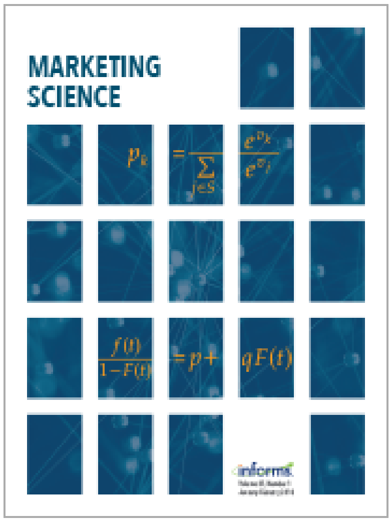

Marketing Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Frontiers: How Effective Is Third-Party Consumer Profiling? Evidence from Field Studies

Nico Neumann, Catherine E. Tucker, Timothy Whitfield

MARKETING SCIENCE
Vol. 38, No. 6, November–December 2019, pp. 918–926
ISSN 0732-2399 (print), ISSN 1526-548X (online)
http://pubsonline.informs.org/journal/mksc

Frontiers: How Effective Is Third-Party Consumer Profiling? Evidence from Field Studies

Nico Neumann,[a] Catherine E. Tucker,[b,c] Timothy Whitfield[d]

[a] Melbourne Business School, Carlton, Victoria 3053, Australia; [b] Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142; [c] National Bureau of Economic Research, Cambridge, Massachusetts 02138; [d] Burst SMS, Sydney, New South Wales 2000, Australia

Contact: n.neumann@mbs.edu, http://orcid.org/000-0003-3094-6238 (NN); cetucker@mit.edu, http://orcid.org/0000-0002-1847-4832 (CET); twhitfie@gmail.com (TW)

Received: January 29, 2019
Revised: April 24, 2019
Accepted: May 14, 2019
Published Online in Articles in Advance: October 2, 2019

https://doi.org/10.1287/mksc.2019.1188

Abstract. Data brokers often use online browsing records to create digital consumer profiles that they sell to marketers as predefined audiences for ad targeting. However, this process is a "black box"—little is known about the reliability of the digital profiles that are created or of the audience identification provided by buying platforms. In this paper, we investigate using three field tests the accuracy of a variety of demographic and audience-interest segments. We examine the accuracy of more than 90 third-party audiences across 19 data brokers. Audience segments vary greatly in quality and are often inaccurate across leading data brokers. In comparison with random audience selection, the use of black box data profiles, on average, increased identification of a user with a desired single attribute by 0%–77%. Audience identification can be improved, on average, by 123% when combined with optimization software. However, given the high extra costs of targeting solutions and the relative inaccuracy, we find that third-party audiences are often economically unattractive except for higher-priced media placements.

History: K. Sudhir served as the editor-in-chief and Puneet Manchanda served as associate editor for this article. This paper was accepted through the *Marketing Science*: Frontiers review process.
Funding: This work was supported by a National Science Foundation CAREER Award [6923256].
Supplemental Material: Data and the online appendix are available at https://doi.org/10.1287/mksc.2019.1188.

Keywords: digital advertising • data brokers • profiling • algorithms • machine learning • big data

1. Introduction

In the digital era, data has often been described as the "new oil" or "new gold" (*The Economist* 2017). The vast majority of online data are collected via cookies, which are placed on a wide variety of websites by third-party data brokers, such as Acxiom or Eyeota, often with the goal of profiling consumers. For example, 90% of the 500 top websites sent information about their visitors to at least one third party in 2016 (Lerner et al. 2016). The data brokers synthesize such consumer browsing information into anonymized user profiles and then apply proprietary heuristics or machine learning to make inferences about consumers. For example, a person could be identified as female by whether that user profile had browsed beauty or makeup websites. Age could similarly be inferred by whether that user profile had previously browsed retirement websites. This process allows the creation of predefined audiences, such as "sports interested," or "males 25–35." The resulting third-party prepackaged audiences are sold to advertisers to allow targeting digital ads to new consumers with whom an organization has no relationship yet and, hence, has no data.

Investment in third-party targeting services and solutions is estimated at $19.2 billion for the United States alone (IAB and WinterberryGroup 2018). Despite this substantial investment, the exact data sources and profiling processes used to create the predefined audiences are secret, and their reliability is unknown. As a result of this black box creation process, buyers are uncertain about quality: as *New York Times* CEO Mark Thompson asks, "When we say a member of the audience is a female fashionista aged 20 to 30, what's the probability that that's actually true?" (Kelly 2017).

To empirically assess the accuracy of the digital profiles and the performance of the overall audience-delivery process, we carry out three large-scale field tests. We investigate 19 leading data brokers and six buying platforms while looking at more than 90 third-party segments of some of the most popular audience data types: demographic and audience-interest attributes.[1]

In study one, we run an online campaign, which allows optimization of third-party audience selection, and assess whether the ad was seen by the requested demographic segment. In study two, we narrow our focus and simply look directly at whether data brokers are able to accurately determine the age and gender of a specific pair of eyeballs. In study three, we extend our data quality assessment from demographic (e.g., "age 25–34") to audience-interest segments, such as people interested in travel. Table 1 provides a summary.

2. Study One

The objective of study one is to examine the performance of a typical digital advertising campaign using the combined services of data brokers and ad-buying platforms (so-called demand side platforms or DSPs) to deliver ads to a specified audience. In a nutshell, DSPs optimize online campaigns and help select the websites on which ads are placed and the data sources (for details, see Online Appendix A.1).

2.1. Method

Study one was conducted in the first quarter of 2016 in collaboration with a major advertising agency and six DSPs. Because every platform has a different interface, we asked the managed service team of each platform to execute the campaign optimization using their own proprietary technology. This approach reduces concerns that performance discrepancies are driven by differences in interface knowledge. We asked the six DSPs to run a charity campaign in Australia according to a three-part instruction for demographic attributes described in Table 2. Each provider had full authority as to how they selected data sources and website placements to deliver the campaign.

To validate the demographic characteristics of the audiences that were exposed to the ads, we rely on Nielsen digital ad ratings (DAR), which uses both its panel and unique access to Facebook data. We also gather control data from Moat, a leading fraud and brand safety provider, on whether there is nonhuman traffic on the websites on which ads are displayed and the extent to which the websites are "brand-safe," for example, whether they have sexual images. Table 3 summarizes these performance criteria.

2.2. Results

Table 4 suggests significant performance differences among the audience delivery platform providers. Average audience targeting accuracy is 59%. The best provider is able to show ads to the right target market 72% of the time, and the worst provider shows ads to the right market 40% of the time. People saw between 1% and 41% more ads than specified in the brief. Brand safety scores range from 74.4% to 99.8%; the percentage of invalid impressions ranges from 1.4% to 6.5% across the six providers.

2.3. Discussion

There are two key observations in study one.

First, the performance of the automated audience delivery appears disappointing, with an average of 41% of impressions being off target. We compared the increase in audience identification with the natural distribution of the two characteristics. Male Internet users aged 25-54 should make up about 26.5% of the corresponding online users (Statista 2015). This suggests improvement in audience identification relative

Table 1. Three Field Studies and Their Focus

Study	Key question	Attribute focus	Task
1	Can programmatic ad campaigns deliver ads to the right target audience?	Demographic attributes	Identify the right cookie profiles to target for a digital campaign
2	Can data brokers accurately determine characteristics of individual online users?	Demographic attributes	Determine user characteristics based on cookies for which the data broker has data
3	Can data brokers accurately determine characteristics of individual online users?	Audience-interest attributes	Determine user characteristics based on cookies for which the data broker has data

Table 2. Study One: Campaign Criteria Given to Ad Platform

Criteria	Detail
Prespecified audience	Males between the age of 25 and 54
Campaign size	100,000 advertising impressions. Each time a display ad is shown on a website to a user, this counts as an impression.
Frequency	As many unique users as possible. Each user should see one impression, rather than one user seeing multiple impressions.

Table 3. Study One: Variable Definitions

Metric	Explanation
Accuracy	Percentage of impressions that were delivered to an audience that identified as male between 25 and 54 years old
Frequency	Average frequency of the campaign or how many impressions each viewer saw
Brand safe	Percentage of impressions that were served in a brand-safe environment
Nonhuman	Percentage of invalid (bot) impressions

to randomly selecting impressions of about 123% (0.59/0.265 = 2.23), on average.

When framed as offering a relative improvement of 123% rather than having a success rate of around 59%, the use of digital audience delivery seems promising. However, this 123% improvement in accuracy relative to a baseline of delivering randomly to the total population should be set against any additional costs. We discuss the cost–benefit ratio in a later section.

Second, audience accuracy varies significantly across the DSPs. At least some of the variation seems to be linked to quality differences in the buying technology of the DSPs and their managed service teams as all our campaign performance criteria suggest a similar ranking.[2]

3. Study Two

Study one used optimization software, DSPs, to select data sources as well as ad placements. The performance we observe could be driven by the skill in audience selection by the platforms, by the quality of the profiles created by the data brokers, or other unobservable reasons. In our second study, we focus only on the accuracy of data brokers while investigating the same demographic attributes: age and gender.

3.1. Method

For study two, we obtained access to a globally leading data management platform (DMP), which was integrated with another high-quality panel survey, Pureprofile, which is ISO best-practice certified. This setup allows assessing the accuracy of classification of cookies by linking the data brokers' cookies (in the DMP, which basically serves as a connection gateway to data brokers) to one user profile of the panel (for details, see Online Appendix A.3).

First, we look at the ability of data brokers to identify audiences that are male and between 25 and 54 (in line with the brief in study one; see Table 5). This test enables us to compare the audience results with and without using buying platforms and managed services. Then, we examine the accuracy of the attributes individually to better understand potential differences in data quality (see Table 6). For age, we were able to get a sample of the three most popular age tiers: 18–25, 25–34, and 35–44.[3]

Table 4. Study One: Campaign Performance Results

DSP	Accuracy, %	Frequency	Brand safe, %	Nonhuman, %
1	72	1.01	99.8	1.4
3	68	1.20	98.4	2.4
2	66	1.03	92.9	2.8
4	57	1.15	89.3	4.1
5	40	1.41	84.3	5.0
6	50	1.13	74.4	6.5
Average	59	1.15	89.9	3.7

Notes. Demand side platform (DSP) identities are anonymized. Accuracy refers to identifying males between the age of 25 and 54. See Table 3 for precise definitions.

3.2. Results

We find that the average accuracy in identifying males between 25 and 54 is 24.4%. Given the natural distribution of the two attributes is 26.5%, the relative average performance of using third-party data according to our sample is worse than random user selection.

The results for individual attributes show high variation in audience accuracy across data brokers for both age and gender. Gender accuracy ranges from 25.7% to 62.7% with an overall average of 42.3% (see Table 6). Given the benchmark for correct gender classification is about 50%, or the natural distribution of gender diversity in the population,[4] using data brokers to assess online browsing profiles for gender appears, on average, less efficient than using nothing.

In contrast, age precision ranges from 4.3% to 42.5% for our tested data brokers and age tiers. The average accuracy for age 18–25 is 10.7%, for age 25–34 is 25.7%, and for age 35–44 is 32% (see Table 7). According to

Table 5. Study Two: Data Broker Accuracy for Joint Identification of Gender (Male) and Age (25–54)

Data broker	Accuracy, %	Sample size
Vendor A	12.9	319
Vendor D	32.0	388
Vendor E	27.1	63
Vendor F	32.2	90
Vendor G	27.1	155
Vendor I	14.8	1,782
Vendor J	24.1	9,004
Vendor K	12.3	253
Vendor L	22.2	63
Vendor M	20.9	129
Vendor N	42.4	1,392
Average	24.4	1,239.8

Notes. For the majority of gender–age combinations, we were only able to compare the accuracy for males 25–44 instead of males 25–54 as we did not have the right age tier available. For these cases, we discarded the missing age-range data to provide conservative estimates as a comparison with study one.

Table 6. Study Two: Data Broker Accuracy for Gender (Male)

Data broker	Accuracy, %	Sample size
Vendor A	27.5	1,396
Vendor B	25.7	408
Vendor C	35.2	1,777
Vendor D	56.4	495
Vendor E	48.8	527
Vendor F	47.9	480
Vendor G	46.8	562
Vendor H	33.2	1,016
Vendor I	33.6	2,336
Vendor J	42.4	14,342
Vendor K	30.6	346
Vendor L	51.9	547
Vendor M	49.1	456
Vendor N	62.7	5,099
Average	42.3	2,127

Statista (2015), 18- to 24-year-olds should make up about 10% of the online user population; 25- to 34-year-olds and 35- to 44-year-olds each make up about 18% of internet users. Hence, using third-party data for our age audiences, on average, appears to provide an efficiency improvement of around 42% (7% for 18–24, 42.7% for 25–34, for 77.0% 34–44) in reaching the desired audience compared with using no targeting.

3.3. Results Extension: Gender Accuracy for Different Household Types

Our panel provider Pureprofile collects information about whether a household has children. We use this variable as a proxy for smaller and larger households and examine the accuracy of the gender attribute for the two different household types (see Table 8). The average accuracy for households with children is 37.2% and without children is 51.4%. This difference is statistically significant ($M = 14.2$, $t = 333.7$, $p < 0.001$).

We may draw two conclusions. First, having a larger number of people in a household tends to decrease accuracy in identifying the correct characteristics of individuals, such as gender. We assume this reflects multiple people sharing the same devices to go online in a household. Hence, some of the profiling errors can be attributed to the fact that several individuals may share online devices in a household with several members.[5] Second, although households without children have a significantly higher accuracy than those with children, the overall hit rate of 51.4% is still only marginally better than random guessing.

3.4. Discussion

Study two shows that the audience accuracy varies greatly for all tested attributes of our sample of 14 data brokers. Total accuracy (the hit rate) ranges from 4.3% to 62.7% for our data. Using digital audiences rather than random user selection leads, on average, to no improvement for gender alone or an audience described by gender and age, and it leads to an improvement of 7%–77% for age-tier classifications. The greater classification efficiency for age tiers in comparison with gender is surprising as there should be fewer mistakes with attributes with fewer degrees of freedom. However, it may well be that the web activity of consumers is a better indicator of age than of gender.[6]

Overall accuracy is also still disappointing for households with and without children. Thus, there must be additional factors driving a data broker's audience precision besides household size. One reason could be a lack of sufficient integrated websites to classify users based on cookies (Trusov et al. 2016) or profiling challenges as a result of cookie and mobile identifier mismatches (Coey and Bailey 2016, Lin and Misra 2018).

Table 7. Study Two: Data Broker Accuracy for Different Age Tiers

Age tier	18–24		25–34		35–44	
Data broker	Sample size	Accuracy, %	Sample size	Accuracy, %	Sample size	Accuracy, %
Vendor A	226	8	217	30.9	285	42.8
Vendor D					32,724	20.7
Vendor E			211	32.2	367	39.8
Vendor G	155	7.7	221	36.7	341	44
Vendor I			32,769	18.0	1,711	22.1
Vendor J	9,537	11.1	10,849	18.8	8,904	23.6
Vendor K			62	30.6	33,303	20.7
Vendor L	68	10.3	141	15.6	157	36.3
Vendor M	93	4.3	290	20.0	271	33.2
Vendor N	2,521	22.8	2,825	28.8	1,214	36.2
Average	2,100	10.7	5,061	25.7	7,928	32.0

Note. Empty cells mean that the data broker did not have a comparable segment for the corresponding age tier we chose for analysis.

Table 8. Gender (Male) Accuracy for Households with (HHC) and Without Children (HHNC)

Data broker	Accuracy, %	Sample all	Accuracy HHNC, %	Sample HHNC	Accuracy HHC, %	Sample HHC
Vendor A	27.5	1,396	35.7	263	22.4	545
Vendor B	25.7	408				
Vendor C	35.2	1,777	38.4	352	30.4	717
Vendor D	56.4	495	54.8	126	52.3	153
Vendor E	48.8	527	62.9	97	38.1	218
Vendor F	47.9	480	54.3	105	43.5	170
Vendor G	46.8	562	60.4	101	36.4	225
Vendor H	33.2	1,016	44.8	181	28.0	403
Vendor I	33.6	2,336	34.5	473	30.4	940
Vendor J	42.4	14,342	43.7	3,252	39.4	5,725
Vendor K	30.6	346	46.6	58	21.3	122
Vendor L	51.9	547	63.9	97	43.1	216
Vendor M	49.1	456	66.7	84	37.6	189
Vendor N	62.7	5,099	61.7	1,375	61.1	1,962
Average	42.3	2,128	51.4	505	37.2	891

4. Study Three

The relative improvement in audience identification when using third-party targeting seems small to moderate for the demographic attributes in study two. The question is whether this outcome is unique to demographic data. Although age and gender are currently the most widely used targeting attributes online, audience interest–based data represents the attributes for which advertisers anticipate the greatest growth in usage over the next two years (Salesforce 2018). We, therefore, repeat our data broker examination using interest-based audience data.

4.1. Method

The setup of study three is exactly the same as for study two, but this time, we selected the three most common audience-interest segments from the data management platform: "sports interested," "fitness interested," and "travel interested." Specifically, someone would count as sports interested if the person indicated in a survey that the person plays any kind of sports, follows any kind of sports, or attends sports events or directly indicated that the person wishes to read about sports content. To be categorized as fitness interested, a user would need to indicate that the user was interested in fitness content. Similarly, someone would be travel interested if the person indicated a desire to travel at least once, either for business or leisure, or a wish to read about travel content. The results of the data broker validation through the Pureprofile panel are summarized in Table 9.

Table 9. Study Three: Data Broker Accuracy for Audience Interests

	Fitness interested		Sports interested		Travel interested	
Data broker	Accuracy, %	Sample size	Accuracy, %	Sample size	Accuracy, %	Sample size
Vendor A			86.2	571	64.7	697
Vendor B			91.0	1,428	64.0	2,564
Vendor C	81.2	611			74.0	704
Vendor D	78.6	117			83.5	127
Vendor E			89.6	4,371	87.8	1,753
Vendor F	82.1	196	86.0	285	67.5	243
Vendor G	83.2	393	86.3	729		
Vendor H	82.3	327				
Vendor I	82.4	307				
Vendor J			89.5	8,772	78.2	10,936
Vendor K			82.8	128	58.9	124
Vendor L			86.7	360	62.4	412
Vendor M	85.9	199	86.7	495	63.8	574
Vendor N			89.9	5,039	77.5	9,846
Vendor O	80.7	405	89.9	4,459	82.4	9,380
Vendor P			89.6	4,371	87.8	1,753
Vendor Q			86.9	604	67.5	499
Vendor R			82.1	168	78.2	10,904
Vendor S					65.9	857
Average	82.1	320	87.4	2,270	72.8	3,211

4.2. Results

Our validation tests for the three audience-interest audiences show a high total accuracy (hit rate) with an average of 87.4% for sports interested, 82.1% for fitness interested, and 72.8% for travel interested. There is still some variation in accuracy across data brokers for the travel audiences (ranging from 62.4% to 87.8%) but less so for sports (ranging from 82.1% to 91%) and fitness audiences (ranging from 78.6% to 85.9%).

The next question is what the odds are that someone in the population is interested in travel, sports, or fitness if we just distribute ads randomly. We obtained numbers from various published sources (detailed in Online Appendix A.3) that suggest 56% of Australians are interested in travel, 67% sports, and 48% fitness. This suggests that, on average, using third-party audiences to reach interest groups improves targeting for our data by 30% (72.8/56 = 1.3), 30% (87.4/67 = 1.3), and 71% (82.1/48 = 1.71), respectively, relative to showing ads randomly.

4.3. Discussion

Overall, we find higher hit rates (accuracy) for our tested audience interests than for our previously tested demographic attributes. With regards to relative improvement, the audience-interest segments, on average, increase the correct identification of the target audiences for our data by 30%–71%. Therefore, the range of relative improvement in comparison with using no audience data for our three audience-interest segments is similar to the one we have seen for the demographic audiences in study two (average accuracy increase by 7%–77%).

Some of our examined attributes (e.g., sports interest audiences) have high baselines, which naturally limit relative improvements because the maximum accuracy can only be 100%. However, interest segment baselines of around 50% allow a direct comparison with gender, our most solid baseline. Moreover, our low hit rates for travel interest audiences (plus an additional test on two fashion interest audiences in Online Appendix A.5) illustrate that the performance results (0%–77%) hold across many attributes independent of the baseline.

5. A Cost–Benefit Analysis

Companies typically use targeting for marketing-communication purposes to reduce wasted ad spending. To understand the benefits that advertisers receive from using digital audiences, we estimated the relative improvement in accuracy in relation to the odds of finding the desired attribute naturally in the population. Table 10 shows that we find between 0% and 123% average improvements for the use of third-party audiences across our three studies.

In Table 11, we summarize the various cost components of leveraging third-party digital audiences. Total costs comprise a mix of fixed and variable (percentage) costs and were taken from several industry sources (see Online Appendix A.4 for details). In particular, the third-party audience information is a fixed cost that is added to the cost-per-*mille* (CPM) of online ads.

As a result, the final cost ratio of using audience solutions versus not using them strongly depends on the price of the publisher's ad placement. For example, standard display banner ads in Australia or the United States have average CPMs of around $4.20 (see Online Appendix A.4) and would result in a cost ratio of 2.51. That is, third-party audience optimization would result in extra costs of 151%. However, when the ad slots on a publisher site are used for more expensive media, such as online video ads with average CPMs of $18.92 (see Online Appendix A.4), the cost ratio of using audience solutions versus ad buys without targeting decreases to 1.58 (58% extra cost).

If we now compare the cost–benefit ratio for the two types of media, we see that, for standard display banner ads, the additional costs of 151% are higher than the average additional gain of 123% in audience

Table 10. Data Broker Performance Across Studies

	Sample of data brokers	Data broker hit rate, %	Population with attribute, %	Ratio hit rate to population odds	Study
Gender and age optimized	6	59	26.5	2.23	1
Gender and age	11	24.4	26.5	0.92	2
Gender	14	42.3	50	0.85	2
Age 18–24	6	10.7	10	1.07	2
Age 25–34	9	25.7	18	1.43	2
Age 35–44	10	32.0	18	1.77	2
Sport	14	87.4	67	1.30	3
Fitness	8	82.1	48	1.71	3
Travel	16	72.8	56	1.30	3
Average single attributes	11	50.5	38.1	1.35	

Table 11. Cost Components for Using Digital Audience Solutions for Different Media in Dollars

	Display ad		Video ad	
	Targeting	No targeting	Targeting	No targeting
Publisher	1.36	1.36	11.00	11.00
SSP/exchange[a]	0.13		1.09	
Third party data costs	1.33		1.33	
Ad serving and verification	0.20	0.20	0.20	0.20
DSP	0.44		2.00	
Trading desk/execution	0.45		2.04	
Agency of record	0.27	0.11	1.24	0.78
Final cost advertiser	4.20	1.67	18.90	11.98
Cost ratio to no targeting	2.51		1.58	

[a]A supply-side platform (SSP) is a technology platform that enables web publishers and digital media owners to manage their advertising space inventory and sell ads through algorithmic optimization (Hof 2014).

identification. For online video ads, the average relative extra costs of 58% would be much lower than the average additional gain of 123% in audience identification. Hence, using third-party audience solutions seems economically viable for more expensive media placements that dictate higher CPMs, such as online video.

6. Implications

6.1. Summary

Using proprietary methods that are typically a black box, data brokers classify users based on cookies and browsing behavior (Bucklin and Sismeiro 2003, Park and Fader 2004) and sell these data profiles to advertisers for purposes of ad targeting. We empirically examine in three field tests the accuracy of the digital profiling and audience delivery process for third-party data using first-party, self-reported data for validation.

Across our tests, we look at two demographic attributes (age and gender) and three audience-interest segments (sports, travel, and fitness interest) and more than 19 different data brokers (resulting in more than 90 validated digital audiences). Study one tests the performance of the entire audience delivery process, including optimization software that helps select ad placements and data sources. For this process and our two tested demographic attributes, we find an average accuracy of 59%. This result corresponds to an average improvement of 123% in audience identification compared with using no third-party audiences or showing ads with no targeting. In study two, we show that, if we just focus on the underlying audiences that are offered by data brokers for the same two attributes, we find that the audience identification is, for many data brokers, worse than random user selection (on average, 24.4%).

When investigating gender (being male) and age (three different tiers: 18–24, 25–34, and 35–44 years) individually, we find that digital audiences for gender are, on average, less often correct than random guessing (accuracy of 42.3%). Age accuracy depends on the chosen age tier with an average of 10.7% for 18- to 24-year-olds, 25.7% for 24- to 35-year-olds, and 32% for 35- to 44-year-olds). This means that third-party age-tier data leads to an average improvement in audience identification between 7% and 77% in comparison with random user selection. For fitness, travel, and sports interest audiences, we find an average accuracy of 82.1%, 72.8%, and 87.4%. These findings correspond to an average improvement in audience identification of 30%–71% (in comparison with random user selection), which is similar to the range of age audience data.

Audience identification can even be improved, on average, by 123% when marketers additionally use optimization software (DSPs) that helps select the best ad placements and vendors. However, although the final cost–benefit ratio depends on the choice of DSP and the experience of the person running the campaign, the relative extra costs for the various supporting technologies are so often so high that these may outweigh any efficiency gains (e.g., on average, further third-party audience costs of 151% for display banners).

6.2. Limitations and Future Research Direction

Our study is subject to possible limitations. First, our research relies on the success of our validation efforts. We used two well-established panel providers and self-reported data to validate third-party audiences: Nielsen DAR, which has unique access to a global panel and Facebook data, and Pureprofile, which has strict control tests in place as well as ISO best-in practice certification for its services.[7] Although user-reported, first-party data are often regarded in practice as more reliable than third-party data that was aggregated in unknown ways and from unknown sources, we acknowledge that some users may distort information too, leading to possible classification errors.

Second, the estimates of our relative improvements in comparison with using no targeting depend on the choice of natural population distributions, which are hard to define for abstract attributes such as interests. We, however, attempted to rely on conservative baseline estimates to avoid any bias (see Online Appendix A.3).

Third, our cost data represents averages only; actual cost data and cost–benefit ratios strongly depend on the specific media buys and contracts. Every organization is encouraged to check its own cost–benefit ratio and should see our estimates as approximate guidelines.

Fourth, our analysis was restricted by cookie data for which we have sufficient data brokers to test and external data with which to validate it.

Fifth, to the best of our knowledge, all studies included data retrieved through mobile web-based and desktop browsing. However, the provided data did not allow us to specifically distinguish between mobile and desktop PC effects. Investigating any potential differences resulting from the different use and characteristics of these basic device types is a worthwhile undertaking for future studies.

Likewise, we find strong differences in average audience accuracy for gender and age audiences between studies one and two. We can only speculate about the possible reasons behind the performance differences, which could be linked to the different sampling procedures of Nielsen DAR across DSPs or the website and data broker selection of the DSP itself. Future research efforts may help further explain the greater efficiency that can be achieved through campaign optimization software.

Notwithstanding these possible limitations, we believe our paper is a useful first step in calibrating the degree of successes and misclassification in third-party audience profiles.

6.3. Contribution

This paper makes academic and managerial contributions.

In terms of our academic contribution, targeting different customer segments with different marketing messages is at the core of marketing (Narayanan and Manchanda 2006). If firms wish to communicate with new prospective customers or don't have any data on their own customers, they need to obtain data elsewhere to target appropriately. Theoretical work has investigated incentives across stakeholders in data sharing (Murthi and Sarkar 2003, Bergemann and Bonatti 2015) and the consequences of imperfect data (Chen et al. 2001). Empirical work has investigated the incentives for customer data intermediaries in offline settings to maximize data availability (Pancras and Sudhir 2007). More recently, Coey and Bailey (2016), Trusov et al. (2016), and Lin and Misra (2018) have investigated how data fragmentation and incomplete browsing information restrict consumer-profiling accuracy in online settings. In a similar vein, Kim et al. (2005) and De Bruyn and Otter (2019) discuss new algorithmic methods to improve customer segmentation. These studies reveal individual methodological and technological challenges for online data profiling, but little is known about the quality of digital audiences and the economic consequences of using third-party solutions, which is the focus of our paper.

Regarding our managerial contribution, we illustrate the risks of using black box consumer profiling and outline possible negative consequences of unverified data products for advertising.[8] We document the large heterogeneity in audience accuracy across data brokers and DSPs, thus highlighting how important it is to select the right data supplier and buying platform. Without experimentation, the audience quality is hard to assess because of the lack of transparency and available benchmarking statistics.

Because of the questionable economics for some ad placements and the difficulty in assessing audience quality, managers should carefully consider whether leveraging third-party audiences makes sense given their media mix and market experience. Of course, advertisers could also improve the economics by reducing any technology and service costs. For example, they could manually select data suppliers (saving DSP fees) or execute media buys in house (saving trading-desk fees). Media buyers who wish to use some form of audience data but may not have the knowledge to run digital campaigns themselves or are likely to face poor cost–benefit ratios may achieve more accuracy using their own first-party data.[9]

Finally, several industry bodies, such as the IAB and the Association of National Advertisers, have proposed a data-labeling initiative for 2019, similar to nutrition labels for food (IAB 2018). The data labels' goal is to increase transparency and help marketers understand on what information digital audiences are based. Our research underscores the need for such actions and initiatives. As advertising is largely unregulated and any data labeling of audiences would be voluntary, our results show that advertisers should carry out their own validation tests and consider enforcing transparency in media buys whenever possible.

Acknowledgments

The authors are grateful for generous support of this study from Pureprofile, Moat, Nielsen, Sizmek, and AppNexus. They also thank Bernd Skiera, Garrett Johnson, Ujwal

Kayande, and Gerardo Berbeglia for their helpful comments; as well as participants at the 2017 Marketing Science Conference; and the research seminars at the University of Bologna, Goethe University, HEC Paris, Kings College London, London Business School, and University College London. Finally, the authors thank the review team of *Marketing Science* for the helpful feedback. All errors are the authors' own.

Endnotes

[1] Two recent surveys of brand marketers suggest that, even though marketers buy a wide range of audiences, including behavioral and location data, the most popular digital information purchased by the majority of advertisers is the basic demographic data of age and gender (Lotame 2018, Salesforce 2018).

[2] We find significant correlations between audience accuracy and frequency ($r = -0.79$, $t = -2.58$, $p < 0.04$), audience accuracy and brand safety ($r = 0.80$, $t = 2.67$, $p < 0.03$) audience accuracy and non-human impressions ($r = -0.86$, $t = -3.37$, $p < 0.02$).

[3] The data brokers have varying age classification ranges, for example, 18–25, 21–25, 20–29, and so forth, which is why it is difficult to find age buckets that allow tests across multiple data brokers.

[4] For example, in the United States, 89% of men are online and 88% of women (accessed April 30, 2019, https://www.statista.com/statistics/184415/percentage-of-us-adults-who-are-internet-users-by-gender/).

[5] We thank the editor for raising this point.

[6] We thank an anonymous reviewer for this comment.

[7] Pureprofile is also an official partner of the local IAB chapter for providing official ad-blocking statistics.

[8] Our empirical findings on accuracy (hit rate) are supported by anecdotes and mentions of poor targeting and incorrect user classifications by others: Flosi et al. (2013), De Bruyn and Otter (2019), and Mallazzo (2018).

[9] First-party data are often used in advertising methods, such as retargeting, in which, after someone has visited a website, users are then tracked and shown ads for products they browsed on that website (Lambrecht and Tucker 2013, Johnson et al. 2017, Sahni et al. 2017).

References

Bergemann D, Bonatti A (2015) Selling cookies. *Amer. Econom. J. Microeconom.* 7(3):259–294.

Bucklin RE, Sismeiro C (2003) A model of website browsing behavior estimated on clickstream data. *J. Marketing Res.* 40(3):249–267.

Chen Y, Narasimhan C, Zhang ZJ (2001) Individual marketing with imperfect targetability. *Marketing Sci.* 20(1):23–41.

Coey D, Bailey M (2016) People and cookies: Imperfect treatment assignment in online experiments. Bourdeau J, Hendler JA, Nkambou R, eds. *Proc. 25th Internat. Conf. World Wide Web (WWW '16)* (ACM Press, New York), 1103–1111.

De Bruyn A, Otter T (2019) Bayesian customer profiling. Working paper, ESSEC Business School, Cergy-Pontoise Cedex, France.

Economist, The (2017) Data are giving rise to a new economy—Fuel of the future. Accessed April 12, 2018, https://www.economist.com/news/briefing/21721634-how-it-shaping-up-data-giving-rise-new-economy.

Flosi S, Fulgoni G, Vollman A (2013) If an advertisement runs online and no one sees it, is it still an ad? Empirical generalizations in digital advertising. *J. Advertising Res.* 53(2):192–199.

Hof R (2014) OpenX aims to boost publishers' online ads with new SSP technology. *Forbes* (June 9), https://www.forbes.com/sites/roberthof/2014/06/09/openx-aims-to-boost-publishers-online-ads-with-new-ssp-technology.

IAB (2018) Major advertising trade bodies unveil data transparency label. Accessed November 30, 2018, https://iabtechlab.com/press-releases/major-advertising-trade-bodies-unveil-data-transparency-label.

IAB and WinterberryGroup (2018) The state of data. Accessed January 10, 2019, https://www.iab.com/insights/the-state-of-data-2018/.

Johnson GA, Lewis RA, Nubbemeyer EI (2017) Ghost ads: Improving the economics of measuring online ad effectiveness. *J. Marketing Res.* 54(6):867–884.

Kelly C (2017) Inaccurate segments may be costing advertisers billions. Accessed March 12, 2018, https://adexchanger.com/data-driven-thinking/inaccurate-segments-may-costing-advertisers-billions/.

Kim Y, Street WN, Russell GJ, Menczer F (2005) Customer targeting: A neural network approach guided by genetic algorithms. *Management Sci.* 51(2):264–276.

Lambrecht A, Tucker C (2013) When does retargeting work? Information specificity in online advertising. *J. Marketing Res.* 50(5):561–576.

Lerner A, Simpson AK, Kohno T, Roesner F (2016). Internet Jones and the raiders of the lost trackers: An archaeological study of web tracking from 1996 to 2016. Holz T, Savage S, eds. *Proc. 25th USENIX Security Sympos*. (USENIX Association, Berkeley, CA), 997–1013.

Lin T, Misra S (2018) Identity fragmentation bias. Working paper, University of Chicago, Chicago.

Lotame (2018) The new state of audience data: Accuracy matters. Accessed August 5, 2018, https://www.lotame.com/lotame-research-report-the-new-state-of-audience-data-accuracy-matters/.

Mallazzo M (2018) When did flawed data become OK? Accessed November 30, 2018, https://adexchanger.com/data-driven-thinking/when-did-flawed-data-become-ok/.

Murthi B, Sarkar S (2003) The role of the management sciences in research on personalization. *Management Sci.* 49(10):1344–1362.

Narayanan S, Manchanda P (2006) Heterogeneous learning and the targeting of marketing communication for new products. *Marketing Sci.* 28(3):424–441.

Pancras J, Sudhir K (2007) Optimal marketing strategies for a customer data intermediary. *J. Marketing Res.* 44(4):560–578.

Park Y-H, Fader PS (2004) Modeling browsing behavior at multiple websites. *Marketing Sci.* 23(3):280–303.

Sahni NS, Narayanan S, Kalyanam K (2017) An experimental investigation of the effects of retargeted advertising: The role of frequency and timing. *J. Marketing Res.* 56(3):401–418.

Salesforce (2018) Digital advertising 2020: Insights into a new era of advertising and media buying. Accessed August 15, 2018, https://www.salesforce.com/form/marketingcloud/digital-ads-2020-research.jsp?d=cta-body-promo-13.

Statista (2015) Australia: Age distribution of internet users 2015. Accessed January 10, 2018, https://www.statista.com/statistics/259828/age-distribution-of-internet-users-in-australia/.

Trusov M, Ma L, Jamal Z (2016) Crumbs of the cookie: User profiling in customer-base analysis and behavioral targeting. *Marketing Sci.* 35(3):405–426.