How Effective Is Third-Party Consumer Profiling And Audience Delivery? Evidence from Field Studies

Nico Neumann, Catherine E. Tucker and Timothy Whitfield*

June 12, 2019

Forthcoming in Marketing Science - Frontiers

Abstract

Data brokers often use online browsing records to create digital consumer profiles they sell to marketers as pre-defined 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 over 90 third-party audiences across 19 data brokers.

Audience segments vary greatly in quality and are often inaccurate across leading data brokers. In comparison to 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.

*Nico Neumann is an Assistant Professor and Fellow for the Centre for Business Analytics at Melbourne Business School. Catherine Tucker is the Sloan Distinguished Professor of Marketing at MIT Sloan School of Management, Cambridge, MA and Research Associate at the NBER. Timothy Whitfield is the Chief Technology Officer at Burst SMS. We are grateful for generous support of this study from Pureprofile, Moat, Nielsen, Sizmek and AppNexus. We 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 seminar at the University of Bologna, Goethe University, HEC Paris, Kings College London, London Business School, University College, London. Thank you to NSF CAREER Award 6923256 for financial support. Finally, we thank the review team of Marketing Science for the helpful feedback. All errors are our own.

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 is 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 percent 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 pre-defined audiences, such as 'sports interested,' or 'males 25-35.' The resulting third-party pre-packaged 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 US alone (IAB and WinterberryGroup, 2018). Despite this substantial investment, the exact data sources and profiling processes used to create the pre-defined 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 over 90 third-party segments of some

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

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 or not 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 where ads are placed and the data sources (for details see Online Appendix 7.1).

[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).

2.1 Method

Study One was conducted in the first quarter of 2016 in collaboration with a major advertising agency and six DSPs. Since 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.

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

Criteria	Detail
Pre-Specified Audience	Males between the age of 25-54
Campaign Size	100,000 advertising impressions. Each time a display ad is shown on 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.

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 non-human 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

Table 3: Study One: Variable Definitions

Accuracy	% 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.
BrandSafe	% of impressions that were served in a brand-safe environment
Non-Human	% of invalid (bot) impressions

Table 4: Study One: Campaign Performance Results

DSP	Accuracy	Freq.	BrandSafe	Non-Human
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%

DSPs identities are anonymized. Accuracy refers to identifying males between the age of 25-54. See Table 3 for precise definitions.

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 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 will 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 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, who are 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 7.3).

First, we look at the ability of data brokers to identify audiences that are male and between 25-54 (in line with the brief in Study 1, 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

[2]We find significant correlations between audience accuracy and frequency (r=-.79, t=-2.58, p <.04), audience accuracy and brand safety (r=.80, t=2.67, p <.03) audience accuracy and non-human impressions (r=-.86, t=-3.37, p <.02).

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%	1782
Vendor J	24.1%	9004
Vendor K	12.3%	253
Vendor L	22.2%	63
Vendor M	20.9%	129
Vendor N	42.4%	1392
Average	24.4%	1239.8

Note: 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 to Study 1.

differences in data quality (see Tables 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]

3.2 Results

We find that the average accuracy in identifying males between 25-54 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 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%. 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

[3]The data brokers have varying age classification ranges, e.g. 18-25, 21-25, 20-29 etc., which is why it is difficult to find age buckets that allow tests across multiple data brokers.

[4]For example, in the US 89% of men are online, and 88% of women. `https://www.statista.com/statistics/184415/percentage-of-us-adults-who-are-internet-users-by-gender/`

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

Data broker	Accuracy	Sample Size
Vendor A	27.5	1396
Vendor B	25.7	408
Vendor C	35.2	1777
Vendor D	56.4	495
Vendor E	48.8	527
Vendor F	47.9	480
Vendor G	46.8	562
Vendor H	33.2	1016
Vendor I	33.6	2336
Vendor J	42.4	14342
Vendor K	30.6	346
Vendor L	51.9	547
Vendor M	49.1	456
Vendor N	62.7	5099
Average	42.3	2127

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

Data Broker	18-24	Accuracy (%)	25-34	Accuracy (%)	35-44	Accuracy (%)
Vendor A	226	8	217	30.9	285	42.8
Vendor D					32724	20.7
Vendor E			211	32.2	367	39.8
Vendor G	155	7.7	221	36.7	341	44
Vendor I			32769	18.0	1711	22.1
Vendor J	9537	11.1	10849	18.8	8904	23.6
Vendor K			62	30.6	33303	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	2521	22.8	2825	28.8	1214	36.2
Average	2100	10.7	5061	25.7	7928	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.

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%. According to Statista (2015), 18-24-year-olds should make up about 10% of the online user population; 25-34-year-olds and 35-44-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 to using no targeting.

3.3 Results Extension: Gender Accuracy for Different Households Types

Our panel provider Pureprofile collects information about whether or not 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] Secondly, while 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

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

Table 8: Gender (male) accuracy for households with (HHC) and with no children (HHNC)

Data broker	accuracy	sample all %	accuracy HHNC %	sample HHNC	accuracy HHC %	sample HHC
Vendor A	27.5	1396	35.70	263	22.40	545
Vendor B	25.7	408				
Vendor C	35.2	1777	38.40	352	30.40	717
Vendor D	56.4	495	54.80	126	52.30	153
Vendor E	48.8	527	62.90	97	38.10	218
Vendor F	47.9	480	54.30	105	43.50	170
Vendor G	46.8	562	60.40	101	36.40	225
Vendor H	33.2	1016	44.80	181	28.00	403
Vendor I	33.6	2336	34.50	473	30.40	940
Vendor J	42.4	14342	43.70	3252	39.40	5725
Vendor K	30.6	346	46.60	58	21.30	122
Vendor L	51.9	547	63.90	97	43.10	216
Vendor M	49.1	456	66.70	84	37.60	189
Vendor N	62.7	5099	61.70	1375	61.10	1962
Average	42.3	2128	51.40	505	37.20	891

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, while it leads to an improvement of 7-77% for age-tier classifications. The greater classification efficiency for age tiers in comparison to 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 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 due to cookie and mobile identifier mismatches (Coey and Bailey, 2016; Lin and Misra, 2018).

4 Study Three

The relative improvement in audience identification when using third-party targeting seems small to moderate for the demographic attributes in Study 2. The question is whether this outcome is unique to demographic data. While age and gender are currently the most widely

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

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 they played any kind of sports, follows any kind of sports, attends sports events or directly indicated that they wished to read about sports content. To be categorized as 'fitness interested', a user would need to indicate that they were interested in fitness content. Similarly, someone would be 'travel interested' if they 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.

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 7.3) 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%

Table 9: Study 3: Data Broker Accuracy for Audience-Interests

	Fitness interested		Sports interested		Travel interested	
Data Broker	Accuracy (%)	Sample	Accuracy (%)	Sample	Accuracy (%)	Sample
Vendor A			86.2	571	64.7	697
Vendor B			91.0	1428	64.0	2564
Vendor C	81.2	611			74.0	704
Vendor D	78.6	117			83.5	127
Vendor E			89.6	4371	87.8	1753
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	8772	78.2	10936
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	5039	77.5	9846
Vendor O	80.7	405	89.9	4459	82.4	9380
Vendor P			89.6	4371	87.8	1753
Vendor Q			86.9	604	67.5	499
Vendor R			82.1	168	78.2	10904
Vendor S					65.9	857
Average	82.1	320	87.4	2270	72.8	3211

(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 to using no audience data for our three audience-interest segments is similar to the one we have seen for the demographic audiences in Study 2 (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 7.5) 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.

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 & age optimized	6	59	26.5	2.23	1
Gender & 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	

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 7.4 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

Table 11: Cost components for using digital audience solutions for different media (in $)

	Display ad		Video ad	
	Targeting	No targeting	Targeting	No targeting
Publisher	1.36	1.36	11.00	11.00
SSP/ exchange*	0.13		1.09	
3rd party data costs	1.33		1.33	
Ad serving & 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 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).

depends on the price of the publisher's ad placement. For example, standard display banner ads in Australia or the US have average CPMs of around $4.20 (see 7.4) 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 7.4), 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 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 over 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 to 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 18-24-year-olds, 25.7% for 24-35-year-olds and 32% for 35-44-year-olds). This means that third-party age-tier data leads to an average improvement in audience identification between 7% and 77% in comparison to 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 to 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, while 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 their services.[7] While user-reported, first-party data is 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 to 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 7.3).

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 their own cost-benefit ratio and should see our estimates as approximate guidelines.

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

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

Fifth, to the best of our knowledge, all studies included data retrieved through mobile web-based browsing and desktop browsing. However, the provided data did not allow us to specifically distinguish between mobile and desktop PC effects. Investigating any potential differences due to 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 1 and 2. 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 DSPs 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). However, if firms do not have data on consumers 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 (2017) 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 due to 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 DSPs 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 (ANA), have proposed a data labeling initiative for 2019, similar to nutrition labels for food (IAB, 2018). The data lables' goal is to increase transparency and help marketers

[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 (2017); Mallazzo (2018).

[9]First-party data is often used in advertising methods such as retargeting, where, 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).

understand what information digital audiences are based on. Our research underscores the need for such actions and initiaves. 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.

7 Online Appendix

7.1 Online Appendix: Institutional Background

Two digital ad technologies are at the centre of our study: Demand side platforms (DSPs) and data management platforms (DMPs). A DSP, such as Google's Bid Manager (recently rebranded to Google Display and Video 360), MediaMath or DataXu, allows an advertiser to purchase online ad placements in an automated fashion. To do this, the DSP is integrated with many different providers that sell possible ad slots, such as ad exchanges, ad networks or publishers. The Interactive Bureau of Advertising (IAB, 2012b) defines a Demand Side Platform as "a technology platform that provides centralized and aggregated media buying from multiple sources including ad exchanges, ad networks and sell side platforms, often leveraging real time bidding capabilities of these sources."

One of the main premises of digital advertising is that marketers can pick the consumers who are supposed to see their ads based on the buying parameters they provide to the DSP. This specific process is often referred to as 'programmatic advertising.' The buying parameters that are provided to a DSP include the desired pre-specified audience as well as other delivery parameters such as price. The DSP will scan the accessible inventory across the websites that are part of its platform network and execute ad purchases based on the provided parameters.

For advertising purposes, it is important to distinguish between different sources of data. 'First-party data profiles' refer to any consumer information that is collected and synthesized by advertisers themselves. However, many times advertisers lack data on prospective customers and turn to third-party data profiles. 'Third-party data profiles' are based on aggregated information from different data pieces and sources unknown to the advertiser. These third-party data profiles are often supplied by data brokers and are sold on a fixed-price basis for each bundle of ad impressions, which are typically sold in thousands (CPM = cost per mille).

Since it would be too cumbersome to request information from each data broker individually, marketers need a connection database that consolidates the information on users from many different sources and data brokers. This connection function is provided by DMPs, such as Lotame, Oracle BlueKai or Adobe Audience Manager (see Figure 2). A DMP is defined by the IAB as "a system that allows the collection of audience intelligence by advertisers and ad agencies, thereby allowing better ad targeting in subsequent campaigns."(IAB, 2012a)

While both platform technologies (DMP and DSP) can be integrated into the same tool or software, it is important to understand the different functions of these two technologies and why we have examined both the contribution of the DSP (data broker optimization) and the data brokers (through a DMP) individually.

Because a DMP or DSP can only access online populations across websites it is integrated into and because no platform is integrated with every possible website on the internet, the reachable cookies with specific user attributes are a subset of the potential online population that goes online during a campaign (see overlap of yellow and pink circle in Figure 1).

Furthermore, to buy ad placements based on a specific third-party data profile, a DMP also needs to be integrated with a data broker that can access websites where the desired online populations browse. Therefore, only a subset of all existing cookies meet the criteria of a programmatic ad campaign, that is, cookies that are active online at a certain time (pink circle), have been classified by a data broker (blue circle) and can be reached by the DSP/ DMP (yellow circle). If the DMP and DSP would be seperate software platforms,

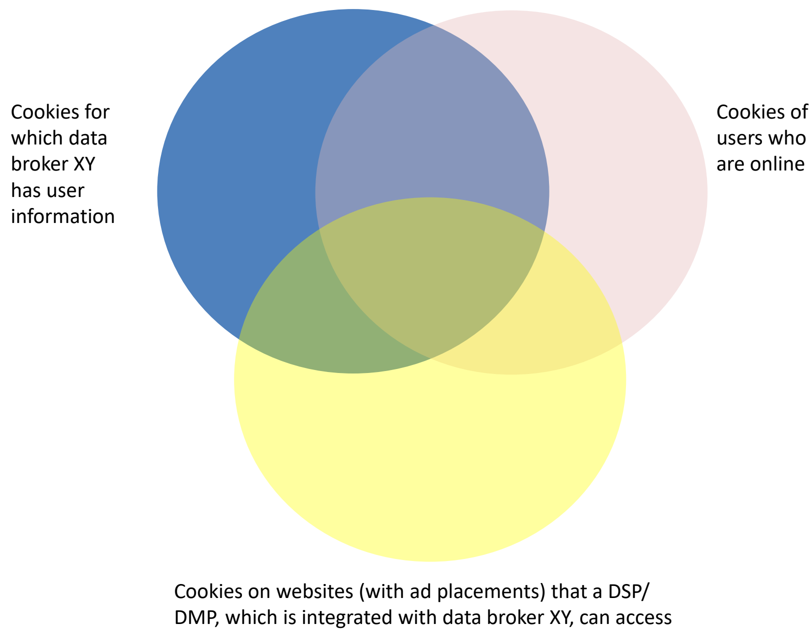

Figure 1: Subsets of Digital Cookie Populations

there would be even a further subset. To maximize reach and allow greater flexibility in ad buying choices, DMPs are typically integrated with many different DSPs and data brokers (see Figure 2). The DMP continually syncs and updates the available attribute lists from all integrated data brokers, typically every 24-48 hours.

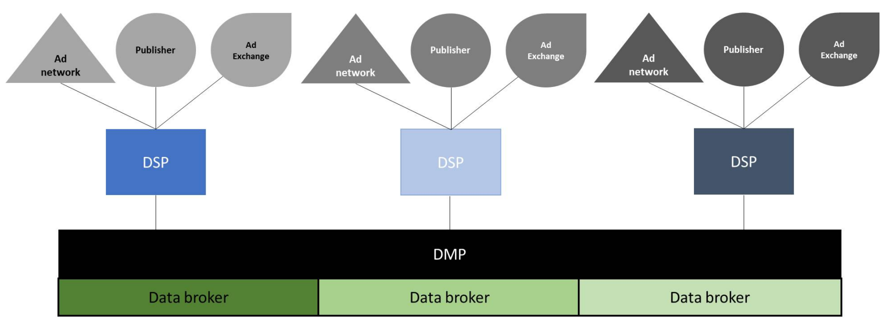

Figure 2: Example of DMP Integration with Multiple DSPs and Data Brokers

7.2 Online Appendix: Details on Method of Study One

Study One has the purpose to mimic a typical targeted online campaign in which advertisers use DSPs to buy ad placements across the web. Because we use Nielsen Digital Ad Ratings (DAR) as our validation source for the live ad campaign in this field study, we must consider another necessary data source that limits the subset of cookies that provide the information of interest for our campaign analysis (see Figure 3). In other words, Nielsen DAR will only have user information on a sub-sample of all data broker characteristics. It should be noted that Nielsen DAR audience data can be purchased as a branded segment for some attributes, such as age and gender, in many DMPs. However, in our case, we use Nielsen DAR services as a validation source for other (to us unobservable) data brokers. For privacy reasons and proprietary IP concerns, we could not obtain any detailed information about the sampling characteristics that underlie our validation test. This is a possible shortcoming that is addressed in Study 2 and 3, where we control for DMP and data-broker access (the blue circle).

7.3 Online Appendix: Details on Method of Study Two and Three

Study Two and Three addressed some of the possible limitations in Study 1 by focussing on the backend of a leading data management platform (DMP). Specifically, the unique setup in these two studies allowed us to retrieve audience samples of data brokers that constantly sync their attribute information with the DMP, using their entire network of partner websites across four countries (US, UK, Australia and New Zealand). We sampled 19 data brokers, whose identities we cannot reveal but are among the leading third-party data suppliers worldwide, providing audience intelligence services to marketers. It should also be noted that the DMP in our tests has a partnership with a bot-detection and quality measurement technology provider to remove questionable website inventory and cookies.

With respect to the sampling criteria, we collected all data from vendors who offered

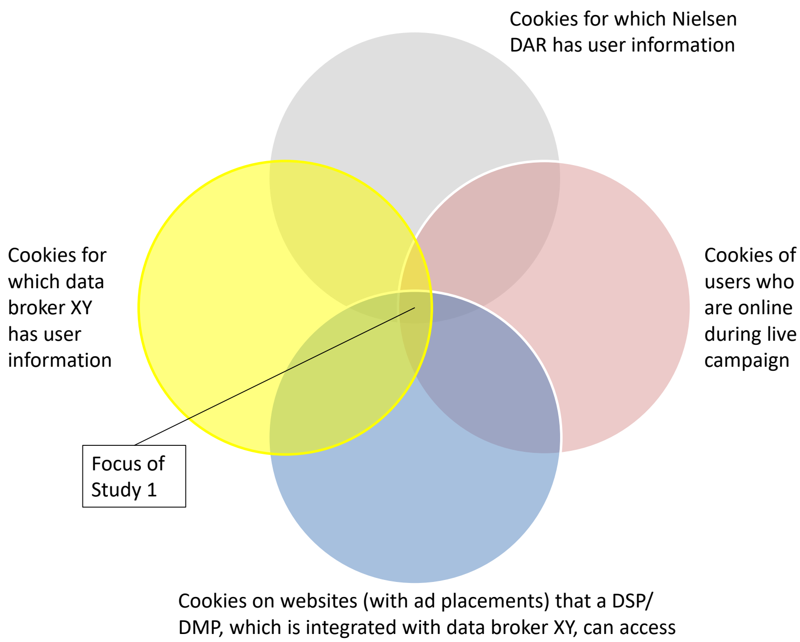

Figure 3: Subsets of Digital Cookie Populations in Study 1

comparable classifications at the time of our test (fourth quarter of 2016). The classifications can vary strongly across vendors; for example, age tiers differ. To compare 'apples to apples' for our selected audiences, we scanned all available data broker audiences for our desired attributes and selected only those where the definitions are exactly the same (e.g. age tier must be 25-34 and not 28-34).

To assess the [purchasable] data-broker attributes in the platform, we synced the platform database for four weeks with the cookies of the panel provider, which serves as our source of attribute validation. The panel provider, Pureprofile, asks each new member to fill out basic demographic information, including age and gender, when signing up for its services to earn money. Providing this information is voluntary, that is, users can opt to share this or not. In addition, many members have filled out surveys that included questions about their interests.

For all tests, the cookie syncing between panel provider and data platform creates the pool of cookies for which we have the desired attributes available. We only analyzed user

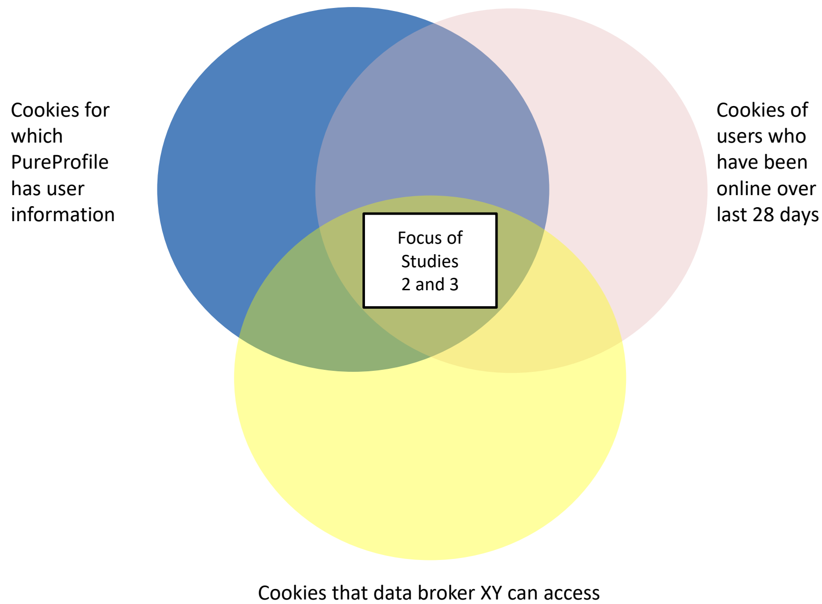

Figure 4: Subsets of Digital Cookie Populations in Studies 2 and 3

data for which we had the required information (see Figure 4). In detail, this means that we first filtered out all cookies that one data broker marked with a certain attribute (e.g. 'sports interested'). We then only used the subset of marked cookies for which we had validated data, that is, a person has indicated their real characteristics. This is an important distinction to Study One. Because data brokers only mark the presence of attributes for web cookies, this means there is no information on 'people who are not interested in sports.' Consequently, there are some limitations in the ways one can validate online profiles and audiences. For example, testing agreement across data broker classifications (how many data brokers marked the same cookie for a distinct attribute) is technically not feasible as one cannot distinguish missing data from wrongly marked web cookies.

We also use survey data to identify the natural distribution of our behavioral characteristics in the population in Study Three. First, according to the Australian Institute of Health and Welfare (AIHW, 2018), 48% of the population were sufficiently fitness active to receive health benefits.[10] Second, 56% of Australians own a travel passport (Australian Government, 2017), thus showing a high interest in being able to travel abroad too.[11] Third, a recent online survey of 1,500 people by the Australian Broadband Network (NBN, 2017) revealed that 67% of Australians are interested in one or more sports.

We choose to rather present conservative baselines (i.e., low natural distributions). For example, the percentage of people who travel may even be larger than the benchmark of 56% (passport owners) if one wanted to include day trips by car or train in the travel definition.

7.4 Online Appendix: Details on Cost Data

For our cost-benefit calculations, we need detailed cost data. Using a variety of sources, we obtain an average CPM of \$4.20 for standard display banner ads and \$18.92 for online video ads (see Table 12). For our calculations for the cost-benefit analysis, we used the averages of all indicated data on placement costs. We consider only data referring to the US and Australia for our average estimate as the CPMs for these countries are higher than global CPMs due to lower prices in developing countries.

Table 12: Cost data - CPM averages for US/Australia

CPM	Beales	AdNews	ANA	Emarketer	Pathmatics	Average
Video			12.64	25.3	13.72	18.92
Display	4.12	5.00	4.80	3.67	3.40	4.20

Sources: eMarketer (2017); Gottlieb (2016); eMarketer (2013); Bennett (2018); ANA (2017). From eMarketer (2017) we used the CPM value for data targeting and from Beales (2010) we used the CPM value for behavioral targeting.

However, these figures mask that advertisers contemplating using any third-party audiences encounter many different cost components (Gertz, 2018), whereby it is crucial to

[10] The health benefits statistic is similar, but slightly lower - hence more conservative - than the reported 55.4% of 18-64 years olds who undertook 150 minutes or more of exercise in the last week according to the most recent National Health Survey of 21,300 people (Australian Bureau of Statistics, 2018).

[11] The passport statistic is very close to the reported 51.3% of Australians who have used air transportation both domestically and internationally to travel between October 2015 and September 2016 according to a survey of 14,416 people by Roy Morgan (2016). We however decided to use the passport statistic as it is a nationally tracked record and does not depend on a fixed time period. For instance, some people may only travel every second year, not every 12 months.

distinguish between fixed and variable costs (percentages of spend):

1. Technology usage or license fees for the software platforms, that is, DSP and SSP (the equivalent of DSPs but used to optimize revenues/ prices for publishers and media owners). This is typically a percentage of the dollar amount the sell side of the advertising delivery process would charge for the service.

2. A fee for the execution services to operate the software platforms. Specialist units for programmatic media buying are referred to as trading desks. They typically also charge a percentage of the amount the sell side would charge.

3. Costs to serve ads and verification technology, which ensure that a possible cookie at a certain website is meeting the desired criteria: For example, that real humans see the ad and that it is placed on a brand safe environment. This is typically a fixed amount per thousand ad impressions (CPM).

4. Usage fees to use some third-party audience data for targeting. This is typically a fixed amount per CPM.

5. A media buying and planning fee for the agency that manages the account. This is typically a percentage of the total ad spend.

To calibrate these, we synthesized 10 publicly available cost data sources (see Tables 13). Based on these data sources, we obtain an average fee related as a percentage of total costs of the campaign of 7% for the media agency of record, 15% for the trading desks (execution of programmatic campaigns), 14.72% for DSP fees, and 9.92% for SSP/exchange fees. There are also fixed $1.33 third-party audience fees and $0.20 ad-serving and verification fees.

Next we calculated the specific cost components using a bottom-up approach such that we achieve a similar display banner ad CPM of $4.20 and an online video ad CPM of $18.90, the average values we calculated earlier.

For example, for a standard display ad, a publisher may expect \$1.36 to fill the respective ad slot. Then the SSP would add a charge of 9.92% of the \$1.36, which is 13 cents. In addition, there would be fixed costs of \$1.33 per CPM for the use of third-party audience information and 20 cents per CPM for ad serving and verification services. These cost components would make up the sell-side costs, totalling \$3.02 for our example. Then, we apply charges of 14.72% (44 cents) and 15% (45 cents) of the sell-side costs for the DSP and trading desk, respectively. These costs would make up the programmatic ad costs, equal to \$3.92. Finally, the media agency would charge 7% (27 cents) on top of the digital ad costs, resulting in a final CPM of \$4.20 for the advertiser.

We repeated the same exercise using the cost data and bottom-up approach for an online video ad example that results in a final CPM of \$18.90 when using audience targeting data.

Importantly, we can compare the same scenario for an ad buy that would not use any audience targeting and be executed only through a traditional media buy through the agency of record. In this case, the only extra costs on top of the ad slot pricing would be the ad serving/verification and the agency media buying fee, resulting in a final CPM of \$1.67 for our display banner example. For our video ad example, the CPM would be reduced to \$11.98.

Table 13: Cost components for digital audience solutions

Cost component	Cost type	WARC	O'Hara	TubeMogul	TradeDesk	Rubicon	Appnexus	Gertz	ANA	MarketingLand	AdExchanger	Average
Agency of record	Variable (%)	5	10						6			7.00
Trading desk/ execution	Variable (%)	15	15					15				15.00
DSP	Variable (%)	10	15	21.6	19.7			10	12			14.72
3rd party data	Fixed ($)	0.97**	1					1.2	0.43	2.38	2	1.33
SSP/exchange	Variable (%)	5	10			11	8.5	10	15			9.92
Verification & ad serving costs	Fixed ($)							0.20*				0.20

Sources: Sluis (2018); WARC (2018); Gertz (2018); O'Hara (2015); Gottlieb (2016); ANA (2017); eMarketer (2017, 2013); Vidakovic (2013).

* The number is based on the example in text, indicating 2% out of $100 ad spend. We changed this to 20 cents which corresponds to the fee range for individual CPMs (note that different values would barely change the cost-benefit ratio as they are added to both targeted and non-targeted ads). ** The WARC study indicated 25% costs for both third-party data and verification. We subtracted the 2% ad serving/verification fees and applied the 23% to our calculated average CPM of $4.20, resulting in 97 cents.

Table 14: Cost data for digital audience solutions

	Display ad costs		Video ad costs	
	Targeting	No targeting	Targeting	No targeting
Publisher	1.36	1.36	11.00	11.00
SSP exchange	0.13*		1.09*	
3rd party data costs	1.33		1.33	
Ad serving & verification	0.20	0.20	0.20	0.20
Sum sell side costs	3.02	1.56	13.62	11.20
DSP	0.44*		2.00*	
Trading desk/execution	0.45*		2.04*	
Sum digital ad costs	3.92	1.56	17.67	11.20
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	

* refers to variable costs. See Table 13 for corresponding percentages.

Sources: Sluis (2018); Gertz (2018)

7.5 Online Appendix: Fashion Interest Audience Data Validation

We were also able to validate two observations for audiences aiming to capture fashion-interested online users. Again, we used Pureprofile panel data, where users must have directly indicated that they wish to see content about fashion, to validate these two segments. While this is a limited sample only - and therefore the average accuracy is only suggestive - we obtain an accuracy (hit rate) average of 31.4% for this attribute.

It is challenging to find a population statistic that may reveal the natural distribution for this rather specific, abstract attribute (being interested in fashion). According to a recent GfK ecommerce report (Goldring, 2017), 41% of Australians bought clothing or apparel in the last six months online. Moreover, readership research by Roy Morgan (2018) suggests that about 4.6% of the Australian population read fashion magazine content (both online and offline). If we take the middle value of these two statistics, we obtain a baseline of 22.8%, resulting in a relative improvement of 37.7% (31.8/22.8=1.4) for using third-party audience targeting versus no targeting. This estimated value actually lies in the middle of our improvement range we find for the other attributes (0-77%) across Study 2 and 3.

Table 15: Data Broker Accuracy for Fashion-Interested Audiences

Data broker	sample	accuracy %
Vendor A	272	30.1
Vendor J	5599	33.5
Average	2935.5	31.8

References

AIHW (2018). Physical activity Overview - Australian Institute of Health and Welfare.

ANA (2017). Programmatic: Seeing through the financial fog.

Australian Bureau of Statistics (2018). National Health Survey 2017-2018.

Australian Government (2017). Passport facts 2016-2017.

Bennett, L. (2018). Australia has the most expensive Facebook ads in the world, study finds. *AdNews*.

Bergemann, D. and A. Bonatti (2015, aug). Selling Cookies. *American Economic Journal: Microeconomics 7*(3), 259–294.

Bucklin, R. E. and C. Sismeiro (2003, aug). A Model of Web Site Browsing Behavior Estimated on Clickstream Data. *Journal of Marketing Research 40*(3), 249–267.

Chen, Y., C. Narasimhan, and Z. J. Zhang (2001). Individual marketing with imperfect targetability. *Marketing Science 20*(1), 23–41.

Coey, D. and M. Bailey (2016). People and Cookies: Imperfect Treatment Assignment in Online Experiments. In *Proceedings of the 25th International Conference on World Wide Web - WWW '16*, New York, New York, USA, pp. 1103–1111. ACM Press.

De Bruyn, A. and T. Otter (2017, July). Bayesian Customer Profiling: Applications to Age and Political Partisanship Estimation. *Working Paper*.

eMarketer (2013). Online Video Advertising Moves Front and Center.

eMarketer (2017). US Display Ad Benchmarks: CPM and CTR, by Targeting Method, 2016 (among impressions served by Choozle) - eMarketer.

Flosi, S., G. Fulgoni, and A. Vollman (2013). If an Advertisement runs online and no one sees it, Is it still an Ad? Empirical generalizations in digital advertising. *Journal of Advertising Research 53*(2), 192–199.

Gertz, O. (2018). Debunking the tech tax and other programmatic myths. *The Drum*.

Goldring, N. (2017). Australian online shopping behaviour. *GFK research report*.

Gottlieb, G. (2016). Programmatic : Quantity and Quality. In *MediaPost's OMMA LA Conference.*

Hof, R. (2014). OpenX Aims To Boost Publishers' Online Ads With New SSP Technology. *Forbes*.

IAB (2012a). Data Management Platform - IAB Wiki.

IAB (2012b). Demand side platform - IAB Wiki.

IAB (2018). Major Advertising Trade Bodies Unveil Data Transparency Label.

IAB and WinterberryGroup (2018). The State of Data.

Johnson, G. A., R. A. Lewis, and E. I. Nubbemeyer (2017, dec). Ghost Ads: Improving the Economics of Measuring Online Ad Effectiveness. *Journal of Marketing Research 54*(6), 867–884.

Kelly, C. (2017). Inaccurate Segments May Be Costing Advertisers Billions. *AdExchanger*.

Kim, Y., W. N. Street, G. J. Russell, and F. Menczer (2005, feb). Customer Targeting: A Neural Network Approach Guided by Genetic Algorithms. *Management Science 51*(2), 264–276.

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

Lerner, A., A. K. Simpson, T. Kohno, and F. Roesner (2016). Internet Jones and the Raiders of the Lost Trackers: An Archaeological Study of Web Tracking from 1996 to 2016. In *USENIX Security Symposium*.

Lin, T. and S. Misra (2018). Identity fragmentation bias. *Mimeo, Chicago*.

Lotame (2018). The New State of Audience Data: Accuracy Matters. https://www.lotame.com/lotame-research-report-the-new-state-of-audience-data-accuracy-matters/.

Mallazzo, M. (2018). When Did Flawed Data Become OK? *AdExchanger*.

Murthi, B. and S. Sarkar (2003, October). The role of the management sciences in research on personalization. *Management Science 49*(10), 1344–132.

Narayanan, S. and P. Manchanda (2006, may). Heterogeneous Learning and the Targeting of Marketing Communication for New Products. *Ssrn 28*(3), 424–441.

NBN (2017). Future of Sport Report.

O'Hara, C. (2015). Programmatic 3.0: The Next Paradigm In Inventory Procurement. *AdExchanger*.

Pancras, J. and K. Sudhir (2007, nov). Optimal Marketing Strategies for a Customer Data Intermediary. *Journal of Marketing Research 44*(4), 560–578.

Park, Y.-H. and P. S. Fader (2004, Aug). Modeling Browsing Behavior at Multiple Websites. *Marketing Science 23*(3), 280–303.

Roy Morgan (2016). Sky high: Australians air travel habits.

Roy Morgan (2018). Over 15 million Australians read magazines across print and online.

Sahni, N. S., S. Narayanan, and K. Kalyanam (2017). An experimental investigation of the effects of retargeted advertising: The role of frequency and timing.

Salesforce (2018). Digital advertising 2020: Insights into a new era of advertising and media buying.

Sluis, S. (2018). Investigation: DSPs Charge Hidden Fees And Many Can't Afford To Stop. *AdExchanger*.

Statista (2015). Australia: Age distribution of internet users 2015.

The Economist (2017). Data is giving rise to a new economy - Fuel of the future.

Trusov, M., L. Ma, and Z. Jamal (2016). Crumbs of the Cookie: User Profiling in Customer-Base Analysis and Behavioral Targeting. *Marketing Science 35*(3), 405–426.

Vidakovic, R. (2013). 5 Ways To Evaluate The Quality Of Audience Data. *MarketingLand*.

WARC (2018). Tech Tax' cost programmatic advertisers over $30bn in 2017.