Prevalence of Third-Party Tracking on Medical Journal Websites

Ravi Gupta, MD[1,2,3]; Ari B. Friedman, MD, PhD[3,4,5]; Matthew S. McCoy, PhD[3,5]

[1] National Clinician Scholars Program at the University of Pennsylvania and Corporal Michael J. Crescenz VA Medical Center, Philadelphia, PA

[2] Division of General Internal Medicine, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

[3] Leonard Davis Institute of Health Economics, University of Pennsylvania, Philadelphia, PA

[4] Department of Emergency Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

[5] Department of Medical Ethics and Health Policy, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

Corresponding Author: Ravi Gupta, MD, National Clinician Scholars Program, University of Pennsylvania, 423 Guardian Drive, 1300 Blockley Hall, Philadelphia, PA 19104 (ravi.gupta@pennmedicine.upenn.edu)

Manuscript Word Count: 650

References: 6

Revision Date: 1/21/2022

Web tracking tools allow third-party advertisers to compile detailed information about individuals based on their browsing behavior.[1] Prior research has found that third-party tracking is prevalent on consumer-facing health information websites,[1] eliciting concerns about privacy risks for patients. Tracking on medical journal websites raises unique ethical and policy considerations because it may help pharmaceutical companies and health care advertisers profile clinicians based on which articles they access. This information can be used to serve clinicians advertising targeted to medical specialties and areas of professional interest inferred from their browsing histories, potentially contributing to undue pharmaceutical industry influence on clinical practice.[2] Thus, we investigated web tracking prevalence and characteristics on medical journal websites.

Methods

In this cross-sectional study, we identified all journals with an impact factor of ≥2.0 in clinically relevant subcategories of the Web of Science's life sciences and biomedical category.[3] We visited each journal's home page using webXray, a tool that detects third-party tracking on websites.[1] For each journal's home page, we recorded third-party data requests, which are significant because they initiate data transfers from a user's computer to third parties. We also recorded the presence of third-party cookies, data stored on a user's computer that frequently serve as persistent identifiers, enabling third parties to track individuals across multiple websites.

We calculated the percentage of journals with a third-party data request or cookie and the median number of data requests and cookies per journal homepage, overall and by journal impact factor. We calculated the most prevalent tracking entities across all webpages. We performed Google searches for the top 5 most prevalent tracking entities' advertising policies and marketing segment disclosures to determine whether they allowed pharmaceutical advertising and medical profession-specific ad targeting. Data were analyzed in October 2021. This study followed the

Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline for cross-sectional studies. Analyses were conducted in Stata version 16 (StataCorp LP).

Results

Overall, 1,599 of 1,605 (99%; 95% CI, 99%-100%) medical journal home pages included a third-party data request and 1,239 (78%; 95% CI, 75%-80%) included a third-party cookie, without significant differences by impact factor (**Table**). Journal home pages had a median of 8 third-party cookies (IQR, 1-17) with no differences by impact factor. The median number of third-party data requests per journal was 32 (IQR, 12-45), with higher impact journals significantly associated with fewer third-party requests ($P<0.001$).

Nearly all (99%) journal home pages included a data request from a third-party entity owned by Alphabet, Google's parent company. Data requests from entities owned by Twitter, Facebook, Oracle, and Adobe occurred on at least 40% of journal home pages (**Figure**). Marketing segment disclosures were found only for Oracle[4] and Adobe Inc.[5], both of which allow medical profession-specific ad targeting. However, all 5 top tracking entities allow pharmaceutical advertising.

Discussion

This study demonstrated that 99% of medical journals with an impact factor ≥2.0 expose visitors to third-party tracking by entities that work with pharmaceutical advertisers. Though similar levels of tracking have been found in health-related websites, tracking on journal websites raises distinctive policy concerns because it may facilitate targeted advertising to clinicians. While targeted advertising can increase knowledge of new therapeutics, it can also sway clinicians' prescribing patterns towards therapies with limited evidence of efficacy and cost-effectiveness.[6]

This study had limitations. Results may not be generalizable to medical journals with an impact factor ≤2.0. Additionally, marketing segment disclosures for 3 of 5 top tracking entities could not be located, and those that were located may be out of date. Finally, this study did not assess how accessing articles through library proxies may alter tracking.

Individual clinicians attempting to limit tracking may be insufficient.[3] Given growing concerns over digital health privacy risks and pharmaceutical advertising to clinicians, medical journal editors and publishers should monitor and assess the potential impact of third-party tracking on journal websites. Further research is needed to determine how tracking information influences targeted advertising to clinicians.

Table. Prevalence of Third-Party Tracking on Medical Journal Websites

	Overall	Impact Factor				P value
		2-5	**5.01-10**	**10.01-15**	**15.01-24**	
No. (%) of websites	1605	1278 (80)	236 (15)	40 (2)	51 (3)	<0.01
Websites, No (%) [95% CI]						
With a third-party data request	1599 (100) [99-100]	1273 (100) [99-100]	236 (100) [98-100]	39 (98) [87-100]	51 (100) [93-100]	0.11
With a third-party cookie	1239 (78) [75-80]	978 (77) [75-79]	192 (81) [76-86]	32 (82) [69-95]	37 (73) [60-85]	0.32
Third-party cookies per journal home page, median (IQR)	8 (1–17)	8 (1–21)	8 (1.5–14)	8 (1–22)	8 (0–14)	0.94
Third-party requests per journal home page, median (IQR)	32 (12–45)	33 (10-47)	26.5 (16-42)	20 (13-41)	19 (5-33)	<0.001

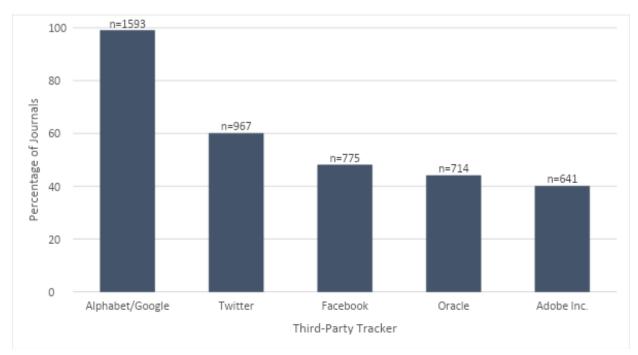Abbreviation: CI, confidence interval; IQR, interquartile range

Figure. Most Prevalent Third-Party Tracking Entities on Medical Journal Websites

[a] Axis represents the percentage of medical journal websites with a third-party data request from a given

tracking entity.

References

1.   Libert T. Privacy implications of health information seeking on the web. *Communications of the ACM.* 2015;58(3):68-77.

2.   Manz C, Ross JS, Grande D. Marketing to physicians in a digital world. *N Engl J Med.* 2014;371(20):1857-1859.

3.   Friedman AB, Miller E, McCoy MS. Prevalence of Medical Journal Websites That Deny Access to Users Who Block Browser Cookies. *JAMA Netw Open.* 2021;4(3):e213492.

4.   Oracle. Health and Wellness Preference Data Segments. 2021; https://www.oracle.com/us/assets/health-wellness-data-segments-2537888.pdf. Accessed December 1, 2021.

5.   Adobe Inc. Adobe Advertising Cloud - Available Third Party Health Segments 2019. 2019; https://www.adobe.com/content/dam/acom/en/privacy/pdfs/Adobe-Advertising-Cloud-Health-Segments-2019.pdf. Accessed December 1, 2021.

6.   Sinha MS, Kesselheim AS, Darrow JJ. Pharmaceutical Advertising in Medical Journals: Revisiting a Long-Standing Relationship. *Chest.* 2018;153(1):9-11.