Best Practices for Conducting Online Ad Effectiveness Research

June 2011

Conducted for the IAB by Marissa Gluck, RadarResearch

Acknowledgements

The IAB is grateful to Marissa Gluck of radarresearch for developing and writing this best practices document. We would like to acknowledge the contributions made by the IAB Research Advisory Board. In addition, the IAB is grateful to the industry experts who were interviewed for this research effort.

IAB Contact Information

Sherrill Mane SVP Industry Services, IAB 212 380-4702 Sherrill@iab.net

Contents

Executive Summary

Online ad effectiveness research is an important tool for marketers seeking to understand how their campaigns perform. However, it is challenged by serious methodological limitations. Questions around recruitment, sample bias and deployment are hampering the validity of this research and undermining the industry as a whole. The growth in online advertising spend requires sound measurement and reliable methodologies to prove effectiveness.

By examining each component of online ad effectiveness research, the IAB hopes to determine best practices along a range of methodologies. To accomplish the goals of developing and articulating best practices, the IAB commissioned Marissa Gluck of radarresearch. This document provides a set of recommendations so that the industry can develop more rigorous research while employing better standard operating procedures during the scientific inquiry period.

While the intent is not to endorse one methodology over another, there are some clear trends emerging as industry dissatisfaction with live intercept recruitment increases. With live intercepts falling increasingly into disfavor due to lowered response rates, greater site clutter, and misalignments between campaign delivery and sample, panels are gaining favor with agencies, publishers and vendors. However, panels are far from an industry panacea today. If properly validated, panels can solve for some of the methodological deficiencies of intercept studies as we know them.

This paper looks across the spectrum of available methodologies to assess best practices in each phase of online effectiveness research, from planning to recruitment, deployment and finally, optimization and analysis. Within each phase, we examine the challenges the industry faces and propose prescriptive remedies to each of these challenges. The issues we look at include organizational hurdles such as planning and staffing, as well logistical and technological impediments such as cookie deletion and declining response rates.

Best practices and associated issues covered include:

Planning	Recruitment	Deployment	Optimization
When to use an ad effectiveness survey	Declining response rates	Survey timing	Statistical differences between demographic groups
Optimal timing	Cookie deletion	Optimal survey length	Data integrity
Lack of staff research experience	Representative samples		Role of digital in ad ecosystem
Cost of true experimental design			

As the supply chain for online advertising has become increasingly complex, it has also opened up possibilities for better methods. As industry frustration grows, there is also an upside: greater willingness to experiment, more innovation, and an investment in developing alternatives.

Background and Methodology

BACKGROUND AND PURPOSE OF THE REPORT

In August 2010 the IAB published a report looking at the strengths and weaknesses of online ad effectiveness research methodologies conducted by Paul J. Lavrakas, PhD (http://www.iab.net/insights_research/947883/interceptstudies). Building on that research, the organization wanted to create a set of IAB best practices for this kind of research—a set of recommendations for agencies, publishers, and vendors to ensure that ad effectiveness research has the greatest possible validity.

This document is not meant to be exhaustive in the scientific aspects of improving the research methodologies. Rather, it is intended as a best practices guide for the marketplace while the study of improvements to ad effectiveness research methodologies occurs. The document builds upon Dr. Lavrakas' previous report, and incorporates newly conducted interviews with relevant stakeholders.

METHODOLOGY USED TO CONDUCT THE EVALUATION STUDY

Multiple interviews with key stakeholders were conducted over the period of two months in February and March 2011. Executives on both the publisher and agency side were interviewed, as well as vendor executives. Both larger and smaller agencies were interviewed, and both research and analytics executives were represented to get a more comprehensive view of the landscape as well as how data from these studies is incorporated into the overall value chain.

These interviews also led to several experts sharing confidential print and online materials. Other relevant information was gathered via internet and academic searches for nonproprietary literature that had been published or presented on how the Internet advertising effectiveness is, and should be, measured.

Introduction

Online advertising effectiveness research is a critical component of the marketing ecosystem, used to validate the efficacy of campaigns and as a tool for both creative and placement optimization. According to The Interactive Advertising Bureau and PwC, Internet advertising revenues in the U.S. in 2010 were $26 billion. The growth in online advertising spend requires sound measurement and reliable methodologies to prove effectiveness. However, the industry faces considerable methodological challenges, many of which were examined in the IAB's *An Evaluation of Methods Used to Assess the Effectiveness of Advertising on the Internet* last year. The paper, which looked exclusively at site intercept studies, sounded the warning bell that the validity of interactive advertising effectiveness (IAE) research is «threatened...by the extremely low response rates achieved in most IAE studies.»

While the Web provides a greater level of accountability and measurement than other media, it is also hamstrung by multiple competing technologies, stakeholders and a complex supply chain. Early online advertising effectiveness research emulated the methodologies established in other media, such as television and print. A reliance on traditional test and control methodologies stems from traditional media, where randomized experimental design allows marketers to compare responses. True experimental design rests on the notion that both groups are identical in every way except exposure to advertising. Effectiveness is measured by calculating «lift,» or the difference between the test and control group. Metrics such as awareness, aided and unaided recall, brand favorability and intent-to-purchase have historically been measured between the two groups.

As a result, procedures were established that have evolved little in the past 15 years, while the underlying technologies, business practices and even consumer behavior have changed drastically. While this paper looks at the challenges wrought by these changes, it is not intended to simply expose and lament the Gordian knot of online ad effectiveness research, but rather to propose best practices, given the state of the industry today as well as recommendations for areas of future investigation.

Planning

Challenge: Advertisers have difficulty determining when using an online advertising effectiveness survey is most appropriate.

All too often, advertisers and agencies add a brand measurement research component at the last minute, without enough forethought or planning. The decision to use an ad effectiveness study often seems driven more by fear than a structured effort at understanding campaign efficacy. Advertisers and their agencies should evaluate the use of an ad effectiveness survey as early in the planning process as possible.

Best Practices:

- **Agencies and Advertisers**
    - o The decision to employ an online ad effectiveness study should primarily be driven by campaign goals, such as increasing brand awareness or favorability. A brand-driven advertiser should never rely chiefly on click through rates as a measure of effectiveness. For advertisers whose main objective is direct response, online ad effectiveness studies can be used as a complement to other metrics.

    - o As Dr. Lavrakas previously explained, online ad effectiveness surveys «should only be used in conjunction with campaigns of a significant size.» Yet the size, he argued, will vary from site to site. To use online ad effectiveness surveys appropriately to make optimization decisions on an individual site, a study will optimally include at least 200 respondents. However, this number may be smaller for some sites, especially as the number of concurrent surveys increases, making recruitment a greater challenge. Ultimately, sites need to balance the user experience with the need for robust data, sometimes resulting in a lower number of respondents. In these cases, sites should strive to recruit at least 100 respondents into a brand effectiveness survey.

    - o In order to reach that threshold, publishers should provide guidance to agencies on the minimum number of impressions necessary. Campaigns that don't meet the minimum threshold risk a lower number of respondents and consequently, a potential misuse or misunderstanding of the available data. Publishers should not facilitate agency studies that do not meet their thresholds for statistically stable research.

- **Publishers**
    - o Publishers should determine a minimum spending and impression threshold for running an online ad effectiveness study. In terms of impressions, publishers interviewed believed a 15 million impressions campaign should be the minimum before a site should consider running a study on behalf of advertisers. However, that minimum will vary from site to site and campaign to campaign based on the recruitment method, response rates and number of studies on the site at any given time. While 15 million impressions is a general guideline, publishers should determine their own rules for feasibility and ability to execute based on their learning from past studies.

    - o One rule of thumb is the cost of an effectiveness study should be less than 10 percent of the overall buy.

Challenge: Advertisers and agencies often request online advertising effectiveness studies immediately prior to a campaign launch.

Online advertising is a dynamic business, optimized often in real time, and subject to myriad externalities beyond a single stakeholder's control. However, agencies and advertisers often let publishers know they'd like to run an online advertising effectiveness study with very little notice at the launch of a campaign. Multiple publishers recounted agencies asking for a brand measurement study with less than 24 hours to go before a campaign launched.

There seem to be two primary reasons this occurs. First, agencies (or their clients) decide just prior to launch that a brand effectiveness study is necessary as a complement to other campaign metrics to justify the ad spend. In these cases, the advertising effectiveness is treated as an after-thought, sending publishers scrambling to meet their demands. The second reason is somewhat more mercenary and, it should be noted, occurs very rarely. Some agencies have an irrational fear that informing a site of plans to run a client's effectiveness study too early will result in the site allocating its premium inventory to that marketer and relegating their other clients to less premium inventory. By informing a site last minute of their intention to deploy an effectiveness survey, they reason, the site will not have the time to re-traffic and re-distribute their clients' campaigns.

While the former reason indicates agencies take for granted their ability to run a survey on any site, at any time, the latter speaks to ongoing mistrust between some agencies and publishers. However, the former can be addressed through several recommendations:

Best Practices

- **Agency & Advertiser:**
  - Online advertising effectiveness research should be considered and included (when appropriate) in the planning stages of a campaign. Including the objective and parameters on Requests for Proposals should be standard. For agencies and advertisers that consistently use the same vendors, RFP's should also inform sites that the agency or advertiser intends to use one of the listed vendors. Additionally, the RFP should include as much information as possible on publisher responsibilities such as guaranteed sample size, tag implementation and including control ads in random rotation with exposed ads.

  - Agencies should include recruitment impressions in their insertion orders prior to the start of the campaign. Additionally agencies and advertisers should ask permission from the publisher to run an effectiveness survey rather than assume it's an inviolable right.

  - Four weeks is ideal to allow ample time for tag implementation and tagging, and the amount of time most requested by media owners. Agencies should remind publishers at least fourteen days prior to the start of a campaign or flight of their previously agreed upon intention to run an online ad effectiveness study.

  - Agencies and advertisers should allow publishers to be able to review and approve surveys before fielding them live.

- **Publisher:**
  - Publishers need to become more proactive in the planning stages as well. When responding to an RFP, publishers should ask agencies and advertisers to identify their chosen research vendor prior to the start of a campaign. Additionally, if extra impressions are needed for the survey, that should be noted on the insertion order. Publishers should also ask advertisers and agencies to share contact information at their vendors.

o Publishers should require pre-approval for advertiser surveys to ensure the questions do not conflict with their editorial policies or media brand. To avoid conflict, publishers should provide advertisers, agencies and vendors with guidelines for survey research.

o Publishers should provide advertisers with a list of preferred or approved vendors for brand effectiveness studies in the planning stages of an advertiser campaign.

- **Vendors**:
  o In addition to greater involvement from the publishers prior to the start of a campaign, vendors should also proactively reach out to publishers included in a measurement survey. This provides a degree of redundancy to help publishers prepare for an upcoming study.

Industry Opportunity

The IAB and the 4A's are in a unique position to offer a standard boilerplate, similar to the industry Terms & Conditions for media buys (http://www.iab.net/iab_products_and_industry_services/508676/tscs3). The industry should develop a boilerplate document for inclusion in agency RFP's that helps mediate terms for this type of research between advertisers, their agencies and publishers.

Challenge: All too often, agency and publisher staffers have a lack of research experience and education.

The media business, particularly in the planning and buying departments at agencies and sales departments at publishers, depends on relatively young, sometimes inexperienced staffers. While many of these staffers receive training on media math, negotiation tactics and agency or site processes, they are often less knowledgeable about research methodology and implementation. «Research education is a huge issue,» says Forbes's Bruce Rogers. «We need to have a more educated conversation about rolling out research.» Agencies seem to agree. «The demand for talent is extraordinary,» says Andy Fisher, EVP Global Data & Analytics Director at Starcom MediaVest.

Best Practices:
  o Agencies should include training and education for their media staffers not just on the mechanics and math of planning and buying but also on the basics of research methodologies and implementation.

  o Training should emphasize how research objectives are married to campaign objectives, what the limits of certain methodologies may be and how best to use the data for optimization. Agency staffers should understand how to use these studies as a creative optimization tool and be able to evaluate if their campaign is communicating to their target audience the intended messages.

  o Training should include orientation to the in-house experts for advice and guidance so that decisions are made using the greatest degree of understanding possible.

Challenge: True experimental design is both costly and operationally inefficient.

The challenges of using experimental design in online ad effectiveness research were explored by Dr. Lavrakas previously. The complexity of online advertising's ecosystem makes experimental design very difficult to implement rigorously. Additionally, it can add substantial cost, both in terms of media spend and operational costs. One vendor estimated the cost of implementing a rigorous test would be $50,000-80,000 versus the typical cost of studies in the $20,000-25,000 range. As one vendor explained: «Low cost experimental design does not exist.» While some may disagree with this assessment, it is certainly true that true experimental design is typically more expensive than today's studies. Advertisers and agencies must also be prepared to encounter technological challenges in implementing experimental design. While third-party ad servers are theoretically designed to segment audiences, the technology is typically used for targeting and sequential messaging. The platforms were not built specifically to accommodate this kind of research. Vendors estimated that five to ten percent of media inventory should be allocated to the control group (using a «placebo» ad, such as a PSA) and eight to fifteen additional hours of work is necessary to train staff and set up the campaign.

  Best Practices:
  o As Dr. Lavrakas explained, true experimental design is almost non-existent online due to the cost and complexity. While most advertisers are engaged in what he terms «quasi-experimental design» the ugly truth is that the industry has resigned itself to accepting a sub-par research methodology in exchange for lower cost and greater efficiency. While the industry made concessions to accommodate the technology, stakeholders and processes involved in online advertising over a decade ago, the industry should strive to develop better, more efficient methodologies.

  o Given the complexity of online advertising's ecosystem, it's unlikely there will ever be a «perfect» methodology – the industry will always be forced to accept methodologies that are «good enough.» However, the standard for what is deemed «good enough» needs to evolve and rise rapidly. One agency interviewed is working on a solution that may point to a potential alternative. Using raw log file data from its third party ad server, the agency regularly de-duplicates the data to identify unduplicated reach and frequency and assess recency. Attempting to match behavioral and attitudinal behavior, the agency is able to recruit respondents into effectiveness surveys using user ID's and information on who has been served what ad, in a particular sequence. The agency targets exposed and control groups by purchasing profiles from data exchanges via the user ID's. While the agency concedes the methodology isn't perfect, it believes the ability to merge third party ad server data with consumer profiles through the ad exchanges and demand-side platforms might provide a hint towards the future of online ad effectiveness research. «This wasn't possible a year ago,» explains the director of analytics for the agency. While the work required to accomplish this depends heavily on manual processes, increasing man hour costs, it points to possible future methodologies available to agencies.

o   The fact that agencies, marketers and publishers must experiment with ad hoc work-arounds (such as the one discussed above) underscores the need for industry commitment to find better solutions. The industry should explore funding more research on developing new methodologies. Greater investment, transparency and structured experimentation are necessary to find and fund industry standards for ad effectiveness research.

Recruitment

Challenge: Response rates to intercept invitations are declining.

With the increase in brand effectiveness studies, consumers are starting to show signs of fatigue. Almost every agency and publisher interviewed admitted response rates were declining for site intercept invitations. Yet marketers also need to collect enough responses per each reportable data cut. The potential increase in surveys is creating clutter and having the adverse effect of lowered response rates. While this is an industry-wide problem with implications for multiple kinds of market research, there are some potential solutions.

Best Practice:
- **Agency & Advertiser:**
    o   There are two potential approaches to addressing this issue. The first, adding incentives, would not be recommended in most cases due to the additional cost and complexity. In rare exceptions, incentives may be desirable when the target audience is unlikely to respond without them. For instance, for highly specialized professionals or C-level respondents an advertiser may look to incentives to motivate response, but incentives should be considered a last resort.

    o   Another potential alternative is the use of panels to recruit respondents. By pre-tagging panel members, the advertiser and vendor can identify potential respondents who match the campaign's target audience, triggering an ad exposure and emailing a survey invitation. While today live recruitment via intercepts is still the most common recruitment method, the industry will likely use panels more and more often as concerns about test and control samples grow. (See Sidebar: *When to Use a Panel*). **Panels are not a magic bullet for the problems the industry is experiencing with live intercepts, yet do offer some promise if properly constructed.** As industry acceptance of recruitment through panels grows, scientific investigation should grow accordingly to better understand potential weaknesses as well as opportunities.

    o   Results should be weighted to actual campaign delivery and, when possible, to campaign demographics as the panel sample base can vary widely from the site population.

- Further research should be conducted on the impact of timing and decay. With respondents taking surveys at a different period of time post-exposure, another variable is added to the analysis. Research should be conducted by advertisers, publishers and vendors to determine what effect, if any, survey timing has on results in order to model timing effects.

- **Publisher:**

  - Publishers need to take some responsibility for declining response rates as well. Publishers should try to limit the frequency of survey invitations to protect the user experience and reduce clutter. This will require more coordination between sales, research and operations. Often, the sales department is not aware of how many surveys are running simultaneously on its site. Better communication between these departments with frequent updates on survey research can help alleviate the burden of concurrent studies.

  - In order to gain a better sense of survey activity on their sites, publishers should create a brief exit report for each study. The report should summarize activity and metrics such as total number of completes, ad spend, cost of survey, impressions and reach/frequency.

  - A panel approach may also benefit publishers as well, since it eliminates the operational effort on the publisher side as well. Additionally, with fewer surveys running live on their sites, the risk of survey «burnout» is lowered. It also reduces the cost for implementing the research since the publisher doesn't need to allocate control impressions for the survey.

- **Vendors:**

  - Panels need to be representative and not biased. To build industry support for panel solutions, vendors should conduct validation studies. The industry should create auditable standards for what constitutes a valid sample for purposes of this kind of research and require that vendors meet the standards.

When To Use A Panel

Benefits of Using a Panel: *Validated* panels provide advertisers the ability to gain longitudinal data over time to assess long-term branding impact. The use of panels also eliminates some of the operational inefficiencies caused by live intercepts such as QA'ing web intercept invites since the recruitment is done via email. Previously, the size and scale of the panels was a challenge for advertisers serving billions of impressions in a campaign but as the size of panels have grown in the past few years, it is becoming increasingly feasible to use them as an alternative to traditional live intercepts to recruit respondents, provided the panels are representative and scientifically validated.

There are several other benefits to this approach including the ability to weight against people who saw the campaign, eliminating variance between test and control groups and the ability to cut the data with greater granularity. A panel solution also eliminates much of the burden on users, has much higher cooperation rates than live intercepts and a large sample can be recruited fairly quickly in comparison to other methodologies. «We can get a robust sample and [the survey] doesn't have to be in field that long,» explains Bruce Rogers, Chief Brand Officer for Forbes.

While the use of panels to recruit respondents may be an attractive alternative to live intercept recruitment, there are certain types of campaigns and circumstances that lend themselves more easily to the use of panels:

- **High volume campaigns**. Given the size of panels (while growing) can still be constrained, advertisers with a high volume of impressions are best served by panel recruitment. Smaller, more targeted campaigns may be constrained by the size of the panel.
- **Multi-platform campaigns**. For advertisers running campaigns on multiple platforms (Web, mobile, tablets, game consoles) the use of a panel is essentially the only viable option for sample recruitment. For multi-platform campaigns, marketers should include questions on the survey about device and app usage to filter respondents.
- **Customized ad placements.** A campaign with customized ad units such as home page takeovers, roadblocks or integrated sponsorships can avoid the biases inherent in live intercepts for these units by recruiting via panel.

Challenge: Recruiting a clean control group is increasingly difficult due to cookie deletion, emerging ad formats and a complex ad delivery value chain.

Cookie deletion is an increasing phenomenon as consumers are given more tools to easily control their cookies. A 2010 study by former IAB CEO and President Greg Stuart stated 36 percent of Internet users delete cookies on a monthly basis. Other recent estimates by Jupiter and Belden put the percentage closer to 30 percent; though it is not clear how often users delete their cookies. The increased prevalence of flash cookies (which are somewhat harder to delete) may offer some short-term aid, but the growth of mobile devices and tablets further complicate the landscape, as these devices don't consistently set cookies for their users and most don't accept flash.

As Dr. Lavrakas previously pointed out, the actual rate, while still unknown, is «likely to be high enough to bias coverage to at least a small extent.» As a result, many people who are counted in a control group may in fact have been exposed to the ad. According to ComScore's Anne Hunter, deletion levels vary by site. Sites with heavy repeat usage are more prone to deletion, «which means errors in response can negate any value in optimization.» Additionally, longer campaigns are also more likely to face cookie deletion. As a result, with cookie deletion levels varying from site to site and by campaign, no single weight can compensate for this issue. In addition to cookie deletion, the use of shared computers between multiple users and a single user with multiple machines also complicates advertisers' and publishers' ability to recruit a clean control group.

The emergence of new ad formats and placements also creates greater difficulty in guaranteeing representative test and control groups. Sponsorships that may not be delivered through a third party ad server, homepage takeovers, roadblocks and niche-targeting are increasingly sought after inventory, yet makes it more challenging for advertisers and publishers to allocate inventory for a control group that matches the exposed group. On sites where share of voice is high, recruiting a control group can be extremely difficult (though some sites have used interstitial invites in these circumstances with success). Similarly, when ads are concentrated in a particular section of a site, as in a roadblock or sponsorship, advertisers and publishers are challenged to find an equivalent control group from the same inventory.

Another issue for control group recruitment is publisher reluctance to allocate equivalent inventory. Often the inventory a publisher will use to recruit a control group is run-of-site and less premium than the inventory used in the exposed ads. Essentially, the two groups (control and exposed) are not the same, resulting in methodological biases.

One other frustration voiced by agencies is the sample composition of their ad effectiveness surveys. Too often, the sample in the survey doesn't match the target audience, frequency and distribution of the buy. «If 50 percent of our impressions and reach comes from the *NY Times* and 10 percent comes from *Vice*, then 90 percent of our respondents can't come from *Vice*,» explains Rudy Grahn, former VP, Director of Analytics at Zenith-OptiMedia. Additionally, agencies and publishers believe there is an excessive burden to police vendors throughout the process. «For a $20,000 survey, we spend $100,000 worth of man-hours babysitting the vendors,» said one agency executive.

Best Practices:

- o If test/control brand lifts are calculated, control exposure should not be contaminated by people who deleted their exposure cookies. Test/control lifts can only be calculated if it can be guaranteed the control respondent has not been previously exposed to the campaign. Cookies are rapidly becoming an obsolete method to designate a control group.

- o Too often, publishers bemoaned the fact that vendors didn't seem to be aware of their responsibility to ensure more equitable recruitment distribution. Vendors need to play a more proactive role in the research value chain to ensure recruitment rates are aligned with the media plan.

- o Advertisers, publishers, and vendors should conduct further research to better understand the methodological flaws and potential opportunities afforded by scientific sampling as an alternative to today's sampling techniques.

Is There a Solution to Cookie Deletion?

As Dr. Lavrakas explored in some depth, the cookie deletion issue is a somewhat intractable problem for effectiveness research. One potential alternative can be found in scientific sampling, whereby an advertiser can measure the number of times a person was eligible to be sampled and then later weight the data based on the probability of selection into the sample. However, there are potential problems with this methodology too, as Dr. Lavrakas acknowledges. One issue is relying on self-reported data on site visitation to get an accurate measurement of the number of times a person could be sampled. Dr. Lavrakas contends the relative differences across people would provide the information needed to make accurate weighting corrections. This opens up a potential area of further examination for advertisers, publishers, vendors and the IAB.

Another possible alternative is a modeled lift approach to online ad effectiveness research. Similar to traditional market model mix research, it eliminates the possibility of a contaminated control group due to cookie deletion. Traditional media planning has used advertising response curves to model the impact of advertising on consumer attitudes and behavior. This model could potentially be used to predict the probability of a response based on exposure levels and can include demographic and psychographic information as control variables as well. This approach also allows advertisers to measure ad placements and units with total exposure such as home page takeovers, roadblocks and sponsorship. However, more research should be conducted to understand the correlation between frequency of exposure and results, as well as impact of placement.

Challenge: There is a strong industry suspicion that respondents are not representative of target audiences.

Almost every agency and publisher reiterated this suspicion: «Survey takers are a slightly different demographic than non-survey takers.» It's also a belief that both parties have resigned themselves to as a «fact of life for all research.» While this may be true, there are two paths the industry can pursue to address this issue.

Best Practices:
- o Agencies and publishers should pressure vendors to undergo independent, third party evaluations of their statistical methods and recruitment methodologies. The Media Ratings Council (MRC) can provide impartial validation. Several vendors have already undergone or been in the process of undergoing audits by these third parties.

- o To understand if there is an egregious difference between survey respondents and the campaign audience, surveys should at the very least include demographic questions in order for marketers to compare sample to audience. Brand advertisers, publishers and vendors need to ensure enough responses are collected for each reportable data cut in order to compare.

- o Vendors should weight results to actual campaign delivery, using third-party audience verification companies as a benchmark.

- o A second option is to forego traditional intercept surveys in favor of the panel approach discussed earlier. While there is still likely to be a bias based on the respondents' willingness to be included in a panel and respond to surveys, as third-party panels increase in size, marketers are more likely to find a sample that closely matches their campaign delivery. Slightly more challenging is finding the comparable control group. Dr. Havlenas at Dynamic Logic recommends marketers «make sure web visitation and behavior patterns look similar» to the exposed group. One way to do this is by using exposure to similarly targeted campaigns as a proxy for matching the control to the exposed group.

- o Ideally, control and exposed groups should be recruited from the same sites and target audience. However, in the absence of this scenario, behavioral questions such as internet usage and category engagement should be asked in the survey and used to weight the control to the exposed group if recruitment from the same placements wasn't successful at reaching a match between the groups.

Deployment

Challenge: Surveys often take place during exposure.

Another issue plaguing online ad effectiveness surveys is the timing of the survey intercept invitation, which can occur while the measured ad is still visible on the screen. This is not always due to negligence on the part of publishers, but often due to the preponderance of ad networks and exchanges. As a result, if the survey question is shown on the same page as the measured brand ad, the exposure is typically biased positively. With sites using different servers, ad networks and data exchanges, one site or sites may experience this more often than others, thus negating value for optimization. This is particularly an issue with in-banner surveys if other ad units display on the page with the survey. Multiple ad units for a single brand typically display simultaneously in video environments and in-banner surveys shouldn't be used in these environments unless that exposure can be controlled. It would be extremely difficult (if not impossible) for an agency or advertiser to measure the varying levels of contamination, rendering them effectively unable to compare site performance.

Best Practices:

- o All survey questions should be asked in an environment where the measured brand ads are not visible. This would require vendors to deliver surveys that cover the entire screen or develop a technology that can measure all of the ads appearing on the page (both within and out of iframes) while a survey is deployed.

Challenge: Excessively long surveys can depress completion rates.

Designing online surveys is both art and science. Designing a survey usually requires a tacit trade-off between the number of relevant questions that can be asked and the optimal length consumers are willing to respond. For certain categories, such as B2B or surveys aimed at C-level professionals, finding the optimal length is critical to completion. When surveys are too long, not only does it affect the consumer, but it also reflects poorly not on the advertiser but on the publisher, who is ultimately responsible for the user experience on its site. In addition, lower completion rates require the publisher to allocate extra inventory to the survey in order to meet guarantees, creating an undue burden on sites. Lastly, excessively long surveys can impact validity as respondents become tired and distracted, resulting in less thoughtful answers. Data from Dynamic Logic reveals indexed completion rates begin to drop significantly on surveys longer than four minutes in length. Additionally, note that the sample size for longer surveys is significantly smaller than that of shorter surveys.

Average Time (in minutes)	Indexed Completion Rate (<4 min =100)	N
1:00-3:59	100	80
4:00-6:59	90	237
7:00-9:59	76	249
10:00-12:59	70	66
13:00-15:59	70	32
16:00+	69	32

Source: Dynamic Logic, May 2011

Best Practices:
- o Surveys aimed at business professionals should take less than ten minutes to complete.

- o Four minutes or less is ideal and will improve completion rates.

- o For consumer-targeted products, a longer survey may be possible though advertisers should be wary of exceeding seven minutes. The «sweet spot» for survey length seems to be five to seven minutes long or approximately 20-25 questions.

- o Surveys should show respondents their progress throughout the questionnaire, visibly demonstrating the percentage of questions answered at any given time.

- o Agencies and publisher should extensively QA surveys to ensure median response times are within the recommended time. Additionally, vendors can be helpful in pinpointing optimal times. Vendors should also be able to eliminate consumers who have responded with a «straight line» through the survey, indicating they were not truly answering the questions.

- o While shorter surveys are preferable, marketers should include a minimum of qualifying questions unrelated to weighting. For instance, a survey for an IT product can ask basic technology questions in order to qualify respondents as appropriate to the target market.

o   Though there has been an increase in the number of vendors offering single question studies, the value of this method is limited. Without the ability to segment audiences, the insight provided by these studies is extremely narrow. Additionally, the inability to weight the control group to the exposed sample makes the results of these studies potentially invalid since all of the variables between groups cannot be controlled for and therefore the changes between the groups can't be isolated to the single variable of exposure.

Optimization and Analysis

Challenge: There are statistical differences between groups, such as men and women.

According to several vendors, there are certain trends in demographic groups that are consistent across campaigns and advertisers. For instance, women tend to have higher levels of awareness and purchase intent lifts than men. If a campaign is over-represented by female respondents, there can be significant discrepancies between reported lifts and actual effectiveness.

Best Practices:
o   As mentioned previously, vendors need to weight results to actual campaign delivery to avoid misleading statistical lifts.

Challenge: Marketers need greater faith in their data in order to optimize their campaigns.

Brand marketers are reliant upon their online advertising effectiveness research (often in conjunction with other metrics such as click through rate or interaction rates) to optimize their creative and placements. They are increasingly demanding more precise cuts of their data to produce actionable results and to better understand impact by publisher site/category, demographic group, creative format, and placement.

Best Practices:
o   When feasible, marketers should marry their campaign effectiveness research to salient online and offline metrics to get a more comprehensive view of their advertising efficacy. This may include visits to their site, time spent, branded or trademarked search behavior and offline sales.

o   Another area for future investigation looks at the role online advertising plays in the overall brand marketing campaign. Digital media is one component of the overall branding strategy. Marketers, agencies and publishers should invest in research that provides insight into how online advertising fits into a comprehensive marketing plan.

o Vendors and publishers should provide better post-survey information on methodology. Ideally, they will provide a one-pager that addresses not only the sample frame of exposed and control (as they do today) but also data on recruitment invitations, sampling rate, click-throughs, number of surveys started, surveys completed, total number of questions, time to complete and flat-liners (when the time between answers drops to <1 second). More importantly, there is a need to provide normative comparisons in order to identify outliers. If possible, publishers should also provide a reach-frequency curve of both the campaign and the survey respondents.

o Even as optimization has been trending towards real time, brand marketers should be wary of making optimization decisions based on real time ad effectiveness data. Branding is a long-term endeavor, so marketers should exercise caution before making hasty, possibly ill-informed optimization decisions. For instance, creative that shows positive results at a single exposure may not perform as well with a higher number of exposures or in combination with multiple exposures. Advertisers need to take frequency, sequencing and exposure over time into account before making optimization decisions. The minimum amount of time will vary by advertiser but for products with more complicated decision processes, more time in the field is necessary to assess brand impact. For instance, automotive or financial services (both categories with complex decision processes) generally take longer to move the needle on branding metrics like favorability than consumer packaged goods.

o Sample size should be large enough for advertisers to distinguish «noise» from statistically reliable data. The size of a sample will vary from product to product, brand to brand and advertiser to advertiser but should be large enough to provide actionable insight. Agencies and marketers should involve their analytics team(s) to determine optimal sample size and relay that to publishers.

Conclusion

While online ad effectiveness research is an important tool for marketers seeking to understand how their campaigns perform, it is challenged by serious methodological limitations and irresponsible study management. Questions around recruitment, sample bias and deployment are hampering the validity of this research and are undermining the industry as a whole. The growth in online advertising spend requires sound measurement and reliable methodologies to prove effectiveness.

More research needs to be done to better understand many of the outstanding questions related to brand effectiveness research, particularly as the industry increasingly looks towards panels as an alternative to today's methodologies. The best practices in this document are intended to provide antecedent guidelines for standard operating procedures in the interim, as the industry continues to engage in empirical investigation and scientific inquiry.

By examining each component of online ad effectiveness research, the IAB hopes to determine best practices along a range of methodologies. While the intent is not to endorse one methodology over another, there are some clear trends emerging as industry dissatisfaction with live intercept recruitment increases. One of the major shifts emerging is the use of panels to recruit respondents. As live intercepts fall increasingly into disfavor due to lowered response rates, greater site clutter, and misalignments between campaign delivery and sample, panels are gaining favor with agencies, publishers and vendors. Yet more research is needed to better understand the strengths and potential flaws in panel recruitment. The Best Practices provide some fundamental guidelines for when and how to approach the use of panels.

By examining each phase in the delivery of online ad effectiveness research, we hope to not only understand the complexities and challenges inherent in this research but also best practices to remedy the industry's most vexing problems. While the supply chain for online advertising has become increasingly complex, it has also opened up possibilities for better methods. As industry frustration grows, there is also an upside: greater willingness to experiment, more innovation and an investment in developing alternatives.

Appendix

The following is a list of executives interviewed for this research effort:

Yaakov Kimelfeld, Ph.D.
Senior Vice President, Digital Research and Analytics Director
Starcom Mediavest Group

David Cohen
SVP, Interactive Media Director
Universal McCann

Jim Nichols
Partner, Strategy
Catalyst SF

Andy Fisher
EVP Global Data & Analytics Director
Starcom Mediavest Group

Rudy Grahn
Former VP, Director of Analytics
Zenith-Optimedia, recently moved to Clearsaleing

Dan Murphy
SVP, Research & Web Analytics
Univision Interactive Media

Keith Berkelhamer
Vice President, CNN Ad Sales Research
CNN.com

Sylvia Barney
Global Research Director
Microsoft Advertising

Bruce Rogers
Chief Brand Officer
Forbes

Jim Dravillas
Head of Advertising Research
Google

Taylan Yildiz, Ph.D.
Display Research Lead
Google

Anne Hunter
VP Ad Effectiveness Products
comScore

Drew Lipner
VP, Group Director
InsightExpress

William J. Havlena, Ph.D.
Vice President, Research Analytics | Client Solutions, CrossMedia
Dynamic Logic - A Millward Brown Company