# How Often Rather than How Many

## Ad Latency for Media Planning

**Steven Millman, comScore**
**Alex Wakim, comScore**

## Abstract

Digital media planners spend a great deal of time and energy attempting to optimize the return on investment (ROI) from their digital campaigns. One of the most commonly used optimization tools is frequency capping, which provides the ability to prevent consumers from being exposed to an individual campaign more than a certain number of times. This research suggests that a more sophisticated approach using ad latency – the effective period of an ad exposure – may provide substantial improvements in ROI over frequency caps. Key findings of this research include: the ability of ad latency periods to be measured on a campaign by campaign basis; that an ad's effectiveness appears to follow a concave logarithmic decay curve rather than the commonly assumed convex half-life (exponential) decay; that video ads appear to have a significantly longer latency periods than display ads; and that changing the way ad servers determine whether to fire an ad using latency periods would offer media planners with a variety of effective new approaches to maximizing ROI.

## Introduction

Every year, comScore conducts thousands of advertising effectiveness studies focused on a variety of media measurement topics. Within these expansive efforts, the most common questions driving research are centered on how to optimize media plans around frequency of exposure.

It is an expected practice that ad servers provide the ability to identify how many times a web browser has been served a particular ad impression and make ad hoc decisions regarding whether or not to serve the next. There is a long-standing belief which has been well documented in market research, that the value of advertising follows a diminishing marginal utility curve. In other words, ads get less effective the more times an individual is exposed to them. Additionally, as ad effectiveness decreases, so does overall ROI of the campaign. Further, if too many ad impressions are served to a small group of individuals, fewer potential consumers are exposed and campaign reach is constrained. These concerns demonstrate why it is critical to determine the optimal frequency of exposure for any given campaign.

Underlying this question of an optimal frequency cap is a key assumption about how digital ad campaigns impact consumers. It is assumed that the timing of advertising is of little consequence to the declining utility of the ad. It seems reasonable to believe that the tenth ad seen in a day would have less impact than the first. But does that relationship hold when the first nine ads are seen on one day, and the tenth three months later? If ads are not retained indefinitely in the mind, then the presumption of a lower utility for subsequent ads is not universal. Using frequency caps, there is no differentiation between an individual that is exposed ten times in the first few days, exposed over large gaps of time, or exposed evenly over the course of the campaign. In order to answer the question of optimal delivery of digital ads, it is important to consider the latency period of an ad - the amount of time it takes for an effective ad to lose its influence on the individual exposed.

comScore measures ad effectiveness in a number of ways, including through its AdEffx Brand Survey Lift™ product, which quantifies the lift in survey responses to brand metric questions using a non-linear regression approach to measurement called Smart Lift™, currently under patent review. Rather than leveraging a more traditional methodology which compares responses between a test and control group, Smart Lift treats each impression as an independent event with individual characteristics. These characteristics include, but are not limited to, creative, publisher, placement, strategy, and, viewability. Since there is a distribution of advertising exposures across the survey sample, we can construct a curve that defines the declining marginal utility of the ad campaign. Unique to the market and similar in design to widely-accepted Market Mix Models, Smart Lift models a baseline response which provides an estimate for the non-exposed, eliminating the need for an actual control group. This allows comScore to accurately estimate what the expected lift of a campaign would be for any average frequency of exposure.

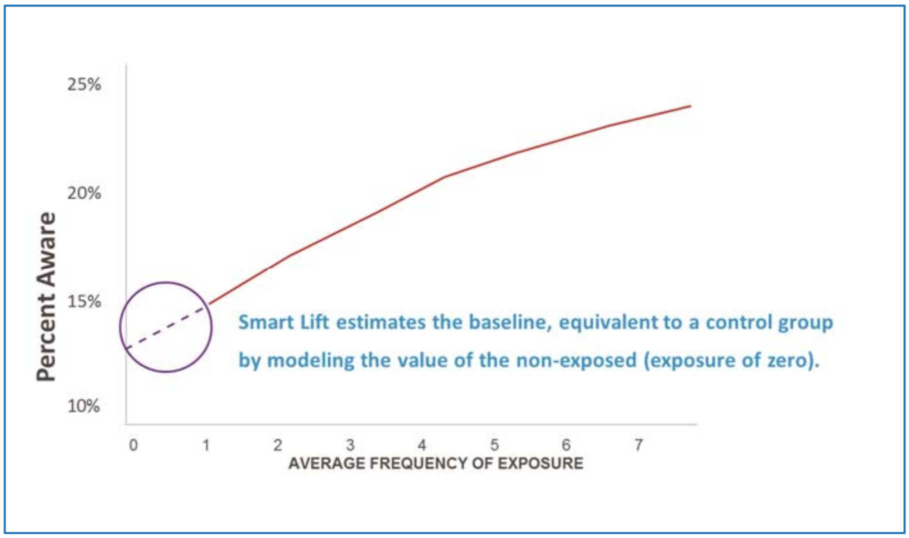


**Figure 1:** Smart Lift Model Lift Measurement

Because this methodology is able to look at each impression as an independent event, comScore is able to also consider the amount of time elapsed between each survey exposure and the time that the survey was completed. Ads are discounted in their capacity to affect survey response based on their time since exposure. In television advertising, it is common to consider the half-life of the effectiveness of an ad, which assumes a decay curve that rapidly decreases in effectiveness initially, and then maintains a long minimally effective tail, as illustrated in **Figure 2:** ***Ad Latency as Half Life***.

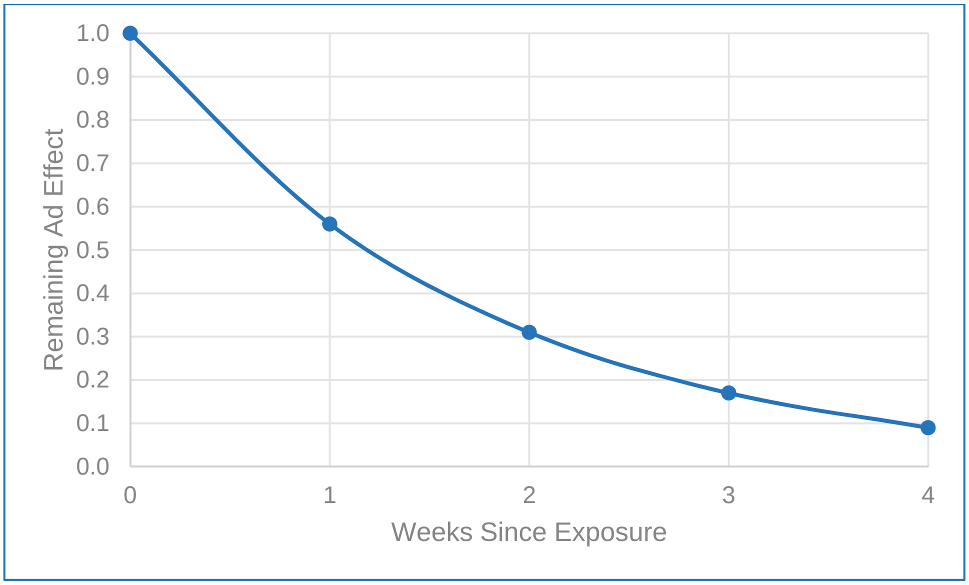


**Figure 2:** Ad Latency as Half Life

This notion of an ad's half-life decay has largely been transferred over to the digital advertising arena. Consider the two examples below, both of which have an individual exposed five times to the same campaign. In the first case, the ads are spaced evenly one week apart. When taking into account the effect of decay, these five exposures would have about the same effectiveness as having been exposed 2.13 times on the day of survey.

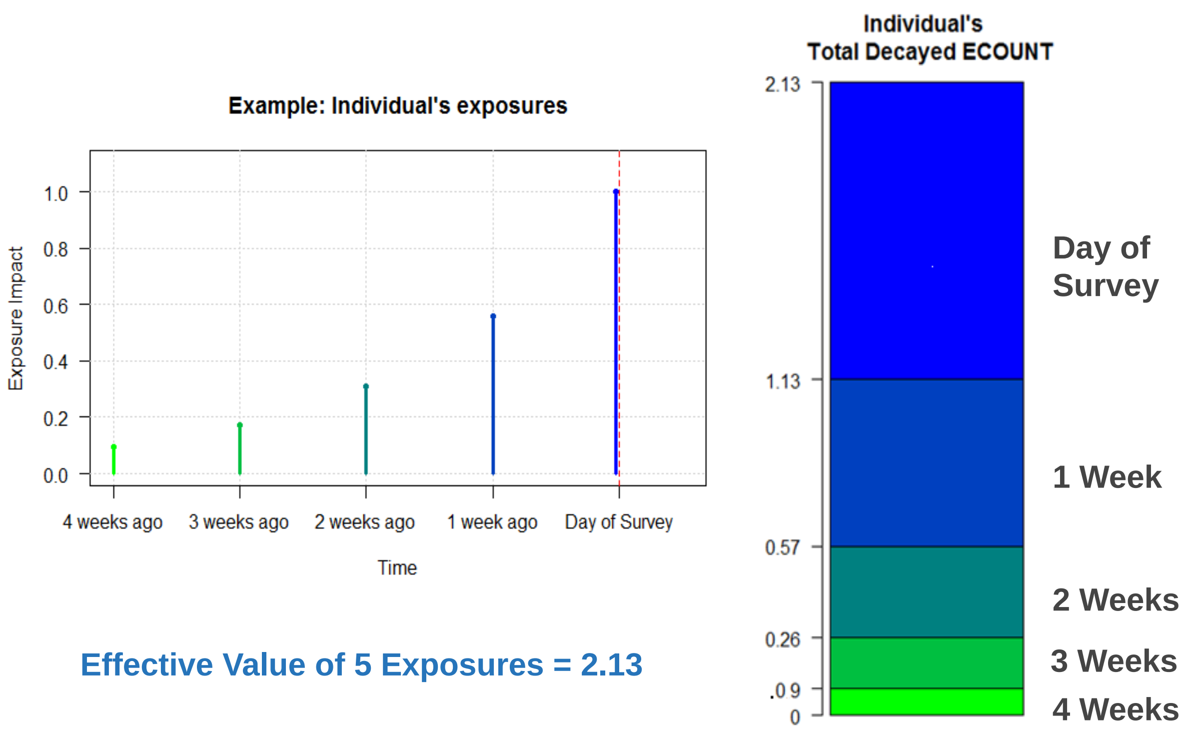


**Figure 3.1:** Sample #1 Survey Exposure to Effectiveness

In the second example, the same five exposures are, on average, considerably closer to the time that the survey was completed. In this case, the five exposures would have about the same level of effectiveness as 2.73 exposures seen on the day of the survey.

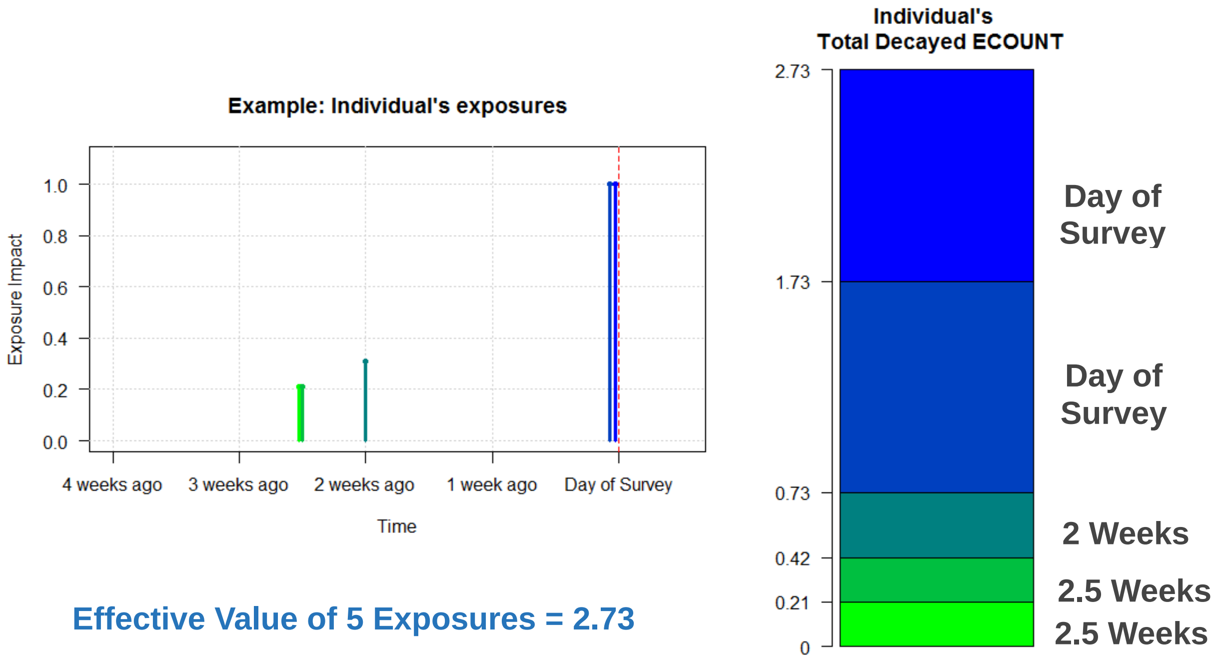


**Figure 3.2:** Sample #1 Survey Exposure to Effectiveness

What remains to be empirically determined, is both the shape of the ad latency curve and the typical length of the latency period. In order to identify the shape of the latency curve, we looked at a variety of functional forms and compared their goodness of fit in the form of an Akaike information criterion (AIC). The AIC is a measure of the relative quality of statistical models for a particular set of data in which the model that minimizes information loss (is most explanatory) will have the lowest AIC value.

comScore looked at eight campaigns selected due to their sufficient variance in the distribution of time between survey and exposure, and relatively large sample size. The variation in exposure times and the large sample sizes allow for better estimates regarding the rate of decay. To measure the rate of decay, comScore's Smart Lift model was leveraged and included a rate of decay parameter. In other words, we looked at the number of ad exposures a respondent sees and when they were seen, to determine whether they had an impact on the measurement of ad effectiveness. The Smart Lift model was run considering different shapes of decay, looking for a variety of different functional forms that an ad could display. Of these decay models, we focused on exponential (half-life), logarithmic, logistic, and linear functions. All of these candidate shapes assume that an ad's effectiveness decreases across time. In order to choose among the candidate models, we ran the SmartLift methodology against each of the candidates and considered which had the greatest likelihood of representing the true shape of the relationship.

Our first criterion was how often model estimation failed. Failure to estimate indicates that a gross model misspecification has likely occurred. If a specific decay shape tends to lead to more estimation failures, a gross misspecification for that data set can be inferred. There were few model failures, but the fewest came from the logarithmic model. None of the model candidates had significant issues with gross misspecification. Our second and third criteria had to do with how well the data fit to each of the different decay shapes as determined by the likelihood. To evaluate this factor, we initially looked at all the data to theorize which shape was most likely to result. Then, we used only those respondents whose most recent exposure was more than six days prior to the survey. We did this because we wanted to select a model that fit, not only recent exposures typical of online ad measurement, but also aged exposures. Surprisingly, the model that had the highest likelihood was the logarithmic model. The best fit curve in every campaign tested was not exponential or logistic in nature as many, including the authors, had assumed to be true. Additionally, the curve that best fit the data overall is concave rather than convex. Instead of having a rapid decline in effectiveness followed by a long trail of minimal effectiveness, the underlying form of the relationship is better described as a modest period of effectiveness followed by a rapid decline to zero value.

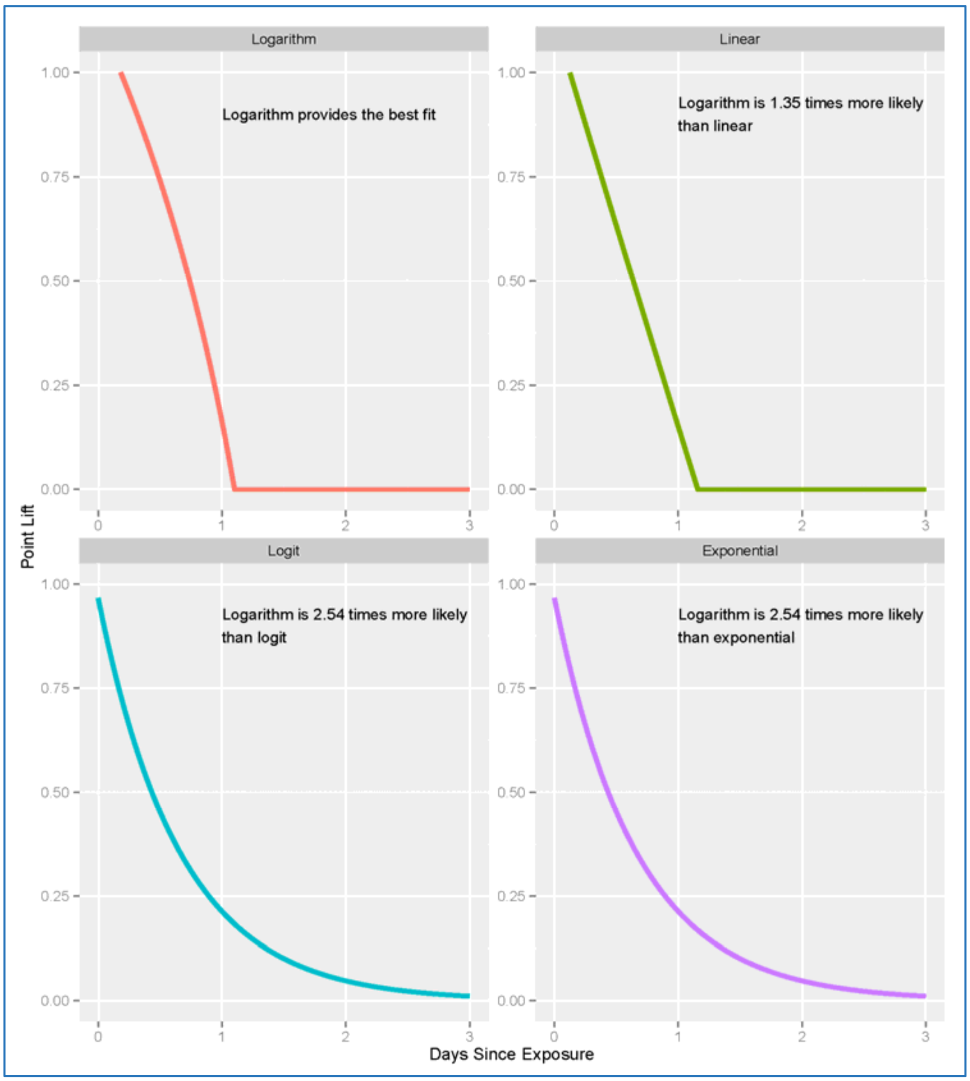


**Figure 4:** Likelihood of Best Fit Candidate Model for Ad Decay

The logarithm being X times more likely than the other candidate models means that the logarithmic decay model is X times as probable as the exponential model to be the best decay function among the candidate decay functions where best is defined as being closest to the true, unknown lift generating process. The logarithmic model was two and a half times as likely as either the logistic or exponential (half-life) models, and nearly one and a half times as likely as the linear model.

With a logarithmic function best defining the decay of an ad, the data suggest that digital advertising, much in contrast to the research on television advertising, extinguishes rather than lingers. Across campaigns, the latency period, as defined by the time it takes for an ad to go to zero effectiveness, was relatively short for display ads and longer for video ads. Among the eight campaigns studied, display ads with statistically significant lifts had latency periods that were generally not more than two or three days long. Some video ads had latency periods typically well over a week long. **Figure 5: *Latency Periods for Video and Display Ads*** demonstrates an example from one of the campaigns running video and display.

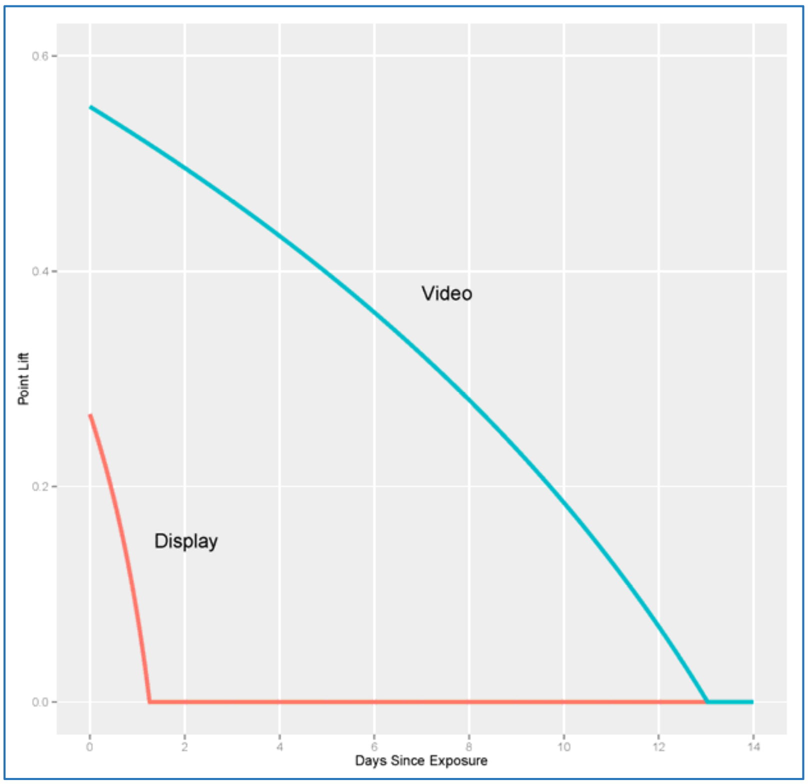


**Figure 5:** Latency Periods for Video and Display Ads

So what do these results mean for media planning and determining optimal frequency of exposure? First, every campaign is different. There is not a standard latency period for ads or ad types – some are more effective than others and some linger longer. Importantly, since these ads do not appear to be effective for long periods of time, there is a substantial difference in the value of serving 10 ads on one day versus one ad every nine days over the course of a 90 day campaign. In the former, although lift is high in the short term, the individual served the large number of ads all at once is at once more likely to see a diminished return on the tenth ad and is likely to see no effect from the ad whatsoever over the course of the lion's share of the campaign. With the latter case, however, "wearout" of the ad is less likely because the ads don't combine in the same way. This hypothesis is supported by the fact that in hundreds of Brand Survey

Lift studies evaluated, there has been no detectable “wearout” found in campaigns with a typically average frequency of two to seven ads per exposed individual. Even more importantly, there is some effect from the ad campaign throughout the period of the media plan with this approach. **Figure 6:** ***Example of Video Ad with Thirteen Day Latency Period Served Every Five Days,*** illustrates the sawtooth-shaped effect of a video ad served every five days, across 30 days total, on self-reported brand metrics. Figure 6 describes the lift that would have occurred had one of our measured campaigns fired their ads to an individual respondent exactly once every fifth day.

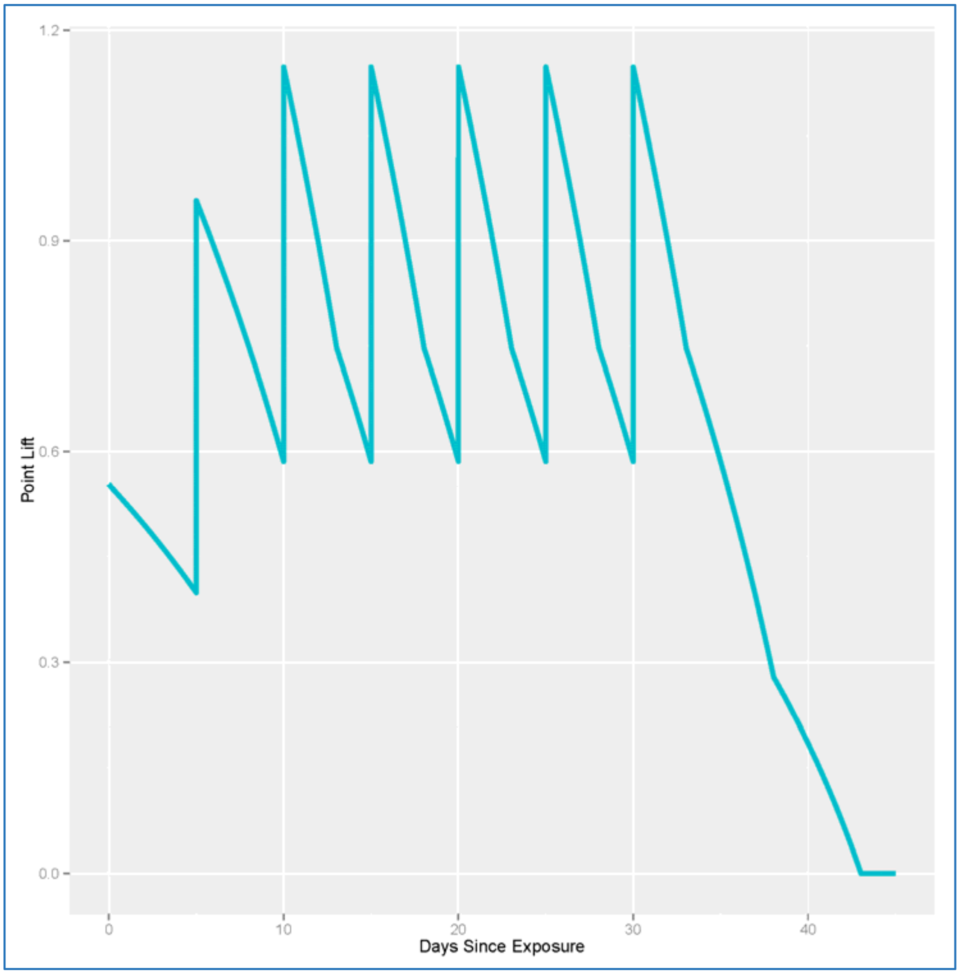


**Figure 6:** Example of Video Ad with Thirteen Day Latency Period Served Every Five Days

Taken together, these results suggest that frequency caps, while valuable for maintaining reach, may serve as too blunt an instrument for optimizing the effectiveness of a media plan. More effective planning would require the ad server to determine how long it has been since the last exposure, rather than simply how many exposures a potential consumer has seen before deciding whether to fire the next ad. If ad servers were to offer this ability to media planners, it would open up a variety of new options to maximize ROI. The following four examples are a few of the potential cases in which optimizing to ad latency periods- rather than frequency of exposure- would produce better results for the relevant brands.

1. **Brand is an impulse purchased item.** For brands which are bought unplanned and at low cost per unit, a sale can come at any time. As a result, such brands would want a consumer affected by the ad at all times. Identifying the latency period for their advertisements and spacing the period between ads served to just under that period would help to extend the value of the campaign to any time in which a consumer might have the opportunity to purchase.
2. **Brand is generally purchased at predictable times.** In some cases, a brand may know that it is primarily purchased at particular times. For example, some brands (e.g. a beer brand) may understand from their purchase data that people are disproportionately more likely to buy their brand on weekends. If the latency period from an ad unit is less than a week, then ads served on a Monday will provide them with far less ROI (if any) than those served on Thursday or Friday. Understanding the latency period of an ad campaign will permit timing that maximizes ROI.
3. **Brand is promoting a special event.** Similar to #2 above, if the brand is a special event such as a one-time promotion, a holiday or weekend sale, or a movie's opening weekend, understanding the latency of the ad campaign will allow the media planner to construct a campaign with a sawtooth graph that builds up the lift in the target audience to its peak just as the event arises. Understanding the latency period will also allow the media planner to save money by identifying the right time to turn off their campaign. Turning it off too soon will reduce sales, and turning it off too late will waste advertising dollars.
4. **General campaign optimization.** Finally, ad latency itself can be evaluated and optimized in just the same way as campaigns can be optimized by publisher or placement. Two ads that create the same initial lift but have very different latency periods will have very different impact. An ad that creates a lift for twice as long is more valuable to the advertiser. Once a wide variety of ads have been measured in this way, guidance for what drives longer ad latencies can be developed to inform creative agencies.

## *Methodology*

Let $Y_i = 1$ if person $i$ has a positive response to a specific survey question and $0$ otherwise. We assume $Y_i$ is a random variable that comes from a Bernoulli distribution with probability $p_i$ of occurring. This probability can be modelled according to the following,

$$p_i = \varphi\left(\alpha + \beta_D \vec{D}_i + \beta_O O_i + \beta_E \sum_{j \in \vec{X}_i} f_\lambda(j)\right)$$

where $\alpha$, $\beta_D$, $\beta_E$, $\beta_O$, and $\lambda$ are all parameters to be estimated; $\vec{D}_i$ is a vector of demographic variables (age, gender, income) for person $i$; $\vec{X}_i$ is a vector of times between survey and exposure for each of person $i$'s exposures to the specified ad-type; $0_\mathrm{i}$ is the number of exposures to the campaign but not the specified ad-type; $\varphi$ is the inverse-logit function used as the link function in logistic regression; $f_\lambda$ is an ad latency function that assigns a value of an exposure based on when it occurred. This value depends on rate of decay parameter $\lambda$.

The following functions $f_\lambda$ were considered:

$$f_\lambda(j) = (1-\lambda)^j \qquad \lambda \in [0,1]$$

$$f_\lambda(j) = \max(log(-\lambda j + e), 0) \qquad \lambda \in [0,\infty)$$

$$f_\lambda(j) = \max(1 - \lambda t, 0) \qquad \lambda \in [0,\infty)$$

$$f_\lambda(j) = 2\ \frac{1}{\exp(\lambda j + \log(2))} \qquad \lambda \in [0,\infty)$$

**Estimation of Parameters**

Assuming $Y_i$ is a Bernoulli random variable with probability $p_i$ and that $Y_i$ are independent across $Y_i$, we may write a likelihood function for the entire set of respondents. Holding the candidate function $f_\lambda$ constant, we can then find the maximum likelihood estimators (MLE) for each of our parameters.

**Selection of Candidate Functions**

We analyzed five campaigns each across several metrics resulting in 85 models ran for each candidate ad latency function. In order to select an appropriate ad latency function $f_\lambda$, we compared how each candidate function performed based on a number of criteria. Since the exact same data was used for each candidate function, differences in estimation or model fit are due to differences in the model specification. Namely, differences in the candidate ad latency function.

### Using the Model

Using our estimated values and setting $\vec{D}_i$ and $O_i$ to their average values, we may estimate the probability of a positive response as a function of the exposure times to a specified ad type. That is,

$$p(\vec{X}) = \varphi\left(\alpha + \beta_D \vec{D}_i + \beta_O \bar{O} + \beta_E \sum_{j \in \vec{X}} f_{\hat{\lambda}}(j)\right)$$

and we define the lift function as,

$$\mathrm{L}(\vec{\mathrm{X}}) = \mathrm{p}(\vec{\mathrm{X}}) - \mathrm{p}(0)$$

Specifying different values of $\vec{\mathrm{X}}$, you may write lift as a function of time for a specified ad with a certain exposure pattern. For example, you may write the lift as a function $g$ of time $t$ for an ad shown every week for three weeks as,

$$g(t) = \begin{cases} L(t)\ for\ t \in [0,7) \\ L(\langle t, t-7 \rangle)\ for\ t \in [7,14) \\ L(\langle t, t-7, t-14 \rangle)\ for\ t \in [14,21) \end{cases}$$