effect size we would have been able to conduct these tests with sufficient statistical power. Table 9 shows the predicted minimum effect size needed to find significance at $\alpha = 0.05$ with power 0.8, and the observed effect size (effect sizes in Cohen's d). Because in each case, the observed effect size was less than the required effect size, our tests were underpowered, and suggest that either the effect was too small to observe, or there was no difference between groups.

Number of comorbidities vs. health-related webpages. We ran a regression analysis to test whether there was a correlation between the number of comorbidities a participant had, and the proportion of webpages they visited that were health-related. The dependent variable was the proportion of health-related sites visited, and the independent variable was the count of self-reported health conditions. A sensitivity power analysis showed that our sample size was sufficient to detect effects with coefficient <0.006.

We initially fit a simple linear regression model, and found no significant association between comorbidities and proportion of sites visited that were health-related. However, some linear regression assumptions were violated: a Shapiro-Wilk test detected non-normality in the residuals (p<0.001).

We conducted a sensitivity analysis with nonlinear models to assess the robustness of this finding. We conducted a Poisson regression, but found that the data was overdispersed (dispersion ratio=92.05, χ^2=9389.146, p<0.001). To account for underdispersion,

Table 9: Mann Whitney U test results and sensitivity power analysis results for hypothesis tests comparing health-related sites visited by participants with and without health conditions. The p-values are the corrected p-values after performing a Holm-Bonferroni correction. "Observed *d*" is Cohen's d calculated on the observed data. "Required *d*" is the predicted minimum effect size that the study need to find a significant effect, at α=0.05 and power=0.8 and with the actual sample sizes we were able to obtain. No results were significant and the observed effect sizes were smaller than the predicted required effect sizes.

	Participants w/Condition		Participants w/o Condition					
ICD-10-CM Condition	*n*	Pages (mean)	*n*	Pages (mean)	U	p	Observed *d*	Required *d*
Any Condition	73	166.40	34	132.88	1305.5	1.000	0.208	0.477
Circulatory	15	22.00	92	4.28	726.5	1.000	0.458	0.641
Digestive	16	3.50	91	1.76	940.5	0.317	0.251	0.623
Ear	13	8.23	94	4.55	674.0	1.000	0.263	0.681
Endocrine	23	17.43	84	25.32	998.5	1.000	0.243	0.542
Genitourinary	15	10.87	92	4.67	737.5	1.000	0.270	0.641
Health Services	36	12.58	71	25.24	1114.0	1.000	0.192	0.471
Musculoskeletal	12	16.08	95	7.34	598.0	1.000	0.255	0.705
Neoplasms	9	2.56	98	2.57	467.5	1.000	0.003	0.802
Respiratory	32	15.34	75	23.59	1078.0	1.000	0.306	0.486
Skin	9	5.78	98	2.46	481.5	1.000	0.314	0.802

Table 10: Quasi-binomial regression model outputs for comorbidities vs. number of health-related pages in history.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-3.50322	0.15887	-22.052	<0.001***
Comorbidities	0.04313	0.05862	0.736	0.464
Observations	107			

we conducted a negative binomial regression. Table 10 shows the results of this regression. We again did not observe a significant effect of comorbidities on number of health-related sites visited confirming the findings of the linear model. We report the results of the negative binomial regression in the body of the paper.

B.2 Targeting of Health-Related Ads

In Section 5.3, we analyzed whether participants' health conditions, health-related browsing histories, and demographics affected the number of health-related ads their profile received in the targeted advertisement experiment.

Total health-related ads observed. We conducted a regression analysis to test whether health-related browsing, health conditions, or demographic characteristics affected the number of ads observed that were health-related. A sensitivity power analysis found that the sample size was sufficient to detect effect sizes of $f^2 = 0.07$, which is between a small and medium effect size.

We initially used a multiple linear regression, with the formula:

$$CountHealthAds/CountAds \sim Age + Gender + NumHealthSites + Circulatory + Musculoskeletal + Digestive + Genitourinary + Neoplasms + Endocrine + Skin + Ear + HealthServices + Respiratory$$

We tested the linear regression assumptions. Linearity and independence of residuals were satisfied. A Shapiro-Wilk test detected analysis with nonlinear models, to assess the non-normality in the residuals (p=0.029). A Breusch-Pagan test detected heteroskedasticity in the residuals (p=0.021). We used robust standard errors to mitigate the impact of heteroskedasticity. Table 12 shows the results of the multiple linear regression.

To assess whether the violation of non-normality in the residuals had an effect on our estimates and p-values, we conducted a sensitivity with nonlinear models. We initially conducted a Poisson regression, which models count data, with the formula:

$$CountHealthAds \sim offset(log(CountAds)) + Age + Gender + NumHealthSites + Circulatory + Musculoskeletal + Digestive + Genitourinary + Neoplasms + Endocrine + Skin + Ear + HealthServices + Respiratory$$

A test of dispersion detected overdispersion (dispersion ratio=4.023, χ^2=362.09, p<0.001); to account for this we conducted a negative binomial regression with the same formula, which models overdispersed count data. Table 13 shows the results of this regression. We found that health-related history size continued to be significant (p=0.005), gender was significant (p=0.021), and digestive system conditions were significant (p=0.023). All other predictors were not significant. The directionality of the coefficient estimates and the p-value magnitudes were roughly similar to the linear model, but the linear model's confidence for gender was lower.

These results suggest that despite non-normality in the residuals of the linear model, the estimates and confidence intervals for the predictors are not meaningfully different from that of a nonlinear model, and our conclusions remain the same. To improve the interpretability of the results, we report the results of the multiple linear regression in the main text of the paper.

Health Conditions vs. Health-Related Ads for Specific Conditions. For each of the health condition categories specified by ICD-10-CM, we compared how many ads related to that condition were observed in profiles based on participants with that condition, versus profiles based on participants who did not have that condition. Because the

Table 11: Mann Whitney U test results and sensitivity power analysis results for hypothesis tests comparing ads received by profiles for participants with and without health conditions. The p-values are the corrected p-values after performing a Holm-Bonferroni correction. "Observed *d*" is Cohen's d calculated on the observed data. "Required *d*" is the predicted minimum effect size that the study need to find a significant effect, at α=0.05 and power=0.8 and with the actual sample sizes we were able to obtain. No results were significant and the observed effect sizes were smaller than the predicted required effect sizes.

	Participants w/Condition		Participants w/o Condition					
ICD-10-CM Condition	*n*	Ads (mean)	*n*	Ads (mean)	U	p	Observed *d*	Required *d*
Any Condition	70	40.46	34	39.85	1172.5	1.000	0.045	0.481
Circulatory	14	4.71	90	3.46	853.5	0.347	0.595	0.661
Digestive	15	1.60	89	2.73	581.0	1.000	0.474	0.642
Ear	13	1.77	91	1.09	672.5	1.000	0.288	0.682
Endocrine	23	7.43	81	8.43	868.0	1.000	0.250	0.544
Genitourinary	14	3.43	90	3.24	628.0	1.000	0.060	0.661
Health Services	35	8.31	69	6.87	1327.0	1.000	0.287	0.477
Musculoskeletal	11	2.36	93	3.35	428.5	1.000	0.477	0.734
Neoplasms	9	1.44	95	1.86	384.0	1.000	0.259	0.802
Respiratory	32	3.28	72	3.46	1060.5	1.000	0.054	0.489
Skin	9	14.22	95	5.45	592.0	0.566	0.528	0.802

Table 12: Linear regression model output for factors that affect the percentage of health-related ads served to a profile.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	4.218	0.647	6.521	<0.001***
Age	0.014	0.011	1.312	0.193
Gender (Male)	-0.736	0.416	-1.768	0.081
Health-related History Size	0.003	0.001	2.490	0.015*
Circulatory	-0.230	0.467	-0.493	0.624
Musculoskeletal	-0.339	0.592	-0.573	0.568
Digestive	-0.919	0.455	-2.020	0.046*
Genitourinary	0.551	0.649	0.849	0.398
Neoplasms	0.216	0.859	0.252	0.802
Endocrine	-0.413	0.506	-0.816	0.417
Skin	0.916	0.760	1.205	0.231
Ear	-0.410	0.644	-0.636	0.526
Health Services	-0.439	0.387	-1.136	0.259
Respiratory	0.205	0.349	0.586	0.559
Observations	104			
Adjusted R^2	0.0803			

Table 13: Negative binomial regression output for factors that affect the count of health-related ads served to a profile. Used for a sensitivity analysis to assess the robustness of the linear model in Table 12.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-3.149	0.113	-27.960	<0.001***
Age	0.003	0.002	1.459	0.145
Gender (Male)	-0.157	0.068	-2.315	0.021 *
Health-Related History Size	0.000	0.000	2.786	0.005**
Circulatory	-0.041	0.102	-0.402	0.688
Musculoskeletal	-0.062	0.127	-0.486	0.627
Digestive	-0.201	0.088	-2.273	0.023*
Genitourinary	0.086	0.103	0.835	0.404
Neoplasms	0.003	0.128	0.022	0.982
Endocrine	-0.081	0.090	-0.893	0.372
Skin	0.183	0.114	1.604	0.109
Ear	-0.054	0.098	-0.548	0.584
Health Services	-0.090	0.072	-1.246	0.213
Respiratory	0.049	0.070	0.705	0.481

data was not normally distributed, we conducted a Mann-Whitney U test for each condition. To correct for multiple comparisons, we conducted a Holm-Bonferroni correction. The results and corrected p-values are in Table 11. We did not detect significant differences for any health condition.

Because the sample size of participants per health condition was low, we conducted a sensitivity power analysis, to determine at what effect size we would have been able to conduct these tests with sufficient statistical power. Table 11 shows the predicted minimum effect size needed to find significance at $\alpha = 0.05$ with power 0.8 and the sample sizes we were able to obtain, and the observed effect size (effect sizes in Cohen's d). Because in each case, the observed effect size was less than the required effect size, our tests were underpowered, and suggest that either the effect was too small to observe, or there was no difference between groups.

Health Related Browsing vs. Health-Related Ads for Specific Conditions. For each ICD-10-CM category, we ran a re-weighted least squares regressions to test whether there was a correlation between the number pages about a condition a participant visited, and the number of health-related ads observed from their profile. The dependent variable was the proportion of ads observed about a specific health condition (out of all ads observed), and the independent variable was the number of pages relating to that condition in the participant's history. A sensitivity power analysis showed that our sample size was sufficient to detect effects with coefficients <0.0001 to 0.0004.

We used re-weighted least squares regressions, because diagnostic graphs showed outliers that may have spuriously resulted in significant coefficients for the Skin and Ear conditions. Weighted least squares regressions are robust to outliers [35]. Table 14 shows

Table 14: Robust linear regression model outputs for ads about health condition vs. pages visited relevant to the health condition. A different regression model was fit for each condition. Browsing history did not have a significant effect for any health condition.

Condition	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	3.4438	0.2323	14.8231	<0.001***
Circulatory history size	0.0046	0.0088	0.5171	0.6051
(Intercept)	1.9264	0.1717	11.2190	<0.001***
Digestive history size	-0.0013	0.0121	-0.1081	0.9139
(Intercept)	0.9153	0.1243	7.3622	<0.001***
Ear history size	-0.0125	0.0080	-1.5602	0.1187
(Intercept)	7.7492	0.4145	18.6956	<0.001***
Endocrine history size	-0.0041	0.0076	-0.5493	0.5828
(Intercept)	2.9173	0.2621	11.1301	<0.001***
Genitourinary history size	-0.0176	0.0256	-0.6875	0.4918
(Intercept)	6.8676	0.4488	15.3013	<0.001***
Health services history size	-0.0003	0.0184	-0.0184	0.9853
(Intercept)	2.8017	0.1845	15.1892	<0.001***
Musculoskeletal history size	0.0012	0.0069	0.1788	0.8581
(Intercept)	1.4065	0.1930	7.2864	<0.001***
Neoplasms history size	0.0126	0.0611	0.2071	0.8360
(Intercept)	3.9390	0.2912	13.5292	<0.001***
Nervous history size	-0.0330	0.0374	-0.8826	0.3774
(Intercept)	0.7468	0.0990	7.5400	<0.001***
Pregnancy history size	0.0008	0.0016	0.5352	0.5925
(Intercept)	3.0571	0.3345	9.1389	<0.001***
Respiratory history size	-0.0017	0.0075	-0.2269	0.8205
(Intercept)	5.0496	0.3262	15.4787	<0.001***
Skin history size	0.0011	0.0740	0.0148	0.9882

the results of these regressions. We did not observe a correlation between participants' browsing about a condition and the number of ads about that condition the profile received.

We tested the linear regression assumptions in each case, finding heteroskedasticity in regression models for Skin and Respiratory conditions. Thus, we report robust standard errors for all regressions.

We detected non-normality in the residuals for each regression. To assess whether this affected our estimates, we conducted a sensitivity analysis with negative binomial regressions (after detecting overdispersion when using Poisson regressions). The regression outputs are summarized in Table 15. We detected significant associations between history and ads for Skin (β=0.034, p<0.001) and Ear (β=-0.05489, p=0.044) conditions, and no significant effects for the other conditions. However, we previously identified outliers in these conditions that likely affected these estimates. Removing the outliers and refitting the models resulted in no significant associations. We conclude that the findings from our re-weighted least squares regression models were robust to violations of the normality of residuals assumption. To improve the interpretability of the results, we use the coefficients of the linear model to plot fitted lines in Figure 7.

Table 15: Negative binomial regression outputs for ads about health condition vs. pages visited relevant to the health condition. A different regression model was fit for each condition. History had an effect for Skin and Ear conditions, but removing one outlier point in both conditions made the estimates non-significant.

Condition	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-5.4407	0.0636	-85.548	<0.0001***
Circulatory history size	0.0015	0.0025	0.596	0.5510
(Intercept)	-5.7635	0.1052	-54.77	<0.0001***
Digestive history size	-0.0078	0.0147	-0.53	0.5960
(Intercept)	-6.4161	0.1608	-39.902	<0.0001***
Ear history size	-0.0549	0.0272	-2.019	0.0435*
Ear condition outlier removed				
(Intercept)	-6.51373	0.15530	-41.942	<0.0001***
Ear history size	-0.05126	0.02617	-1.959	0.0501
(Intercept)	-4.6078	0.0583	-79.073	<0.0001***
Endocrine history size	-0.0002	0.0013	-0.155	0.8760
(Intercept)	-5.4880	0.0882	-62.198	<0.0001***
Genitourinary history size	-0.0094	0.0071	-1.321	0.1870
(Intercept)	-4.7294	0.0606	-78.101	<0.0001***
Health services history size	0.0001	0.0008	0.188	0.8510
(Intercept)	-5.5326	0.0791	-69.935	<0.0001***
Musculoskeletal history size	-0.0010	0.0035	-0.272	0.7850
(Intercept)	-6.1166	0.0990	-61.790	<0.0001***
Neoplasms history size	0.0010	0.0121	0.083	0.9340
(Intercept)	-5.2032	0.0772	-67.386	<0.0001***
Nervous history size	-0.0173	0.0109	-1.594	0.1110
(Intercept)	-6.7833	0.1391	-48.758	<0.0001***
Pregnancy history size	0.0001	0.0024	0.043	0.9660
(Intercept)	-5.4479	0.1108	-49.171	<0.0001***
Respiratory history size	-0.0019	0.0031	-0.627	0.5310
(Intercept)	-5.0398	0.0807	-62.436	<0.0001***
Skin history size	0.0338	0.0099	3.426	0.0006**
Skin condition outlier removed				
(Intercept)	-4.97464	0.07440	-66.863	<0.0001***
Skin history size	-0.01379	0.01273	-1.083	0.279

Total number of deceptive health-related ads. We conducted a regression analysis to test whether health-related browsing, health conditions, or demographic characteristics affected the proportion of ads observed that were deceptive and health-related. A sensitivity power analysis found that the sample size was sufficient to detect effect sizes of $f^2 = 0.07$, which is between a small and medium effect size.

We initially conducted a multiple linear regression, with the formula:

$$
\begin{aligned}
&ProportionDeceptiveHealthAds \sim Age + Gender + \\
&NumHealthSites + Circulatory + Musculoskeletal + \\
&Digestive + Genitourinary + Neoplasms + Endocrine + Skin + \\
&Ear + HealthServices + Respiratory
\end{aligned}
$$

When checking the assumptions, a Shapiro-Wilk test detected non-normality in the residuals (p=0.029), and a Breusch-Pagan test detected heteroskedasticity in the residuals (p=0.021). Additionally, we observed outliers in a diagnostic plot. Thus, we conducted a

Table 16: Robust regression model outputs for the proportion of deceptive health-related ads observed per profile. No predictors were significant.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	15.648	3.013	5.194	<0.001***
Age	0.068	0.050	1.352	0.177
Gender (Male)	-1.155	1.580	-0.731	0.465
health related history size	0.002	0.006	0.316	0.752
Circulatory	-0.855	2.538	-0.337	0.736
Musculoskeletal	-3.931	3.370	-1.166	0.244
Digestive	-1.250	2.325	-0.538	0.591
Genitourinary	-3.220	3.345	-0.963	0.336
Neoplasms	4.043	3.528	1.146	0.252
Endocrine	0.066	2.719	0.024	0.981
Skin	4.244	2.825	1.502	0.133
Ear	-0.389	3.329	-0.117	0.907
Health services	-1.164	1.986	-0.586	0.558
Respiratory	1.099	1.663	0.661	0.509
Observations	104			

Table 17: Negative binomial regression model outputs for the number of deceptive health-related ads observed per profile. Age and neoplasm conditions had a significant effect.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-3.989	0.141	-28.269	<0.001***
Age	0.005	0.003	2.026	0.043*
Gender	-0.116	0.085	-1.369	0.171
Health-Related History Size	0.000	0.000	1.358	0.175
Circulatory	-0.102	0.128	-0.793	0.428
Musculoskeletal	-0.288	0.163	-1.771	0.077
Digestive	-0.183	0.111	-1.641	0.101
Genitourinary	-0.038	0.131	-0.287	0.774
Neoplasms	0.352	0.157	2.244	0.025*
Endocrine	-0.182	0.114	-1.596	0.110
Skin	0.257	0.141	1.827	0.068
Ear	-0.011	0.122	-0.094	0.925
Health Services	-0.011	0.090	-0.126	0.900
Respiratory	0.115	0.088	1.312	0.189

robust linear regression using weighted least squares, and report robust standard errors. Table 16 shows the results of this regression. This regression did not detect any significant effects on the proportion of deceptive health-related ads a profile received.

Because assumption of non-normality of residuals was violated, we conducted a sensitivity analysis to assess the robustness of the linear model, using a negative binomial regression for overdispersed count data. The formula used was:

$$CountDeceptiveHealthAds \sim offset(log(CountAds))$$
$$Age + Gender + NumHealthSites + Circulatory + Musculoskeletal +$$
$$Digestive + Genitourinary + Neoplasms + Endocrine + Skin +$$
$$Ear + HealthServices + Respiratory$$

Table 17 shows the results of the regression. We found that age and profiles of participants with neoplasm conditions had a significant effect. We report the results of the negative binomial regression in the body of the paper.