Ad Attributes and Attribution:

Large-Scale Field Experiments Measure Online Customer Acquisition

Randall A. Lewis, David H. Reiley, and Taylor A. Schreiner*

First Draft: 30 June 2011
This Version: 13 March 2012

Abstract

Using a large-scale experiment involving 3.7 million treated subjects on Yahoo!, we measure the ability of online display advertising to cause new account sign-ups at an online business. We experiment with two dimensions of media choice: banner-shaped ads versus large rectangular ads, and Yahoo! Mail placement versus Yahoo! Run-of-Network placement. The Run-of-Network advertisements succeeded in generating a statistically significant increase in sign-ups of approximately 11% relative to the control group, but the Yahoo! Mail ads did not produce a statistically significant increase in sign-ups for this advertiser. By contrast with the results for choice of pages, we find no effect of the form factor of the ads (banner versus large rectangular) in this campaign. We note the somewhat surprising fact that even with experimental sample sizes in the millions of subjects, our estimates remain quite noisy: for example, the upper bound of the 95% confidence interval for the statistically insignificant Yahoo! Mail campaign represents a 15% increase in new customers. Most importantly, our estimates call into question click-only attribution models, as restricting attention to the number of users that clicked on an ad and converted is less than 30% of the estimated treatment effect on conversions.

JEL Classification

C93 - Field Experiments, M37 – Advertising, L81 – E-Commerce

Keywords

Online advertising, advertising effectiveness, field experiment, click attribution

---

* Lewis and Reiley: Yahoo! Research, <ralewis@yahoo-inc.com> and <reiley@yahoo-inc.com>. Schreiner: Yahoo! Marketing Insights, <tschreiner@yahoo-inc.com>. We thank Meredith Gordon and Sergiy Matusevych for their work on the experiment and the data. Yahoo! Incorporated provided financial and data assistance as well as guaranteed academic independence prior to our analysis so that the results could be published no matter how they turned out. We acknowledge the helpful comments of Clara Lewis, Glenn Ellison, Jerry Hausman, Stephen Ryan, Sara Ellison, and many participants at the MIT Industrial Organization Workshop and the Economic Science Association meetings in Tucson.

I. Introduction

Advertising is a frontier in measurement. Tens of billions of dollars are spent each year on information and persuasion in a variety of media formats, including email, sponsored search, display advertising, radio, and video. In 2009, advertisers spent $22.7 billion online (IAB Internet Advertising Revenue Report 2009) with display ads accounting for 22% or $5.1 billion. In spite of the large scale of these advertising expenditures, little is known about the effectiveness of these dollars. Few studies have gone beyond "the click" to assess the effects of ads on concrete objectives such as sales or new accounts. We address this issue by running a large-scale, randomized controlled experiment tracking ad exposures and outcomes at the individual level.

The experiment randomly assigned each Yahoo! visitor into a treatment or control group. The treatment group was further divided into four subgroups to examine the impact of two different ad formats on the page and two different placements on Yahoo! websites. The two different ad formats were banner ("north") versus large rectangular (LREC) ad units. These ads were served either on Yahoo! Mail or as Yahoo! Run-of-Network, which is a portfolio of available inventory across many of Yahoo!'s various subdomains (including Yahoo! Mail). In total, 3.7 million treatment-group members were shown 67 million online display ads for an online business.[1] These ads represented approximately 5% of contemporaneous online advertising expenditure by the company. The ads were targeted at the top 10% of scorers in a standard click-targeting model. From this target group of users, 52% were randomly chosen as a control group and deliberately excluded from seeing the ads. The remaining qualified users were randomly split between the four treatment subgroups, totaling 3.7 million Yahoo! visitors exposed to the ad campaign for this advertiser. The objective was to quantify the causal effect of the ads on new account sign-ups and identify how to reach the most responsive users.

The Run-of-Network ads generated statistically significant increases in sign-ups relative to the control group, approximately 10% for both North and LREC ads. However, the ads shown on Yahoo! Mail did not produce a statistically significant increase in sign-ups. Despite being derived using millions of subjects, this estimate is

[1] This company, which prefers not to be identified, was advertising to generate new accounts for a service with the potential for repeated billing.

quite noisy, with the upper bound of the 95% confidence interval estimate being a 15% increase in new customers.

Most importantly, if we had followed a standard industry practice and assumed all conversions due to the ads must have involved a click on one of the ads, we would have missed at least 70% of the causal effect of the ads. In other words, at least 70% of the increased sign-ups came from users who saw but did not click on the ads. Accordingly, click-based attribution models, which fail to account for these "view-through" conversions, would significantly understate the effects of the ads and should be regarded with caution. This contributes significantly to the debate about "attribution" in the online-advertising industry, involving various attempts to determine each publisher's contribution when a user decides to make a purchase. A typical attribution model is the last-click model, in which the ad last clicked before a conversion gets "credit" for that conversion. Proposed replacements for last-click attribution involve different weights for various classes of online media events (views, clicks, etc.), such as Engagement Mapping, created by the Microsoft subsidiary, Atlas (2008). However, the media-event weighting assumptions upon which these models rely have not been scientifically validated and are not universally agreed upon by advertisers and publishers. Publishers seek credit for delivering the advertiser's message to their website visitors while advertisers want to ensure that their advertising works.

This paper contributes to the literature using randomized trials to assess the effectiveness of advertising. Zielske (1959) pioneered the use of field experiments to examine the effects of advertising, mailing print advertisements to potential customers and measuring the impact of the ads over time on consumer recall. Several advertising experiments run by Eastlack and Rao (1989) explored outdoor, print, radio, and television ads from 1975 to 1977 with the Campbell Soup Company. Other experiments have examined the influence of direct-mail advertising on consumer purchases, including Simester, Sun, and Tsitsiklis (2006) and Anderson and Simester (2008) for mail-order catalogs, and Bertrand *et al.* (2009) for short-term loans.

Abraham and Lodish (1990) and Lodish, et al. (1995a,b) pioneered experimental measurements of television advertising in the 1980s by exogenously varying cable television subscribers' exposure to ads for consumer packaged goods. Each experiment tracked 3,000 households per market in several markets, matching ad exposure and purchases at the individual level, using television set-top boxes and supermarket scanners. Their meta-analyses found aggregate evidence of the effectiveness of television advertising across a large number of advertising campaigns. A recent update on this branch of research by Hu, Lodish, and Krieger (2007) confirms the earlier meta-analyses. However, this research also shows that sample sizes of 3,000 were too small to be able to give precise estimates of the effects of each individual advertising campaign: their threshold for a successful campaign was a one-tailed hypothesis test at the 20% level of significance. Our research scales up their sample sizes by several orders of magnitude, providing increased precision.

Most previous research specific to *online* advertising has mostly been unable to tie advertising exposure to actual consumer transactions. Danaher and Mullarkey (2003) studied recall of online advertising in a laboratory experiment, while Chiou and Tucker (2010) studied the effects of search advertising for pharmaceuticals on consumer choices to search for health treatments.

Our own recent work has studied the impact on consumer purchases of online advertising placed by retailers. Lewis and Reiley (2011b) found that the majority of the online advertising effect for a major offline retailer came through their offline storefront. Further, most of this effect could be attributed to the large share of customers who viewed the ad but chose not to click. Lewis and Reiley (2011a) further analyzed the experiment with the nationwide retailer and found that 40% of the impact of the online advertising comes from the 5% of customers who were aged 65 and older. These results came from studying the existing customers of a retail advertiser; by contrast, the present experiment involves studying the acquisition of new customers.

This paper demonstrates that the technological advances of online advertising can facilitate accurate measurements of the marginal effectiveness of advertising. In particular, randomized controlled experimentation is a viable method for ad campaign evaluation and cost-effective targeting improvements. Using these tools, advertisers can identify subpopulations where the ads had the largest marginal effect and only deliver ads to those who are influenced enough to recoup the advertising expenses through increased profits. We also show that with sample sizes in the millions of experimental subjects, we are just on the frontier of being able to measure economically relevant effects of advertising. In our concluding remarks, we discuss improvements in experimental design and data gathering that will improve statistical precision in future experiments.

The remainder of the paper proceeds with an overview of the experiment in section II, a description of the data gathered and used in the experiment in section **Error! Reference source not found.**, the analysis and results of the experiment in section **Error! Reference source not found.**, and concluding remarks in section V.

II. Experiment Overview

We carried out an experiment to assess the effectiveness of online display advertising at attracting new accounts for an online business. The experiment consisted of showing online display advertisements on Yahoo! and then tracking the Yahoo! visitors who signed up for a new account at the advertiser's website. We study a company that does most of its business online and expects a majority of the advertising impact to come through the traceable online channel.

The ads were shown on Yahoo! either as large rectangular units (LREC, 300x250 pixels) or as banner ads (728x90 pixels). Examples of these ad dimensions and placement for AT&T (not the advertiser in our experiment) can be seen in Figure 1. In this Yahoo! Tech page, we see both a banner ad across the top of the page and a complementary LREC ad near the bottom of the page. The ad impressions for this experiment, by contrast, involved just one advertisement per page (either a banner or an LREC).

Figure 1 - Yahoo! Run-of-Network (Yahoo! Tech) Banner and LREC

The experiment used a browser cookie[2] to randomly assign each visitor to Yahoo! to either the treatment group (who could see the ads) or to the control group (designated as ineligible to see this advertiser's campaign, they instead saw ads from other advertisers). In addition, the ad server was programmed to show ads only to a targeted subpopulation: the top 10% scoring visitors[3] in a standard Yahoo! click-targeting model. Then, during the experiment, treatment group visitors were collectively shown 67.4 million ad impressions for this advertiser during their visits to Yahoo! Fifty-two percent of the targeted population was assigned to the control, leaving the remaining 48% to be

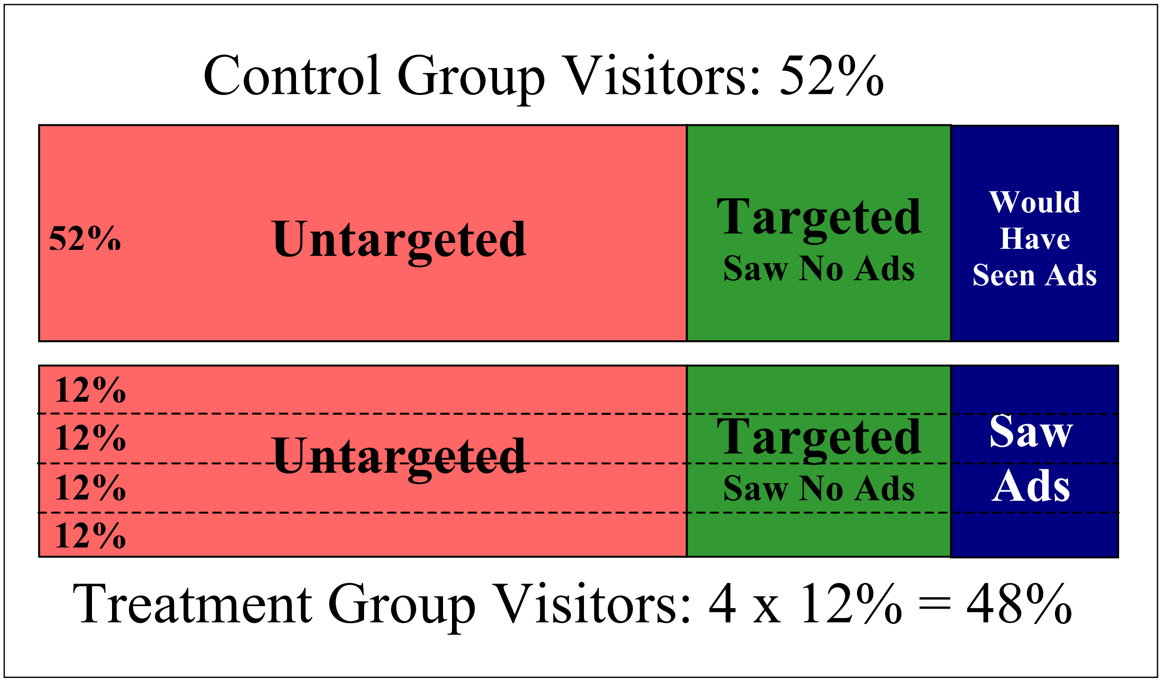
[2] A browser cookie is a piece of data stored in the user's web browser that helps online services customize a user's browsing experience. Each time a user visits a Yahoo webpage, Yahoo checks to see whether a Yahoo browser cookie identification number is stored in the user's web browser. If the user does not already have a number stored in a browser cookie, Yahoo issues a new identification number and stores it in that user's web browser. Approximately 146 million unique users visit Yahoo each month (comScore 2008), and each has a unique, randomly assigned, number. We use this number to select randomized treatment and control groups. Note that cookie "churn" (deletion of cookies by users) is approximately 30% per month for recently generated cookies and 10% for long-lived cookies. This churn will tend to attenuate our estimated effects, because it will reshuffle some users across treatment and control groups during the experiment if they delete their cookies and receive a new Yahoo! cookie the next time they visit.

[3] The top 10% of scores for Yahoo! visitors is restricted to "scoreable" users—visitors for whom we had enough data to compute the "custom" click-targeting model, a targeting service offered by Yahoo!. Hence, these ads were eligible to be shown to less than 10% of all Yahoo! visitors.

treated with ads. This group was broken into four equally-sized treatments (12% each): Yahoo! Mail LREC ads, Yahoo! Mail banner ads, Yahoo! Run-of-Network LREC ads, and Yahoo! Run-of-Network banner ads. While Mail advertising takes place exclusively on the mail.yahoo.com subdomain, Run-of-Network ads are delivered to a variety of subdomains (mail.yahoo.com, news.yahoo.com, tech.yahoo.com, autos.yahoo.com, etc.).

Figure 2 illustrates this experimental design, with an initial random assignment of the entire Yahoo! population, a subset of that population qualified to see those ads, and a smaller subset who happened to browse the appropriate pages at the appropriate time to receive delivery of the ads. Differences between the control and treatment groups represent the impact of the display ads on users' behavior—differences between control-group users who "would have seen ads" if they had been in the treatment group, versus treatment-group users who actually "saw ads." Since we do not have eye-tracking capability, we do not know who literally *saw* the ads but use the word "saw" as shorthand for the event that we delivered an ad to a given user; note that these are the relevant users to the advertiser, since the advertisers pays a fee for each delivered impression. Note also that technological limitations prevented us from recording which individual control-group users would have seen the ads for this campaign, so instead of being able to restrict attention to the treatment effect on the treated (the segment of users on the right of the diagram), we must content ourselves with studying the entire Yahoo! user population (the entire set of users in the figure).

Figure 2 - Treatment and Control: Randomizing the Browser Cookies of the Yahoo! Population



The scale of this campaign is typical for display ad campaigns shown on Yahoo!. However, we note that these ads only represented approximately 5% of the advertiser's expenditure on online advertising during this period. As such, we measure the marginal impact of each of these four advertising strategies, given the users' contemporaneous exposure to other ads purchased with the other 95% of the advertiser's budget (which, through randomization, should be equivalent across treatment and control).

III. Data Description

Basic statistics from the experiment are presented in

Table 1. A total of 3.7 million visitors, divided into four equal-sized treatment groups, received 67.4 million ads over six consecutive weeks during January and February 2008.

Table 1 – Treated Individuals Data Summary

Location	Yahoo! Mail	Yahoo! Mail	Yahoo! Run-of-Network	Yahoo! Run-of-Network
Media	LREC	Banner	LREC	Banner
Number of Impressions	6,613,761	21,506,041	9,169,441	30,099,923
Number of Viewers	794,332	748,730	1,080,250	1,101,638
Avg. Impressions	8.3	28.7	8.5	27.3
Number of Clicks	1,965	6,171	3,992	10,560
Number of Sign-ups	867	762	1,254	1,304
Num. of Clicker Sign-ups	13	16	17	36
Click-Through Rate (CTR) = # Clicks / # Impressions	0.030%	0.029%	0.044%	0.035%
Clickers/Viewers	0.222%	0.715%	0.329%	0.810%
Sign-up Rate	0.109%	0.102%	0.116%	0.118%

Average impressions and click rates differed across media and location. However, the differences in website characteristics and advertising format would lead each of the four combinations of media and location to expose somewhat different populations. For example, the Mail ads would not reach users who did not check their Yahoo! Mail accounts during the relevant period, and the LREC ads would have reached relatively more users who tended to browse the types of Yahoo! pages containing LREC rather than banner ads.[4] Therefore, the four samples of exposed users are not equivalent populations; because they have different browsing behavior, we should also expect them to have different baseline propensities to sign up for new accounts with the advertiser.[5] We therefore make comparisons between the superset populations of Yahoo! browser cookies who could have been exposed to the campaigns, rather than between the subpopulations who were actually exposed. In particular, we do not know how much of the difference

[4] Unlike the illustration in Figure 1, many Yahoo! pages contain only one or the other of these two ad types. Different layout decisions are made by different property managers.
[5] See Lewis and Reiley 2011b for concrete measurements of spurious correlation between advertising exposure and purchase behavior.

between the Mail and Run-of-Network campaigns to attribute to the differences in ad format, differences in webpage, or differences in frequency of impressions.

Users who received at least one ad in the first treatment group (Mail/LREC), received an average of 8.3 ad impressions from this campaign. We find that 0.222% of those users clicked at least once ("clicker rate"), while the sign-up rate for this group was 0.109%. The corresponding sample for the second treatment group (Mail/Banner) saw an average of 28.7 ads per person, with 0.715% clicking at least one ad and 0.102% signing up for a new account with the advertiser.

Users exposed in the third treatment group (Run-of-Network/LREC) saw an average of 8.5 impressions per person, with a clicker rate of 0.329% and a sign-up rate of 0.116%. Finally, users who saw ads in the fourth treatment group (Run-of-Network/Banner) saw an average of 27.3 impressions per person, with a clicker rate of 0.810% and a sign-up rate of 0.118%.

These descriptive statistics suggest some insights about the efficacy and the characteristics of the ads. We observe higher frequency per user with the Banner ads relative to the LRECs, as well as a higher propensity to click. Greater frequency might encourage a higher tendency to notice and engage with the ad. However, diminishing returns could easily set in by the 20th impression, by which point many users might have already noticed the ad and experienced the vast majority of its potential impact on them. This is reflected in the fact that the higher clicker rates are not matched by correspondingly higher sign-up rates for the Banner treatment groups. Of course, these data are suggestive rather than causal, because as noted above, the populations receiving the ads in the different treatments could easily differ in their baseline clicking or sign-up behavior. We refer the user interested in the effects of frequency to Lewis (2010), who exploits a clean natural experiment to understand the causal effects of additional impressions, measuring the extent of diminishing returns across a variety of online ad campaigns, finding the "ad fatigue" or "wear out" to be relatively small for a large number of online advertising campaigns.

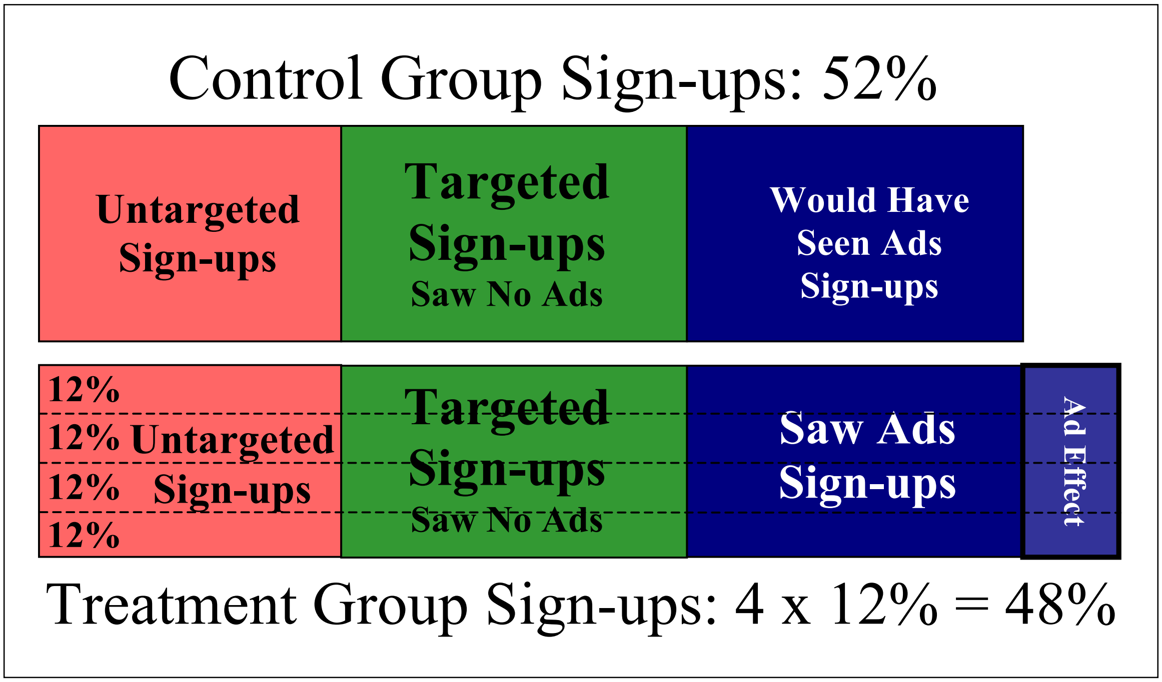
One striking observation is the remarkably small count of users who both clicked on ads and signed up for accounts. For each of the ad campaigns, thousands of individuals clicked on the ads, while fewer than 40 of those clickers signed up for a new account with the advertiser.[6] Since clickers account for less than 1% of exposed visitors, we are also quite interested in the number of non-clicking ad viewers who were influenced to sign up for a new account. We measure the impact on both clickers *and* viewers by using our randomized experiment, which provides a proper control group to compare with the treatment groups.

In order to identify the proper control, we use the pre-experiment random assignment of all Yahoo! visitors. Recall, as depicted in

[6] Note that we do not know how many clickers who signed up would have signed up anyway in the absence of the ads. Since clicks are endogenous (i.e. we cannot force a click in this experiment), we are left with serious selection problems in any causal inference we might like to make about clickers. In particular, it is impossible to construct a control group to measure how many clickers would have signed up had they not been given the opportunity to see or click on the ad.

Figure 2, that while every visitor was assigned to the treatment or control group, only the top 10% of scorers in the targeting model were eligible to be shown the ads. Recall that we were unable to mark the subset of control users who would have seen the ads had they been in the treatment group. We therefore track the entire population of Yahoo! visitors who signed up for an account on the business's website, comparing the entire treatment and control groups regardless of whether those users saw or were even qualified for the targeting of these ads. Using the sign-up counts for all Yahoo! browser cookies during the six-week campaign, we arrive at a valid experimental comparison. The comparison contains an unfortunate amount of statistical noise from the users who never could have seen the ads on Yahoo!, as our treatment effect will be the "needle" displayed on the right hand side of Figure 3, representing the "haystack" of all sign-ups. Fortunately, our sample sizes are large enough to give us some statistical power despite the technological limitations on recording data on counterfactual ad views in the control group.

Figure 3 - Treatment and Control: All New Account Sign-ups on Online Business's Website

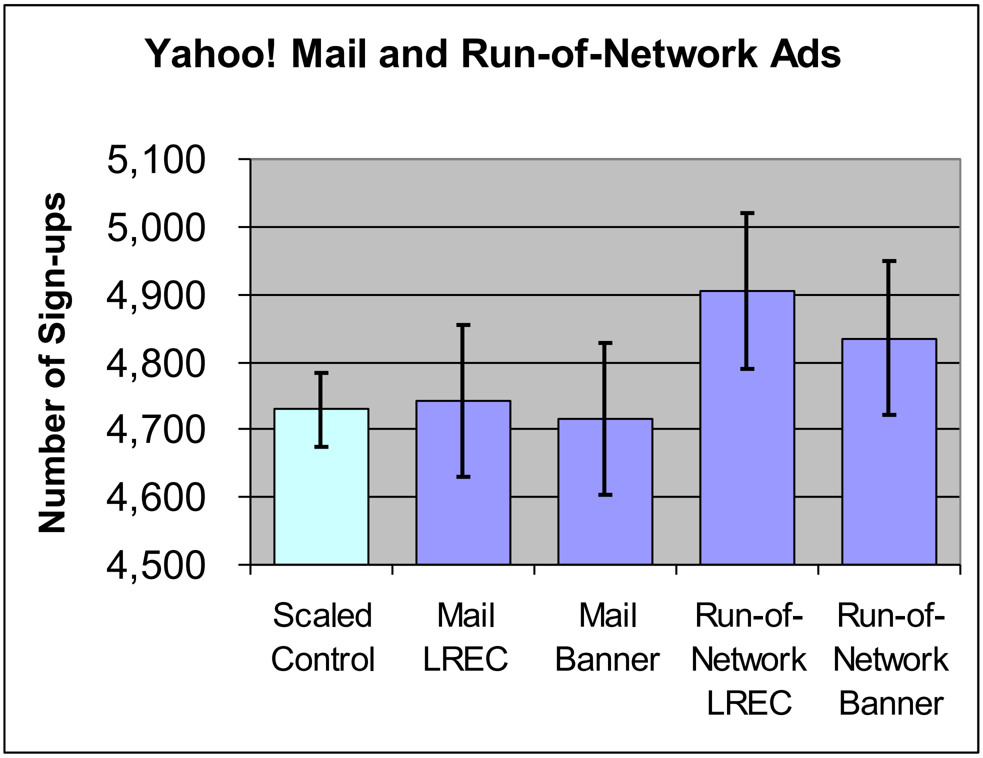


IV. Experimental Results

Given the experiment, our statistical analysis is straightforward. Recall that we tracked the number of new account sign-ups over six weeks for the treatment and control groups. Since the fraction of the population that signed up for these accounts is small (a mere 0.112% of those treated), we choose to approximate our count data as a Poisson random variable. That is, we assume a constant probability per unit time of observing a new account sign-up with the advertiser, and we allow this probability to vary by experimental treatment group. This simplifies the computation of standard errors; in order to compute standard errors for the total count of new-account conversions, we simply take the square root of that count.[7] Further, since our counts of roughly 20,000

[7] An alternative specification would be to assume a binomial distribution for the number of counts, based on the number of exposed users as the potential number of conversions *N*. Of course, for small *p* and large *N*, as in this situation, the binomial and Poisson distributions are approximately equivalent.

Figure 4 - All Four Treatments versus Scaled Control



for the treatment and control groups are quite large, the Poisson approximation is approximately Gaussian, allowing for standard normal-distribution-based inference.[8,9]

The experimental outcomes are presented for reference in Figure 4 with 90% confidence intervals, as well as in Table 2. In total, over the six-week advertising campaign, 20,497 control group individuals (within 52% of the population) and 19,196 treatment group individuals (within 48% of the population) signed up for an account on the business's website.

First, the aggregate effect of the advertising is the simple difference between the sum of the sign-ups for the four treatments and a scaled (48/52) count of the control group's sign-ups (that is, counts per 48% of the total population studied). This effect is

[8] The Poisson model implies that the sampling variance equals the mean for the number of counts. When we compute the variance via another method, we find a lower number, probably due to sampling variation. We compute the variance of the number of counts across fifty-two independent 1% partitions of the control group, obtaining 242.2. By contrast, the sample mean across the fifty-two groups is 22% larger, at we 394.2 (signups per 1% partition). We point this out to demonstrate that in computing standard errors via the Poisson method, we have been conservative and chosen the method giving larger standard errors of our estimates of advertising effectiveness.

[9] This means that we can also use the normal distribution to perform interval estimation for our treatment (difference in Poisson counts between treatment and control). In particular, we need not employ the Skellam distribution's critical values, because the difference between two normal random variables is itself normally distributed.

estimated to be 275.7 (191.5), which is statistically significant at the 10% level (p= 0.075, one-sided). The 90% confidence interval gives us a lower bound of 30 sign-ups due to the advertising. With a total of only 82 conversions coming from visitors that clicked on the ads, we see that non-clicking viewers must have accounted for at least 70% of the causal effects of the ads.

Second, we compare Yahoo! Mail versus Run-of-Network ad placements. We find that Yahoo! Mail ads had a very small estimated impact of -3.2 (117.6) sign-ups, while the Run-of-Network ads had a strong and statistically significant (p=0.01, one-sided) effect of +278.8 (118.8) sign-ups. This strong effect indicates that the Run-of-Network ads performed much better than the Mail ads for this particular advertiser. A 90% confidence interval for the Yahoo! Run-of-Network ads puts the lower bound at 126.6 sign-ups. Because only 53 conversions came from clickers, we see once again that click-based attribution would miss the majority of the ad effect—the large effect from only viewing the display ads.

Third, we examine differences in the form factor of the ads (and their placement on the page). Both Yahoo! Mail ads performed similarly to each other, posting statistically insignificant effects of 11.9 and -15.1 for LREC and banner ads, respectively. While these point estimates are small, we note that their precision is low, with estimated standard errors of 76 sign-ups. While the 90% confidence intervals rule out extreme outcomes, they still include treatment effects as high as 100 sign-ups per 12% of the population. Both Yahoo! Run-of-Network ads performed well, with the LREC ads performing better than the banners, with estimates of 173.9 (77.4) and 104.9 (77.0), respectively. However, these estimated counts are not significantly different from one another. Once again, click-based conversion would grossly understate the effect of the display ads – using the lower bounds of the 90% confidence interval estimates, we find that the minimum amount of underestimate for the Run-of-Network campaigns would be 90% for LRECs and 66% for banners.

Table 2 - All Sign-Up Results

	Control	Treatment			
Location		Yahoo! Mail		Yahoo! Run-of-Network	
Media		LREC	Banner	LREC	Banner
Fraction of Population	52%	12%	12%	12%	12%
Number of New Sign-ups	20,497	4,742	4,715	4,904	4,835
Expected Sign-ups (from Control)	-	4,730	4,730	4,730	4,730
Num. of Clicker Sign-ups	-	13	16	**17**	**36**
Total Ad Effect (Ads)	-	**275.7*** **(191.5)**			
Ad Effect (Location)	-	-3.2 (117.6)		**278.8***** **(118.8)**	
Ad Effect (Location x Media)	-	11.9 (76.4)	-15.1 (76.2)	**173.9**** **(77.4)**	**104.9*** **(77.0)**

Standard errors denoted by parentheses.
*, **, and *** Denotes significance at the one-sided 10%, 5%, and 1% levels, respectively.

In summary, we find that Yahoo! Mail ads failed to produce a large enough effect to be statistically distinguishable from zero. However, the wide confidence intervals, with standard errors of 117.6 sign-ups, do not rule out economically significant numbers of sign-ups. By contrast, Yahoo! Run-of-Network ads outperformed the Yahoo! Mail ads for this advertiser, generating 278.8 new sign-ups for the 2.2 million individuals that were exposed to the ads, relative to an equivalently sized control-group baseline. The appropriate baseline amount is a bit subtle—the table above shows a scaled control baseline of 9,460 for all Yahoo! users, which looks like only a 2.9% increase in conversions—but this is not the relevant baseline since it includes conversions from the entire Yahoo! population rather than merely those exposed to the ad campaign. As we saw in

Table 1, only 2,558 conversions came from those who saw the Run-of-Network ads. With the appropriate denominator in mind, we find a 13.9% increase in conversions among the LREC ad viewers and an 8.0% increase among the Banner ad viewers exposed to the two Run-of-Network treatments.

In order to consider the cost effectiveness of this advertising, we need some appropriate economic context, considering not just the estimated marginal effect of these ads on account sign-ups, but also the possible lifetime value of a newly acquired customer. While precise details about the advertiser's finances cannot be included without compromising the advertiser's anonymity, a back-of-the-envelope calculation (using financial statements to estimate the cash-flow benefit per year of the average customer) suggests that an average customer's value per year is on the order of 25% of the cost per new sign-up[10] expended by the advertiser. This rough calculation of estimated benefit ignores various complications, such as that new customers may be more or less valuable than the average customer, that new customers may not last multiple years with the service, and that ads might have long-run positive benefits, such as subsequent word-of-mouth referrals. Our calculation suggests that the advertiser would recoup the costs of these new-customer-acquisition ads within approximately four years. However, utilizing the knowledge gained by this experiment could halve that figure for future campaigns, by encouraging this advertiser to focus on Run-of-Network rather than Mail placements of these particular ad creatives. Furthermore, recognizing that Run-of-Network ads may include as many as 60% Mail ads, targeting the non-Mail Yahoo! subdomains might substantially improve the cost effectiveness of this particular online advertising campaign on Yahoo!

Our results show the potential for scientific experimentation to help advertisers improve their campaign targeting strategies. For example, looking only at the click statistics in Table 1 might have led the advertiser to believe that Mail Banner ads were more than twice as effective than Run-of-Network LRECs, since the clicker rates were more than twice as high. Looking instead at click-based conversions, the advertiser might have concluded that those two strategies were approximately equivalent in value.

[10] This calculation involves taking the point estimate for the number of incremental conversions, and dividing it by the actual cost expended by the advertiser on the ads in this experiment. The prices were entirely representative of other advertising campaigns with similar targeting at Yahoo!

But both of these standard methods for analyzing effectiveness would give the wrong answer. In fact, using an experiment to measure the true causal effects of the advertising, we reach the conclusion (tentatively, given wide confidence intervals) that the Run-of-Network LRECs outperformed the Mail banners. The experiment helps the advertiser to learn which targeting strategies are best able to find the users most likely to change their actual behavior as a result of the advertising.

V. Conclusion

Using a controlled experiment, we demonstrate how different types of online advertising may have very different effects on generating new customers. Targeting, location, media, creative, frequency, and many other factors can influence a particular campaign's effectiveness. For the parameters granted to us in this experiment, we find that the increase in new customer sign-ups attributable to online display advertising for this online business is approximately 11% for users shown Yahoo! Run-of-Network advertisements. Because the estimates of the Yahoo! Mail ads are bounded from above by around 15%, and since Run-of-Network ads include some Mail ads, more effective targeting and advertising product choice have the potential to deliver even greater efficacy of the ads to this particular business.

We also demonstrate that attribution models not based on experimental data can easily misattribute the causes of online conversions. For one thing, standard click-based models can overstate the effect of advertising by "taking credit" for users who clicked but would have converted in the absence of the advertising anyway. For another, which turns out to be even more important in our setting, click-based attribution can understate the effect of advertising by ignoring the effects on the typical 99%+ of users who view but do not click on the ad. We find that this effect on viewers represents at least 70% of all the incremental conversions due to this advertiser's campaign on Yahoo!

Our future research will focus on improving the technology with which we run advertising experiments, making it possible for us to mark the control-group members who would have seen the ads (perhaps by deliberately running other ads in their place). This will allow us to remove the noise represented by the more than 90% of users whose conversion data could not possibly have been influenced by the advertising and should allow us to shrink our standard errors considerably. Future research will focus on identifying and understanding underlying causes of the difference in performance.

We find that the impact of these ads on the 3.7 million viewers reached was realized on as few as 300 individuals. Because the advertising only influenced 0.01% of those exposed to the ads, potentially four orders of magnitude remain to improve the targeting in order to only deliver the ads to those who will be influenced. Rather than targeting the "core demographic" of customers who purchase the most, advertisers can use randomized experiments to obtain precise measurements and identify the different ad products that are most cost-effective and reach the subpopulations that respond the most.

References

Abraham, M. and L. M. Lodish. "Getting the Most out of Advertising and Promotion." *Harvard Business Review*, May-June 1990, pp. 50-60.

Ackerberg, Daniel. "Empirically Distinguishing Informative and Prestige Effects of Advertising." *RAND Journal of Economics*, vol. 32, no. 2, Summer 2001, pp. 316-333.

Ackerberg, Daniel. "Advertising, Learning, and Consumer Choice in Experience-Good Markets: An Empirical Examination." *International Economic Review*, vol. 44, no. 3, August 2003, pp. 1007-1040.

Anderson, E., Y. Hu, E. Brynjolfsson, and D. Simester. "Dynamics of Retail Advertising: Evidence from a Field Experiment." *Economic Inquiry*, vol. 47, no. 3, July 2009, pp. 482-499.

Anderson, Eric T. and Duncan I. Simester, 2003. "Effects of $9 Price Endings on Retail Sales: Evidence from Field Experiments," *Quantitative Marketing and Economics*, vol. 1, no. 1, March 2003, pp. 93-110.

Bertrand, M., D. Karlan, S. Mullainathan, E. Shafir, and J. Zinman, 2009. "What's Psychology Worth? A Field Experiment in the Consumer Credit Market," Yale University Economic Growth Center Working Paper No. 918, 2009. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=770389

Chiou, Lesley and Catherine Tucker. 2010. "How does Pharmaceutical Advertising affect Consumer Search? Working paper, MIT Sloan School of Management, <http://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=IIOC2010&paper_id=514>.

comScore, 2008. "comScore Media Metrix Ranks Top 50 U.S. Web Properties For December 2008." <http://www.comscore.com/press/release.asp?press=2685>.

Danaher, P.J. and G.W. Mullarkey. "Factors Affecting Online Advertising Recall: A Study of Students," *Journal of Advertising Research*, September 2003, p.252-267.

Dorfman, R. and Steiner, P. O. "Optimal Advertising and Optimal Quantity," *American Economic Review*, Vol. 44 (March 1954), pp. 826-836.

Eastlack, J. O. and A. G. Rao. "Advertising Experiments at the Campbell Soup Company," *Marketing Science*, 8(1) (Winter, 1989), pp. 57-71.

Hu, Y., L. M. Lodish, and A. M. Krieger. "An Analysis of Real World TV Advertising Tests: a 15-Year Update." *Journal of Advertising Research*, vol. 47, no. 3, September 2007, pp. 341-353.

Lewis, R. 2010. "Where's the 'Wear-Out'?: Online Display Ads and the Impact of Frequency," Yahoo! Research working paper.

Lewis, R. and D. Reiley, 2011a. "Advertising Especially Influences Older Users: A Yahoo! Experiment Measuring Retail Sales," Yahoo! Research working paper.

Lewis, R. and D. Reiley, 2011b. "Does Retail Advertising Work? Measuring the Effects of Advertising on Sales via a Controlled Experiment on Yahoo!" Yahoo! Research working paper.

Lodish, L. M., Abraham, M., Kalmenson, S., Livelsberger, J., Lubetkin, B., Richardson, B., and M. E. Stevens. "How T.V. Advertising Works: a Meta-Analysis of 389 Real World Split Cable T.V. Advertising Experiments." *Journal of Marketing Research*, vol. XXXII, May 1995a, pp. 125-139.

Lodish, L. M., Abraham, M., Kalmenson, S., Livelsberger, J., Lubetkin, B., Richardson, B., and M. E. Stevens. "A Summary of Fifty-Five In-Market Experiments of the Long-Term Effect of TV Advertising." *Marketing Science*, vol. 14, no. 3, part 2, 1995b, pp. G133-140.

Microsoft, 2008. "Engagement Mapping: A new measurement standard is emerging for advertisers," atlas Thought Papers, Microsoft Advertiser and Publisher Solutions.

PriceWaterhouseCoopers, L.L.C. 2010. "IAB Internet Advertising Revenue Report 2009." <http://www.iab.net/insights_research/1357>.

Simester, Duncan I., Sun, Peng, and John N. Tsitsiklis. "Dynamic Catalog Mailing Policies," *Management Science*, vol. 52, no. 5, May 2006, pp. 683-696.

Strong, Edward C. "The Use of Field Experimental Observations in Estimating Advertising Recall," *Journal of Marketing Research*, Vol. 11, No. 4 (Nov., 1974), pp. 369-378.

Zielske, Hubert A. "The Remembering and Forgetting of Advertising," *Journal of Marketing*, 23 (January) 1959, 239-43.